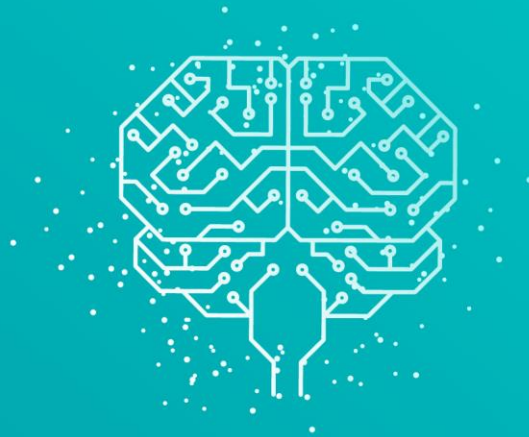


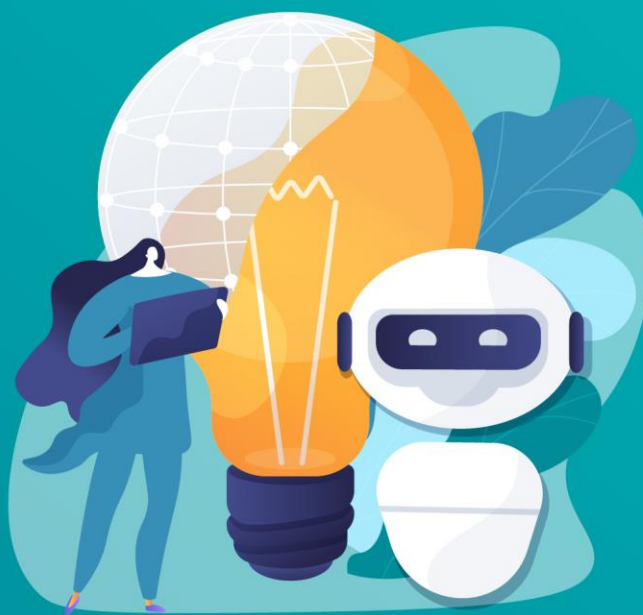


February 2022



Study on copyright and new technologies:

Copyright data management and artificial intelligence



Internal identification

Contract number: LC-01528857

VIGIE number: 2019/0038

EUROPEAN COMMISSION

Directorate-General for Communications Networks, Content and Technology

Directorate I Media Policy

Unit I.2 Copyright

Contact: CNECT-I2@ec.europa.eu

European Commission

B-1049 Brussels

Study on copyright and new technologies: copyright data management and artificial intelligence

SMART 2019/0038

LEGAL NOTICE

The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the Commission. The Commission does not guarantee the accuracy of the data included in this study. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use which may be made of the information contained therein. More information on the European Union is available on the Internet (<http://www.europa.eu>).

PDF

ISBN: 978-92-76-39967-4

doi: 10.2759/570559

KK-09-21-269-EN-N

Manuscript completed in January 2022

First edition

The European Commission is not liable for any consequence stemming from the reuse of this publication.

Luxembourg: Publications Office of the European Union, 2022

© European Union, 2022



The reuse policy of European Commission documents is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

For any reproduction or use of elements on which the European Union does not own the rights, permission may need to be sought directly from the rightholders indicated as the source of the third party content, including: Cover photo: ©Andrey Suslov and ©Visual Generation - iStock Getty Images Plus; CMU Insights (Figure 1); Stage Enterprises (Figures 72, 78, and 79); Rightscom (Figure 75); County Analytics (Figure 80); Mediasert and EBU (Figures 82, 83).

TABLE OF CONTENTS

ABSTRACT	16
EXECUTIVE SUMMARY (EN)	17
EXECUTIVE SUMMARY (FR) / RÉSUMÉ.....	23
1. INTRODUCTION.....	30
2. FIRST PART: COPYRIGHT DATA MANAGEMENT IN EUROPEAN CREATIVE INDUSTRIES ..	31
2.1. Introduction.....	31
2.1.1. Context	31
2.1.2. Scope and approach of the study	33
2.1.3. Methodological remarks	34
2.2. A technical overview of the relevant legal framework	36
2.2.1. International framework	36
2.2.2. Acquis communautaire	37
2.2.3. Preliminary conclusions.....	43
2.3. Literature review on metadata issues and digital technologies	44
2.3.1. Issues related to copyright metadata	44
2.3.2. Impacts of the metadata issues	47
2.3.3. Use of new technologies	48
2.4. An overview of initiatives addressing rights metadata issues	53
2.5. Summary of sectoral interviews on pain points and priorities related to copyright metadata	56
2.5.1. Costs related to rights management	57
2.5.2. Efficiency issues in the context of licensing	63
2.5.3. Challenges concerning payment processes.....	66
2.5.4. Issues related to misappropriation and other rights infringements.....	69
2.6. Empirical findings on challenges related to copyright metadata.....	75
2.6.1. Film & television industry	75
2.6.2. Music industry	81
2.6.3. Publishing industry (books, press, journals, and images)	89
2.6.4. Summary of the empirical findings in the different industries.....	100
2.7. Avenues for future action - strengthening the momentum of ongoing initiatives addressing rights metadata issues	101
2.7.1. Raising general copyright awareness	103
2.7.2. Mastering metadata skills.....	104
2.7.3. Making use of new technologies relevant for metadata management ..	105
2.7.4. Integrating the existing rights data framework(s)	107
3. SECOND PART: COPYRIGHT AND ARTIFICIAL INTELLIGENCE IN THE FIELD OF CREATIVE INDUSTRIES	110
3.1. Introduction.....	110
3.1.1. Context	110
3.1.2. Policy context and motivation of the study.....	110

3.1.3.	Scope of the study	113
3.1.4.	Methodological remarks	115
3.2.	Current use of artificial intelligence in selected creative sectors	117
3.2.1.	Visual arts	118
3.2.2.	Music	126
3.2.3.	Audiovisual and Film	134
3.2.4.	Gaming	140
3.2.5.	Transversal analysis of the use of AI.....	146
3.3.	Challenges of the use of artificial intelligence within the current copyright framework.....	148
3.3.1.	A first overview of copyright, related and sui-generis rights.....	148
3.3.2.	Protected subject matter – when are artificial intelligence inputs and outputs protectable?	150
3.3.3.	Authorship and Ownership – Who is the holder of rights to AI output and how are these rightholders identified?.....	171
3.3.4.	Economic rights and infringement – Which acts in an AI process have copyright relevance?	177
3.3.5.	Moral rights – Which actions in an AI process are relevant for the moral rights under copyright and the performers’ rights?.....	188
3.3.6.	Exceptions (text and data mining) – To which extent do the TDM exceptions apply to AI relevant operations?	192
3.3.7.	Transversal issues – How do the copyright rules relate to other AI-relevant legal regimes?.....	201
3.4.	Policy scenarios.....	203
3.4.1.	Considering the copyright status of the creative input used for AI training	204
3.4.2.	Considering the copyright status of output generated by artificial intelligence	232
4.	SUMMARY AND CONCLUSIONS	270
4.1.	Conclusions on copyright data management in the European creative industries.....	270
4.2.	Conclusions on copyright and artificial intelligence in the field of creative industries	272
5.	ANNEXES	275
5.1.	Annex 1: Exemplary technical challenges related to the <i>acquis communautaire</i> on rights metadata – textual references.....	275
5.1.1.	Copyright in the information society (EC/2001/29) – Article 7 §2	275
5.1.2.	Collective management of copyright (EU/2014/26) – Article 24	276
5.2.	Annex 2: List of sources for the literature review.....	280
5.3.	Annex 3: List of current and ongoing initiatives in data interoperability within the content rights infrastructure.....	282
5.4.	Annex 4: Impact Model	321
5.5.	Annex 5: Some characteristics of rights data management	322
5.5.1.	Prevalence of the content value network	322
5.5.2.	Purposes of rights management information	324
5.5.3.	Works and metadata flows	326
5.5.4.	Drowning in data	329
5.5.5.	Outline of a generic functional architecture	329

5.5.6.	The need for a holistic integration	331
5.5.7.	The lowest common denominator to frame rights metadata	332
5.5.8.	The potential role of Artificial Intelligence	333
5.6.	Annex 6: Outline of an Open Rights Data Framework	334
5.6.1.	Functionality	334
5.6.2.	Technology	335
5.6.3.	Characteristics.....	336
5.7.	Annex 7: An interoperable network of identifiers and schemas	337
5.7.1.	Context of this proposal	337
5.7.2.	What is data interoperability?	338
5.7.3.	ORDF: a network of linked Identifiers.....	339
5.7.4.	Why is the Right “missing”?.....	341
5.7.5.	Why is an independent ORDF needed?	343
5.7.6.	The scope of an ORDF	343
5.7.7.	Specification: LCC-based Rights Declaration schema	344
5.7.8.	Requirements: characteristics of rights management	345
5.7.9.	Requirements: characteristics of rights data	346
5.7.10.	Governance of the ORDF.....	347
5.7.11.	Implementation of the ORDF	348
5.8.	Annex 8: Exemplary potential benefits of an Open Right Data Framework.....	349
5.8.1.	Addressing inaccurate & slow payments to songwriters.....	349
5.8.2.	Addressing the high costs of duplication	351
5.8.3.	Addressing inefficient markets through micro-licensing.....	352
5.8.4.	Addressing misappropriation of film and television programmes.....	353

LIST OF FIGURES

Figure 1: The rights framework in the music industry and implications for music licensing	60
Figure 2: Number and shares of respondents having identified metadata issues (publishing industry)	90
Figure 3: Estimated extent of missing metadata (publishing industry, qualitative estimations).....	91
Figure 4: Estimated extent of missing metadata, by segments books and newspapers/magazines (qualitative estimations)	92
Figure 5: Estimated extent of missing metadata, by segments books and newspapers/magazines (percentages)	92
Figure 6: Quantitative estimations of impacts of missing metadata in the publishing sector	93
Figure 7: Metadata issues experienced by respondents (publishing industry).....	94
Figure 8: Extent to which lack of interoperability and lack of authoritative sources are perceived as an issue (qualitative estimations).....	96
Figure 9: Estimation types impacts by respondents (qualitative estimations).....	96
Figure 10: Impacts perceived due to metadata quality issues.....	97
Figure 11: Professional activity of Delphi survey participants (N=64)	116
Figure 12: Field of expertise of Delphi survey participants (N=64)	117
Figure 13: Screenshot presenting a fraction of the 8,159 results of the “front-facing brown hair joy female with blue eyes”	124
Figure 14: Screenshot from Generated Photos	125
Figure 15: Screenshot from Generated Photos	130
Figure 16: Screenshot from AIVA.....	130
Figure 17: Screenshot from AIVA.....	131
Figure 18: Screenshot from AIVA.....	131
Figure 19: Screenshot from AIVA.....	132
Figure 20: Screenshot from AIVA.....	132
Figure 21: Screenshot from AIVA.....	133

Figure 22: Overview of the style-preserving visual dubbing approach.....	138
Figure 23: TrueSync technology applied to Jack Nicholson’s performance in “A Few Good Men”	138
Figure 24: TrueSync technology applied to Jack Nicholson’s performance in “A Few Good Men”	139
Figure 25: Screenshot from AI Dungeon	143
Figure 26: Screenshot from AI Dungeon	143
Figure 27: Screenshot from AI Dungeon	144
Figure 28: Screenshot from AI Dungeon	144
Figure 29: Screenshot from AI Dungeon	145
Figure 30: Screenshot from AI Dungeon	146
Figure 31: Stakeholders' opinion on the impact of copies of style made by AI vs. by humans	152
Figure 32: Stakeholders' awareness on the absence of protection for AI-generated output	157
Figure 33: Stakeholders' opinion on whether people using AI would protect their creations with related rights.....	165
Figure 34: Stakeholders' opinion on the attribution of authorship when using AI for the production of creative output.....	176
Figure 35: Experts' opinion on the use of moral rights to oppose the use of works to train AI.....	191
Figure 36: Stakeholders' awareness on the TDM exception	197
Figure 37: Stakeholders' opinion on whether the training of AI leads to some copyright-related obstacle (N=64).....	198
Figure 38: Opinion of ECS experts on the cumulative application of exceptions (Total votes=11)	199
Figure 39: Stakeholders' opinion on copyright uncertainties raised by AI practices (N=64)	204
Figure 40: Stakeholders' opinion on scenarios of the copyright scope of protection (N=56 for the first three scenarios, N=28 for the fourth and N=27 for the last one)	207
Figure 41: Stakeholders' opinion on the impacts of the status quo of the copyright scope of protection	211

Figure 42: Stakeholders' opinion on the impacts of policy scenarios of the copyright scope of protection	212
Figure 43: Stakeholders' opinion on whether the opt-out possibility will become an extended practice (N=64)	218
Figure 44: Stakeholders' opinion on scenarios for the communication of the opt-out decision (N=59)	219
Figure 45: Stakeholders' opinion on the suitability of different implementation of the opt-out (N=24)	222
Figure 46: Opinion of ECS experts on the expression of the opt-out decision (Total votes = 17)	223
Figure 47: Stakeholders' opinion on scenarios on the opposition to AI training based on moral rights (N=59)	228
Figure 48: Opinion of ECS experts on the conditions to opposed AI training based on moral rights (Total votes = 10)	228
Figure 49: Stakeholders' opinion on the impact of scenarios of the exercise of moral rights	230
Figure 50: Stakeholders' opinion on scenarios on the protection of AI-generated output (N=57)	233
Figure 51: Opinion of ECS on whether AI-generated output should be offered IP protection	234
Figure 52: Stakeholders' opinion on the consequences of the status quo of the protection of AI output (N=14)	236
Figure 53: Stakeholders' opinion on the consequences of protecting AI-generated cultural output (N=14)	237
Figure 54: Stakeholders' opinion on the consequences of introducing another right next to copyright for AI-generated output (N=22)	239
Figure 55: Stakeholders' opinion on the attribution of rights in case AI-generated output would be protected (N=28)	240
Figure 56: Stakeholders' opinion on the scenarios for the protection of style against AI-generated output and performance rights (N=53)	243
Figure 57: Stakeholders' opinion on the impact of policy scenarios on the protection of style (N=21)	245
Figure 58: Stakeholders' assessment of the effectiveness and manageability of policy scenarios to combat AI mimics (N=21)	247
Figure 59: Opinion of ECS experts on a reform of performers' right (Total votes = 10)	249

Figure 60: Experts' opinion on the scenarios on the protection of AI-generated output and performance rights (N=45)	250
Figure 61: Stakeholders' opinion on the impacts of policy scenarios on the protection of AI-generated output and performance rights (N=17).....	251
Figure 62: Experts' Stakeholders' opinion on the scenarios on the protection of AI-generated output and related rights (N=50)	254
Figure 63: Opinion of ECS experts on reforming of related rights (Total votes = 10) ..	255
Figure 64: Stakeholders' opinion on the impact of adding a condition of investment to related rights (N = 20)	256
Figure 65: Experts' opinion on scenarios on false authorship and authorship presumption (N=52)	262
Figure 66: Opinion of ECS experts on a reform of the presumption of authorship (Total votes = 13)	263
Figure 67: Experts' opinion on the scenarios for the obligation of information on the human/AI generation of creative output (N=59).....	266
Figure 68: Opinion of ECS experts on the imposition of an information obligation towards consumers (Total votes = 12).....	267
Figure 69: Stakeholders' opinion on the impact of imposing an obligation to inform customers that a work is AI-generated (N=26)	268
Figure 70: Exemplary impact model for the music industry - potential metadata challenges and impacts	322
Figure 71: Purposes of RMI.....	325
Figure 72: Flow of data pertaining to work and related subject matter, and counterflow of data pertaining to remuneration – example 1: music industry	327
Figure 73: Flow of data pertaining to work and related subject matter, and counterflow of data pertaining to remuneration – example 2: theatre playwriting	328
Figure 74: Functionality and regulation.....	331
Figure 75: ORDF Entity Model	339
Figure 76: Computer-eye view of an ORDF.....	340
Figure 77: ORDF Entities – public identifiers and schemas	341
Figure 78: Current dataflows of music streaming	350
Figure 79: An attribution and verification middleware	351

Figure 80: Possible new deduplication services	352
Figure 81: Retrieving metadata through open content-dependent identification	353
Figure 82: Marking the TV programme	354
Figure 83: Enforcing rights	354

LIST OF TABLES

Table 1: Categories of metadata	38
Table 2: Overview on current and ongoing initiatives in data interoperability within the content rights infrastructure	283

Abstract

The study addresses two topics related to copyright and new technologies in two parts of the report.

The first part focuses on rights metadata, analysing the current situation in different creative sectors and compiling information on ongoing initiatives in this area. Key findings include that challenges with respect to rights data management persist, despite considerable progress in many domains. The study indicates some broad steps that could contribute to improving the efficiency and transparency of the copyright data system, including measures to build awareness, master metadata, use new technologies and integrating the existing rights data frameworks.

The second part of the study analyses the implications of the use of Artificial Intelligence technologies within the cultural and creative industries for copyright and copyright's related rights. The study finds that new AI solutions are currently developed and applied not only for repetitive tasks but also tasks traditionally considered as involving creative choices. Identified challenges for the copyright system concern, *inter alia*, the scope of the reproduction right, text and data mining exceptions, and challenges of false attribution of AI-created works to a human or of reliance on related rights to circumvent the absence of copyright protection for this type of output.

Executive Summary (EN)

In October 2020, DG CNECT commissioned a consortium of Technopolis Group, Philippe Rixhon Associates, UCLouvain, Crowell&Moring and IMC University of Applied Sciences Krems with the execution of a 9-month “Study on copyright and new technologies: copyright data management and artificial intelligence”.

The study addresses two topics related to the copyright system and new technologies, dealt with in two dedicated parts of the study.

- The **first part** of the study systematically takes stock of the current situation with respect to rights metadata in different creative industries. It attempts to **identify and describe the economic impact of the current situation related to rights metadata**. It also compiles information on the **most important ongoing initiatives** to address some of the identified problems. Finally, the study indicates broad avenues which could contribute to improving functioning of the **copyright data ecosystem**.
- The **second part** of the study focuses on a) **uses** of copyright-protected content as input to feed AI technologies and b) the *copyright implications* of the **production of cultural outputs by or with the assistance of AI**. Furthermore, the study discusses possible **policy scenarios** which might be needed to react to these developments.

All information and views set out in this publication are those of the authors and do not necessarily reflect the official opinion of the Commission.

Summary Part One

The **first part** of this project is focusing on the topic **rights metadata or “rights management information”** as defined in the European Directive 2001/29/EC on the harmonisation of copyright in the information society:

“Rights Management Information means any information provided by rightsholders which identifies the work or other subject matter [], the author or any other rightsholder, or information about the terms and conditions of use of the work or other subject-matter, and any numbers or codes that represent such information. [This] shall apply when any of these items of information is associated with a copy of, or appears in connection with the communication to the public of, a work or other subject matter []”¹.

It analyses the extent to which different **challenges** related to rights metadata can be empirically substantiated, including the availability of rights metadata attached to content, the interoperability between different systems for exchanging metadata, or the authority (i.e., trustworthiness) of sources. In doing so it provides supporting information in the context of the stocktaking document on developing improving the [Copyright Infrastructure](#) issued by the Council of the European Union under the Finnish Presidency

¹ European Community (2001). “Directive 2001/29/EC of the European Parliament and of the Council on the harmonisation of certain aspects of copyright and related rights in the information society”. Article 7 §2.

in December 2019 and the [Action Plan on Intellectual Property](#) adopted by the European Commission in November 2020.

The **empirical basis** for this part of the study consists of interviews and surveys among industry stakeholders as well as analysis of existing studies and further secondary research by the study team. We conducted more than 80 interviews covering three creative industries (film & television, music, and publishing, i.e., books, press, journals, and images) in different Member States of the European Union, Canada, the United Kingdom, and the United States. The interviewees were stakeholders of the digital value network in these creative sectors. The surveys were rolled out in spring 2021 and targeted the main trade associations of four creative industries (music, publishing, film, and TV broadcasting) as well as their members. The surveys consisted of two modules: one general part for industry experts without detailed insights into metadata challenges (mostly open questions on the respondents' perspective on the topic) and one detailed part targeted at metadata experts. Due to the high complexity of the topic, but also to the varying relevance of the topic in different industries, the number of respondents responding to the industry surveys differed widely (between 7 responses for broadcasting and 124 for publishing, with further item non-response for specific questions). Due to the low number of responses, quantitative estimations on the study topics were difficult to obtain. This led the study team to mainly concentrate on reporting qualitative indications.

For the **literature review**, we thoroughly analysed more than 20 core documents, including recent research and working papers, studies, position articles, and communications from the European Commission. In the selection process, we chose recent documents that cover different industries, technologies, and perspectives.

The study team additionally compiled an impressive **list of EU and industry initiatives** that currently play a role, or propose to play a role, in the **content rights infrastructure**, be it on issues of concrete data access and exchange (interoperability), global standards, and identifiers of parties or content or the overall governance of the copyright infrastructure. Any serious future action in the area of the copyright infrastructure must take these into account, either because (like the identifier and metadata standards and many of data access/exchange systems) they already play an essential and established role which must be integrated with others where necessary; or because (like frameworks, working groups and reports) they provide guidance, tools or potential tools for solving aspects of the interoperability challenge. The majority of these initiatives come from specific content or cultural sectors and are thus not cross-sectoral by design.

Key findings of the report suggest that rights data management is – for many reasons – challenging in all creative industries. Based on interviews, surveys and literature review, it can be summarised that the analysed creative sectors are facing data-related challenges in the following four areas:

- **Costs**, in the area of *rights management*
- **Efficiency issues**, in the domain of *licensing*
- **Challenges concerning payments processes**, in the field of *rights remuneration*
- **Risks of misappropriation and other rights infringements**, in the sphere of *rights enforcement*

To be able to describe the challenges in a nuanced way, the study team took a **sectoral approach** in analysing metadata challenges, holding a differentiated perspective on the film, music, and publishing industry. The reason for this is that there are large **differences between industries, but also within industries**, with respect to the identified

importance of metadata challenges and how high they are currently ranked on the industry or policy agenda. The same applies to the current status quo of metadata initiatives in the different industries.

As an example, in the music industry findings from existing evidence (information from e.g. CMOs) and expert interviews for this study among various stakeholder groups suggest that – while the work of many initiatives (for example CISAC and DDEX) has improved data exchange processes – challenges of imperfect or disputed rights metadata information are still a challenge. This implies, for example, that at least in some cases the music industry “spends an inordinate amount of time correcting errors and resolving disputes which hold up payments to music creators”². This is less the case for newer recordings since awareness of the importance of rights metadata has improved, but in any case still seems to be an issue for older works. However, stakeholders from the recorded music sector also signalled that rights metadata issues are for them nowadays less problematic than, for example, the inaccuracy or lack of usage metadata provided by online platform services.

In the publishing sector, the results of interviews and surveys suggest that in areas such as the digital news publishing sector, challenges are more prevalent than in book publishing. In the former, issues such as a lack of granular attribution of copyright ownership for photographs used on news websites can lead to challenges regarding the remuneration of rightholders. In the latter, rightholder identification (e.g., the author of a book) is more straightforward and rights management processes therefore easier. The granularity and the degree to which copyright-protected content is embedded in creative works as well as rightholder structures of a creative work (through iterative contributions or co-authorships) make a substantial difference.

In the audiovisual sector, problems of compatibility of descriptive as well as rights management data were identified – despite recent developments to increase compatibility through a harmonisation of the registration process between the standard content identifiers developed within the industry, EIDR and ISAN. Also, a transparent exchange of usage metadata seems currently not always to be a given. Niche players in the film and television industry (independent film producers), but not the major studios we spoke to, raised concerns in this regard. In film production, stakeholders mention interoperability issues in rights management systems. Moreover, in most cases, there is no obligation to use the EIDR or ISAN standard identifiers and not all players are using them to identify their works. Such standard identifiers do not cover the rights, which does not facilitate licensing.

The study concludes that different avenues for future action could help improve the current situation: **raising copyright awareness** in general (e.g., on the side of creators and rightholders, and on the side of users and consumers) would help clarify the importance the copyright system in general for creatives in all industries. More specifically, initiatives to help raising **awareness and skills specifically related to rights metadata** seem to be important. The study results suggest that individuals in the creative content industries (especially creators themselves) have a relatively weak understanding of what metadata are and how to handle them. This lack of expertise and attention ends up in non-exhaustive metadata at various distribution stages of works and related subject matters. A “metadata education programme” could therefore be promoted in order to improve metadata right from the start of the creative process. In order to improve metadata quality and rights management processes the study authors also suggest to further take into account new

² <https://www.prsformusic.com/m-magazine/news/auddly-launches-metadatapays-campaign/>

technological developments such as Artificial Intelligence and Distributed Ledger technologies (for example via the European Blockchain Services Infrastructure). Finally, the study authors are of the opinion that, in the long run, a **cross-sector rights data network** could bridge gaps between standard content identifiers such as ISRC, ISWC, ISBN or ISAN and digital manifestations of the content they denote. This would increase interoperability also between different media or content sectors. The ultimate objective of this endeavour would be to break the silos between different creative industries and improve the efficient rights data management and licensing across sectors. It could help to release even more of the digital potential of Europe's creative sectors.

Summary Part Two

The **second part** of the study analyses the implications for copyright and the related rights of the increasing use of AI technologies within the cultural and creative industries. Building on a review of several experiences with AI tools and concrete use cases, on a traditional study of the legal sources, on interviews with legal experts and stakeholders and lastly on a Delphi survey among legal experts and industry stakeholders, this part of the study focuses on the challenges raised by those tools and use cases for the EU copyright and related rights framework.

Over the last few years, AI solutions have been deployed across **different industries** and in a **wide range of applications**. The cultural and creative sector is no exception: some AI tools assist or complete the highly human process of creation; more often, they are used for improving the production of successful cultural artefacts or the consumer experience, e.g. through well-targeted recommendations. The reliance on AI technologies *for or during the creative process* might yet challenge copyright and/or related rights. The study distinguishes upstream or **input issues**, i.e. those related to the use of protected content as inputs for an AI application, from the downstream or **output issues**, i.e. those related to the musical, graphical, audiovisual or other cultural content that results from the use of the AI application. On the input side, AI applications may be trained with large datasets of creative content enjoying protection under copyright and related right, which prompts the question whether the rightholders' authorisation is needed for such use. On the output side, the AI applications can generate cultural content without any significant human contribution, which raises the question whether such outcome is protected under copyright or a related right. Other issues come to mind: Is there a need for additional incentives (in the form of copyright-like rights) to use AI tools for generating cultural outputs? Should the related investments in AI solutions be protected by an exclusive right or just promoted through funding? Are there authorship or ownership issues?

The study is structured as follows. It first shows how some AI applications are used in practice (part 3.2). The illustrations help to understand how this developing practice may impact the various stakeholders (creators, artists, producers, distributors, etc.) in the cultural and creative industry. This assessment is done in four creative sectors, namely visual arts, music, audiovisual & film, and video games. This allows then to identify possible issues with and challenges to copyright and related rights (part 3.3). Finally, some policy options are examined (part 3.4) to address these challenges.

Concerning the deployment of AI solutions in the cultural sector, the research demonstrates on the basis of concrete examples, that the overall reliance on these tools is increasing even if the **degree of adoption of AI solutions** varies significantly from sector to sector, for instance with a clear use of AI tools for upgrading video games, as well as for generating photos and faces for advertisements or "elevator music". Furthermore, AI solutions are not only deployed for repetitive or mechanical duties but also for tasks leading

to outputs which appear original and imaginative – and therefore, traditionally considered as within the sole realm of humans. The study focuses on the creations that are produced autonomously by the AI solution, with no or little human intervention.

Most of the AI applications appear to be **marketed online “as a service”** (one might even refer to the nascent market for “Creation as a Service” or “CaaS”). This business model confers factual control of the use of the AI solution and the production of AI output to the AI developers, who can consequently protect their revenues (and terminate the service in case of non-payment). The features of this business model should be considered when assessing the need for protecting the AI output under copyright or related rights.

On the **input side**, several challenges are examined, including the limits of the exclusivity conferred by copyright and the related rights.

Firstly, **the scope of the reproduction right** is still in the process of being defined by the European courts, especially when purely technical or intermediate copies are made such as within the process for training an AI algorithm through the analysis of protected elements. The recent teleological interpretation of the reproduction right and of the extraction right under the database right by the Court of Justice of the EU (CJEU) opens new avenues for the interpretation and application of those exclusive rights. It remains to be seen how the case law will interpret the notion of reproduction and whether it applies to intermediate and technical copies made during the process of feeding an AI tool. The scope of the reproduction right might indeed be resized so as to permit some uses that do not lead to an output in which the protected elements contained in the input are visible or audible.

Secondly, it appears from the consultation of the stakeholders that the expected application of the **TDM exceptions**, in particular of the TDM exception for other purposes than scientific research, raises some interrogations, especially concerning the way opt-out decisions should be communicated. In any case, the transposition of the TDM exceptions within the national laws should be carefully monitored to avoid diverging interpretations. Clarification as to the means and modes for expressing the opt-out under Article 4 DSM Directive might result from the development of good practices when the TDM exceptions will come into force.

Thirdly, the **moral rights** attached to copyright (and to the performers’ right) have not been harmonised at EU level. This could lead to diverging situations where some member states allow authors and performers to exercise their moral rights (in particular the right of integrity) to oppose the use of their works or performances as AI inputs. One way to address this is to clarify that the moral rights cannot block the application of harmonised exceptions (such as for TDM). A more ambitious approach would be to (partly) harmonise the moral rights.

On the **output side**, the AI-generated output is not protected under copyright **in the absence of human creative choices**. The research, interviews, and surveys conducted within the study indicate firstly that no incentive for the use of AI tools in the creative process in the form of additional exclusive rights appears necessary. The already broad deployment of AI tools in the creative context confirms this. Also, the feedback received seems to indicate that an additional right in favour of machine-generated outputs might have negative impacts on the traditional creative sectors. The study concludes that a new related right for AI-generated outputs is not desirable.

Secondly, even if advanced AI applications are increasingly capable of approximating the style of human-made works or performances, the scope of copyright should not be extended to offer protection to the **style** of an author or of a performer, unless some significant and recognizable features of a protected work or performance are reproduced in the AI output. The protection of a creator's style would indeed amount to a significant extension of copyright scope and would limit in a disproportionate way the freedom of expression and freedom of art downstream in the artistic process. Under national law, other remedies (e.g. image rights, personality rights or unfair practices) may be available and, some harmonisation at EU level of the claim of parasitism as an unfair commercial practice could be considered.

Thirdly, the absence of copyright protection for an AI autonomously generated output could leave **artistic performances** without protection under the law of certain member states, which may require that a 'work' protected under copyright be performed for the performance to be protected under the related rights. The human performance of an AI creation would not meet that condition and consequently the human performer would then be left without protection. To avoid this, the study suggests to consider a harmonised definition of 'performance' as the subject matter of the performers' right. This definition would not require the performed subject matter to be protected under copyright (while making sure that the scope of the performers' protection is not stretched to cover activities with little cultural interest).

Fourthly, even if autonomously generated outputs are excluded from copyright protection, they might, however, enjoy protection **under the related rights of phonogram producers, film producers, or press publishers** (if the outputs respectively qualify as sound or audiovisual fixation or as press publication), even if the production of an AI output has little to do with the traditional activities of producers or publishers and does not necessarily require a similar investment. To avoid that some of those related rights are used to circumvent the copyright policy trade-off, the study proposes to impose a condition of investment so that only the fixations for which a certain investment (possibly with a de minimis threshold) was made could trigger the application of the related rights of phonogram and film producers or publishers.

Fifthly, in the (still rare) cases of AI autonomously generated outputs, the **false attribution of authorship** to a human might, in practice, allow to circumvent the absence of copyright protection for this type of output. It might indeed suffice to claim authorship (by mentioning a person as the author) for the person to enjoy (unlawfully) the presumption of authorship and consequently, in fact, copyright protection over an AI-generated creation, knowing that this presumption is difficult or even impossible to rebut. However, the status quo concerning the authorship presumption should reasonably be maintained. A restriction or abolishment of this presumption would indeed be excessive and could have negative effects for the human creators on whom the burden of establishing authorship would lie.

Lastly, the study reflects on the evidence that the fact of knowing that an art piece is created by a human or by an AI system might affect the perception and the experience of the public. The study nevertheless points to the conclusion that no **information obligation** concerning the use of an AI solution for the development of the work should be added *within* the copyright framework. Such a legal obligation would indeed raise issues regarding its scope and its impact on the creators' artistic freedom and their personality rights. The study does not enquire about the adequacy of imposing an information obligation in other bodies of law, such as consumer law.

Executive Summary (FR) / Résumé

En octobre 2020, la DG CNECT a confié à un consortium composé de Technopolis Group, Philippe Rixhon Associates, UCLouvain, Crowell&Moring et IMC University of Applied Sciences Krems la réalisation d'une « Étude sur le droit d'auteur et les nouvelles technologies : gestion des données du droit d'auteur et intelligence artificielle » d'une durée de 9 mois.

L'étude aborde deux sujets liés à l'écosystème du droit d'auteur et aux nouvelles technologies, traités en deux parties :

- La **première partie** de l'étude fait le point sur la situation actuelle de l'information en matière de droits dans différentes industries créatives. Elle tente d'**identifier et de décrire l'impact économique de la situation actuelle en ce qui concerne les métadonnées**. Elle compile également une **liste d'initiatives importantes** visant à résoudre certains des problèmes identifiés. Enfin, l'étude indique de grandes pistes qui pourraient contribuer à améliorer le fonctionnement de **l'écosystème des données sur les droits d'auteur**.
- La **deuxième partie** de l'étude se concentre sur a) **l'utilisation** de contenus protégés par le droit d'auteur pour alimenter les technologies d'IA et b) les *implications en matière de droit d'auteur* de la **production de biens culturels par ou avec l'aide de l'IA**. En outre, cette partie examine des **scénarios politiques** qui pourraient être nécessaires pour encadrer ces développements.

Toutes les informations et opinions présentées dans cette publication sont celles des auteurs et ne reflètent pas nécessairement l'opinion officielle de la Commission.

Résumé de la première partie

La **première partie** de ce projet se concentre sur le sujet des **métadonnées de droits ou « informations sur le régime des droits »** telles que définies dans la directive européenne 2001/29/CE sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information :

« On entend par « information sur le régime des droits » toute information fournie par des titulaires de droits qui permet d'identifier l'œuvre ou autre objet protégé [], l'auteur ou tout autre titulaire de droits. Cette expression désigne aussi les informations sur les conditions et modalités d'utilisation de l'œuvre ou autre objet protégé ainsi que tout numéro ou code représentant ces informations. [Ceci] s'applique lorsque l'un quelconque de ces éléments d'information est joint à la copie ou apparaît en relation avec la communication au public d'une œuvre ou d'un objet protégé [] »³.

Elle analyse en quelle mesure différents **défis** liés aux métadonnées de droits peuvent être empiriquement définis, notamment la disponibilité des métadonnées de droits et leur rattachement aux contenus, l'interopérabilité entre différents systèmes d'échange de

³ Communautés européennes (2001). Directive 2001/29/CE du Parlement Européen et du Conseil sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information. Article 7 §2.

métadonnées ou l'autorité (c'est-à-dire la fiabilité) des sources. Ce faisant, elle fournit des informations contextuelles se rapportant au [document d'inventaire sur l'amélioration de l'infrastructure du droit d'auteur](#) publié par le Conseil de l'Union européenne sous la présidence finlandaise en décembre 2019 et au [plan d'action en faveur de la propriété intellectuelle](#) adopté par la Commission européenne en novembre 2020.

La **base empirique** de cette partie de l'étude est constituée d'entretiens et d'enquêtes auprès des acteurs de l'industrie, d'une analyse des études existantes et de recherches supplémentaires menées par notre équipe. Nous avons eu plus de 80 entretiens couvrant trois industries créatives (cinéma et télévision, musique, et édition, c'est-à-dire livres, presse, revues scientifiques et images) dans les différents États membres de l'Union européenne, au Canada, au Royaume-Uni et aux États-Unis. Les interlocuteurs étaient des acteurs du réseau numérique de ces secteurs créatifs. Les enquêtes ont été menées au printemps 2021 et ciblaient les membres des principales associations professionnelles de quatre industries créatives (cinéma, télédiffusion, musique et édition). Les enquêtes se composaient de deux modules : une partie générale destinée aux experts du secteur qui ne disposent pas d'informations détaillées sur les défis liés aux métadonnées (essentiellement des questions sur le point de vue des répondants) et une partie plus détaillée destinée aux experts en métadonnées. En raison de la grande complexité du sujet, mais aussi de la variation de sa pertinence dans les différents secteurs, les nombres de répondants aux enquêtes sectorielles ont été très différents (entre 7 réponses pour la radiodiffusion et 124 pour l'édition). Vu le faible nombre de réponses, il a été difficile d'obtenir des estimations quantitatives sur les sujets étudiés. Cela a conduit les experts à se concentrer sur les rapports d'indications qualitatives.

Pour l'**examen de la littérature**, nous avons analysé de manière approfondie plus de 20 documents fondamentaux, y compris recherches récentes, documents de travail, études, articles de position, et communications de la Commission européenne. Nous avons choisi des documents récents qui couvrent différentes industries, technologies et perspectives.

L'équipe chargée de l'étude a également dressé une **liste non-exhaustive du nombre impressionnant d'initiatives de l'UE et de l'industrie** qui jouent actuellement un rôle, ou proposent de jouer un rôle, dans l'amélioration de l'infrastructure des métadonnées sur les droits, qu'il s'agisse d'actions concrètes sur l'accès et l'échange de données (interopérabilité), sur les identifiants des parties prenantes ou des contenus, ou sur la gouvernance de l'infrastructure. Toute action sérieuse dans le domaine de l'infrastructure des données du droit d'auteur doit tenir compte de ces initiatives, soit parce qu'elles jouent déjà un rôle essentiel qui doit être intégré à d'autres si nécessaire (comme les normes relatives aux identifiants, métadonnées et nombreux systèmes d'accès et d'échange de données), soit parce qu'elles fournissent des orientations, des outils ou des approches pour résoudre certains aspects d'interopérabilité (comme les cadres, les groupes de travail et les rapports). La majorité de ces initiatives viennent de secteurs spécifiques des médias et de la culture et ne sont donc pas intersectoriels.

Les principales conclusions du rapport suggèrent que la gestion des données relatives aux droits est, pour de nombreuses raisons, en chantier dans la plupart des industries créatives. Sur base des entretiens, des enquêtes et de l'analyse documentaire, on peut résumer que les secteurs créatifs analysés sont confrontés à des défis liés aux données dans les quatre domaines suivants :

- Les **coûts**, dans le domaine la *gestion informatique* des droits
- L'**efficacité**, dans le domaine de l'*octroi de licences*

- Les **procédures de paiement**, dans le domaine de la *rémunération des droits*
- Les risques de **détournement et autres atteintes aux droits**, dans le domaine de l'*application des droits*.

Pour décrire les défis de manière nuancée, les experts ont adopté une **approche sectorielle** différenciée dans l'analyse des industries du film, de la musique et de l'édition. En effet, il y a de grandes **disparités entre secteurs, et aussi au sein d'un même secteur**, en ce qui concerne l'importance et les priorités accordées aux métadonnées et à la place qu'elles occupent actuellement dans les agendas industriel ou politique. Il en va de même pour le statu quo des initiatives en matière de métadonnées dans les différents secteurs.

Par exemple, dans l'industrie de la musique, l'analyse de la situation, sur base d'informations communiquées par des sociétés de gestion collective et d'entretiens avec des parties prenantes d'horizons divers, suggère que trop d'informations sur les droits restent imparfaites ou contestées bien que le travail de nombreuses initiatives (par exemple DDEX) ait amélioré les processus d'échange des données s'y rapportant. Cela implique, par exemple, qu'au moins dans certains cas, l'industrie de la musique « passe un temps excessif à corriger des erreurs et à résoudre des litiges qui retardent les paiements aux créateurs et interprètes ». C'est moins le cas pour les enregistrements récents car la prise de conscience de l'importance des métadonnées s'est améliorée ; mais cela reste souvent un problème pour les œuvres plus anciennes. Par ailleurs, les représentants des maisons de disques ont signalé que les métadonnées de droits posent aujourd'hui moins de problèmes que, par exemple, l'inexactitude ou le manque de métadonnées d'utilisation fournies par les services de plateforme en ligne.

Dans le secteur de l'édition, les résultats des entretiens et des enquêtes suggèrent que les défis sont plus fréquents dans le secteur de la presse numérique que dans celui de l'édition de livres. Dans le premier cas, l'absence d'attribution granulaire des droits pour les photographies utilisées sur les sites d'information peut, par exemple, entraîner des difficultés quant à la rémunération des ayants-droits. Dans le second cas, l'identification des ayants-droits (par exemple, l'auteur d'un livre) est plus simple et la gestion des droits est donc plus facile. La granularité et la modularité du contenu protégé par le droit d'auteur ainsi que les relations entre ayants-droits d'une œuvre créée par contributions itératives et collaboratives, définissent le niveau de ces défis.

Dans le secteur de l'audiovisuel, nous avons identifié des problèmes d'incompatibilité des données nécessaires à la description des contenus et à la gestion des droits malgré les récents développements visant à harmoniser par exemple les processus d'enregistrement EIDR et ISAN, les identifiants de contenu standards développés dans l'industrie. De même, l'échange transparent de métadonnées d'utilisation ne semble pas toujours acquis à l'heure actuelle. Ce sont les producteurs indépendants et non les grands studios avec lesquels nous nous sommes entretenus qui nous ont fait part de leurs préoccupations à cet égard. Dans la production cinématographique, les problèmes d'interopérabilité des systèmes de gestion de droits sont mentionnés. De plus, dans la plupart des cas, il n'y a pas d'obligation d'utiliser les identifiants standards EIDR ou ISAN et tous les acteurs ne les utilisent pas pour identifier leurs œuvres. Ces identifiants standard ne couvrent pas les droits, ce qui ne facilite pas l'octroi de licences.

L'étude conclut que différentes pistes d'actions pourraient contribuer à améliorer la situation actuelle : la **sensibilisation au droit d'auteur** en général (du côté des créateurs et des ayants-droits autant que du côté des utilisateurs et des consommateurs) aiderait à

démontrer l'importance des droits d'auteur pour les créateurs de tout secteur. Plus précisément, les initiatives visant à **développer les compétences spécifiquement liées aux métadonnées des droits** semblent importantes. Les résultats de l'étude suggèrent que les acteurs du monde des médias et de la culture (en particulier les créateurs eux-mêmes) ont une compréhension relativement faible de ce que sont les métadonnées et de la façon de les traiter. Ce manque d'expertise et d'attention se traduit par des métadonnées non exhaustives à différents stades de la distribution des œuvres et de leurs représentations. Un « programme d'éducation » pourrait donc être mis en place pour améliorer la qualité des métadonnées dès le début du processus de création. Afin d'améliorer la qualité des métadonnées et les processus de gestion de droits, les auteurs de l'étude suggèrent également de prendre en compte les nouveaux développements technologiques tels que l'intelligence artificielle et les technologies de registres distribués (par exemple via l'infrastructure européenne de services blockchain). Enfin, les auteurs sont d'avis qu'un **réseau transsectoriel de données sur les droits** pourrait relier identifiants standards tels que ISRC, ISWC, ISBN ou ISAN et les manifestations numériques du contenu qu'ils désignent. L'accessibilité et l'interopérabilité s'en trouveraient grandement accrues entre les différents secteurs des médias et de la culture. L'objectif ultime de cette entreprise serait de briser les cloisonnements entre différents secteurs et juridictions et ainsi d'améliorer l'efficacité de la gestion des données relatives aux droits et à l'octroi de licences. Elle pourrait contribuer à libérer encore davantage le potentiel numérique des secteurs créatifs européens.

Résumé de la deuxième partie

La **deuxième partie** de l'étude analyse les effets sur le droit d'auteur et les droits voisins de l'utilisation croissante des technologies d'IA dans les secteurs culturels et de la création. S'appuyant sur un examen de plusieurs expériences avec des outils d'IA et des cas d'utilisation concrets, sur une étude traditionnelle de sources juridiques, sur des entretiens avec des experts juridiques et des parties prenantes et enfin sur une étude *delphi*, cette partie de l'étude se concentre sur les défis amenés par ces outils et cas d'utilisation pour le cadre européen du droit d'auteur et des droits voisins.

Au cours des dernières années, des solutions d'IA ont été déployées dans **différents secteurs** et dans un large éventail d'applications. Le secteur culturel et de la création n'y fait pas exception : certains outils d'IA assistent ou complètent le processus éminemment humain de la création ; ils sont généralement utilisés pour améliorer la production d'objets culturels ou l'expérience du consommateur, à travers des recommandations ciblées par exemple. Le recours aux technologies de l'IA pour ou pendant le processus de création peut néanmoins soulever des difficultés en matière de droits d'auteur et/ou de droits voisins. L'étude fait la distinction entre les **problématiques en amont** ou d'*input*, à savoir celles liées à l'utilisation de contenus protégés comme *inputs* pour une application d'IA, et les **problématiques en aval** ou d'*output*, à savoir celles liées aux contenus musicaux, graphiques, audiovisuels ou autres contenus culturels qui résultent de l'utilisation de l'application d'IA. En ce qui concerne l'*input*, les applications d'IA peuvent être entraînées avec de grands ensembles de données de contenu créatif bénéficiant de la protection du droit d'auteur et des droits voisins, ce qui soulève la question de savoir si l'autorisation des titulaires de ces droits est requise pour une telle utilisation. En ce qui concerne l'*output*, les applications d'IA peuvent générer du contenu culturel sans aucune contribution humaine significative, ce qui soulève la question de savoir si ces produits sont protégés par le droit d'auteur ou un droit voisin. D'autres questions se posent : Faut-il prévoir

des incitations supplémentaires (sous la forme de droits similaires au droit d'auteur) pour utiliser les outils d'IA afin de générer des produits culturels ? Les investissements dans les solutions d'IA doivent-ils être protégés par un droit exclusif ou simplement encouragés par des financements ? Y a-t-il des problèmes de paternité ou de propriété ?

L'étude est structurée comme suit. Elle présente d'abord comment certaines applications d'IA sont utilisées dans la pratique (partie 3.2). Les illustrations permettent de comprendre comment cette pratique en développement peut affecter les différentes parties prenantes (créateurs, artistes, producteurs, distributeurs, etc.) de l'industrie culturelle. Cette analyse est réalisée dans quatre secteurs de la création, à savoir les arts visuels, la musique, l'audiovisuel & le cinéma, et les jeux vidéo. Cela permet ensuite d'identifier les potentiels problèmes et défis liés au droit d'auteur et aux droits voisins (partie 3.3). Enfin, certaines options de politiques sont examinées (partie 3.4) pour relever ces défis.

En ce qui concerne le déploiement de solutions d'IA dans le secteur culturel, la recherche démontre, sur la base d'exemples concrets, que ces outils sont utilisés plus fréquemment, même si le **degré d'adoption des solutions d'IA** varie considérablement d'un secteur à l'autre, avec par exemple une utilisation manifeste des outils d'IA pour l'amélioration des jeux vidéo, ainsi que pour la création de photos et de visages pour les publicités ou de « musique d'ascenseur ». En outre, les solutions d'IA ne sont pas seulement déployées pour des tâches répétitives ou mécaniques, mais aussi pour des tâches aboutissant à des résultats qui se révèlent originaux et imaginatifs - et donc traditionnellement considérés comme relevant du seul domaine des humains. L'étude focalise sur les créations produites de façon autonome par les outils d'IA, sans intervention humaine significative.

La plupart des applications d'IA semblent être **commercialisées en ligne « en tant que service »** (un marché naissant de la « *Creation as a Service* » ou « CaaS » peut être observé). Ce modèle commercial attribue le contrôle factuel de l'utilisation de la solution d'IA et de la production de produits d'IA aux développeurs d'IA, qui peuvent ainsi protéger leurs revenus (et mettre fin au service en cas de non-paiement). Les caractéristiques de ce modèle commercial devraient être prises en compte lors de l'évaluation de la nécessité de protéger le produit de l'IA par le droit d'auteur ou les droits voisins.

En ce qui concerne **l'input**, plusieurs défis sont examinés, notamment les limites de l'exclusivité conférée par le droit d'auteur et les droits voisins.

Tout d'abord, le **champ d'application du droit de reproduction** est toujours en cours de clarification par les juridictions européennes, notamment lorsque des copies purement techniques ou intermédiaires sont réalisées, comme dans le cadre du processus d'apprentissage d'un algorithme d'IA par l'analyse d'éléments protégés. L'interprétation téléologique récente du droit de reproduction et du droit d'extraction dans le cadre du droit des bases de données par la Cour de Justice de l'UE (CJUE) ouvre de nouvelles voies pour l'interprétation et l'application de ces droits exclusifs. Il reste à voir comment la jurisprudence interprétera la notion de reproduction et si elle s'applique aux copies intermédiaires et techniques réalisées au cours du processus d'alimentation d'un outil d'IA. Le champ d'application du droit de reproduction pourrait en effet être réaménagé de manière à permettre certaines utilisations qui ne

conduisent pas à un *output* dans lequel les éléments protégés contenus dans l'*input* sont visibles ou audibles.

Deuxièmement, il ressort de la consultation des parties intéressées que l'application prévue des **exceptions au TDM**, en particulier de l'exception au TDM à des fins autres que la recherche scientifique, soulève certaines interrogations, notamment en ce qui concerne la manière dont les décisions d'*opt-out* doivent être communiquées. En tout état de cause, la transposition des exceptions au TDM dans les législations nationales doit être surveillée de près pour éviter des interprétations divergentes. La clarification des moyens et des modes d'expression de l'*opt-out* en vertu de l'article 4 de la directive DSM pourrait résulter du développement de bonnes pratiques lorsque les exceptions au TDM entrent en vigueur.

Troisièmement, les **droits moraux** attachés au droit d'auteur (et au droit des artistes-interprètes) n'ont pas été harmonisés au niveau de l'UE. Cela pourrait conduire à des situations divergentes dans lesquelles certains États membres autorisent les auteurs et les artistes-interprètes à exercer leurs droits moraux (en particulier le droit à l'intégrité) pour s'opposer à l'utilisation de leurs œuvres ou de leurs prestations comme *input* de l'IA. Une façon de résoudre ce problème est de prévoir que les droits moraux ne peuvent pas empêcher l'application d'exceptions harmonisées (comme pour le TDM). Une approche plus ambitieuse consisterait à harmoniser (partiellement) les droits moraux.

En ce qui concerne l'**output**, les produits générés par l'IA ne sont pas protégés par le droit d'auteur **en l'absence de choix créatifs humains**. Les recherches, entretiens et enquêtes menés dans le cadre de l'étude indiquent premièrement qu'aucune incitation à l'utilisation d'outils d'IA dans le processus de création, sous la forme de droits exclusifs supplémentaires, ne semble nécessaire. Le déploiement déjà large des outils d'IA dans le contexte de la création le confirme. Par ailleurs, les réactions reçues semblent indiquer qu'un droit supplémentaire en faveur des produits générés par des machines pourrait avoir des effets négatifs sur les secteurs traditionnels de la création. L'étude conclut qu'un nouveau droit connexe pour les produits générés par l'IA n'est pas souhaitable.

Deuxièmement, même si les applications avancées de l'IA sont de plus en plus capables de se rapprocher du style d'œuvres ou de prestations créées par les humains, le champ d'application du droit d'auteur ne devrait pas être étendu pour offrir une protection au **style** d'un auteur ou d'un artiste-interprète, à moins que certaines caractéristiques significatives et reconnaissables d'une œuvre ou d'une prestation protégée ne soient reproduites dans le produit de l'IA. La protection du style d'un créateur équivaldrait en effet à une extension significative du champ d'application du droit d'auteur et limiterait de manière disproportionnée la liberté d'expression et la liberté artistique en aval du processus artistique. En vertu du droit national, d'autres moyens (par exemple, les droits à l'image, les droits de la personnalité ou les pratiques déloyales) sont envisageables et une certaine harmonisation au niveau européen de l'action en parasitisme en tant que pratique commerciale déloyale pourrait être envisagée.

Troisièmement, l'absence de protection du droit d'auteur pour un produit généré de manière autonome par l'IA pourrait priver certaines **prestations artistiques** de toute protection en vertu de la législation de certains États membres, qui exigent parfois qu'une « œuvre » protégée par le droit d'auteur soit interprétée pour que la prestation

soit protégée par les droits voisins de l'artiste interprète. L'interprétation humaine d'une création de l'IA ne remplirait pas cette condition, de sorte que l'artiste-interprète humain se retrouverait sans protection. Pour éviter cela, l'étude suggère d'envisager une définition harmonisée de « l'interprétation » comme objet du droit des artistes-interprètes. Cette définition n'exigerait pas que l'objet de l'exécution soit protégé par le droit d'auteur (tout en veillant à ce que le champ d'application de la protection des artistes-interprètes ne soit pas étendu à des activités présentant un faible intérêt culturel).

Quatrièmement, même si les produits générés de manière autonome sont exclus de la protection du droit d'auteur, ils pourraient toutefois bénéficier d'une protection **au titre des droits voisins des producteurs de phonogrammes, des producteurs de films ou des éditeurs de presse** (si les produits peuvent être qualifiés respectivement de fixation sonore ou audiovisuelle ou de publication de presse), même si la production d'un produit d'IA ne présente que peu de rapport avec les activités traditionnelles des producteurs ou des éditeurs et ne nécessite pas nécessairement un investissement similaire. Afin d'éviter que certains de ces droits voisins ne soient utilisés pour contourner l'arbitrage sous-jacent du droit d'auteur, l'étude propose une condition expresse d'investissement, de sorte que seules les fixations pour lesquelles un certain investissement (éventuellement avec un seuil minimum) a été réalisé puissent déclencher l'application des droits voisins des producteurs de phonogrammes et de films ou des éditeurs.

Cinquièmement, dans les cas (encore rares) de produits générés de manière autonome par l'IA, **la fausse attribution de la paternité** à un humain pourrait, en pratique, permettre de contourner l'absence de protection du droit d'auteur pour ce type d'output. Il pourrait en effet suffire de revendiquer la paternité (en mentionnant une personne comme auteur) pour que celle-ci bénéficie (illicitement) de la présomption de paternité et donc de fait de la protection du droit d'auteur sur une création générée par l'IA, sachant que cette présomption est difficile, voire impossible à réfuter. Toutefois, le *statu quo* concernant la présomption de paternité devrait raisonnablement être maintenu. Une restriction ou une suppression de cette présomption serait en effet excessive et pourrait avoir des effets négatifs pour les créateurs humains à qui incomberait la charge de démontrer la paternité.

Enfin, l'étude analyse le fait que savoir qu'une œuvre d'art est créée par un humain ou par un système d'IA peut affecter la perception et l'expérience du public. L'étude aboutit néanmoins à la conclusion qu'aucune **obligation d'information** concernant l'utilisation d'une solution d'IA pour le développement d'une œuvre ne devrait être ajoutée à la réglementation en matière de droit d'auteur. Une telle obligation légale susciterait en effet des questions quant à son champ d'application et à son incidence sur la liberté artistique et les droits de la personnalité des créateurs. L'étude ne porte pas sur l'opportunité d'imposer une obligation d'information dans d'autres branches du droit, comme le droit de la consommation.

1. Introduction

This document is the final report for the project “Study on copyright and new technologies: copyright data management and artificial intelligence” which was carried out by a consortium of Technopolis Group, Philippe Rixhon Associates, the Catholic University of Louvain (UCLouvain), Crowell & Moring LLP and IMC University of Applied Sciences Krems.

The study addresses two broad topics related to the copyright system and new technologies. The questions are dealt with in two dedicated parts of the study.

The **first part** of the study systematically takes stock of the current situation with respect to rights metadata in different creative industries. It **identifies and describes current challenges and the potential economic consequences**. It also takes stock of the most important past and ongoing industry-led initiatives to address the identified problems, including initiatives based on artificial intelligence (AI) or blockchain. Finally, the study indicates ways that could contribute to improving the efficiency of copyright data management.

The main results of part 1 are featured in the main text of this report. Further in-depth analyses and technical details can be found in Annexes 1-6 to this report.

This part of the study was led by Philippe Rixhon and specific (empirical) parts were supported by Florian Berger (Technopolis Group) as well as Alfred Radauer (IMC University of Applied Sciences Krems). The team was advised by industry experts from the music, film and publishing sectors who acted as a sounding board for the project and consisted of Piero Attanasio, Chris Cooke and Sten Saluveer. Further support was also provided by Fabian Waiblinger and Dominik Beckers (Technopolis Group).

The **second part** of the study focuses on (i) **uses** of copyright-protected content as input to feed AI technologies and (ii) the *copyright implications* of the **production of cultural outputs by or with the assistance of AI**. Furthermore, the study discusses possible **policy options** that might be needed to react to these developments. This part of the study was performed by Alain Strowel (UCLouvain), Sari Depreeuw (Crowell & Moring LLP), Luc Desaunettes-Barbero (UCLouvain) and María del Carmen Calatrava Moreno (Technopolis Group).

2. First part: Copyright data management in European creative industries

2.1. Introduction

2.1.1. Context

Socio-technical developments – for example the data economy, the emergence of AI, and growing relevance of digital technologies like blockchain and business models such as interactive online platforms and B2B marketplaces – provide a unique opportunity for economic growth and societal well-being. At the same time, they bring **challenges**, but also **opportunities** to improve the protection of one of the core assets for innovation in the cultural and creative industries and the data economy: **intangible assets** and their manifestations in different products, for example audio-visual works and films, computer programmes or musical works.

In recent decades, there has been significant progress in creating a single European market for intellectual property (IP) in different areas – for example the 2019 Directive on copyright in the Digital Single Market (CDSM Directive). This will bring many benefits to the EU economy⁴. However, there is still room for improvements notably when it comes to the practical functioning of the market. One of the areas where this potential exists is the **data management related to copyright-protected content in European creative industries**.

Since the beginning of the digital revolution, the vision of a free flow of information across borders has been accompanied by the expectation that **digital rights management** would maximise the range of available artistic productions, minimise transaction costs, and pave the way for ubiquitous and differentiated licensing solutions, thus increasing monetisation and remuneration, and enabling the creative industries to flourish.⁵ This expectation has been partially met, for example thanks to many ongoing initiatives on copyright data and the interoperability of copyright data management systems (see section 2.4 and the corresponding annex). Digitisation offers many new opportunities for commercialising artistic productions. However, several data-related problems still prevent the creative industries from realising their full potential.

Recently, there have been important developments around copyright and metadata management in Europe and North America – driven by public actors and private initiatives. Examples include:

- Conceptual explorations at the Council of the European Union (EU) under the presidencies of Malta (January to June 2017), Estonia (July to December 2017) and

⁴ As indicated in the stocktaking document 15016/19 *Developing the Copyright Infrastructure* issued by the Council of the European Union on 20 December 2019.

⁵ Senftleben, M., Margoini, T., Antal, D., Bodo, B., van Gompel, S., Handke, C. et al., 'Ensuring the visibility and accessibility of European creative content on the world market: the need for copyright data improvement in the light of new technologies', *SSRN*, 2021.

Finland (July to December 2019) leading to first attempts to outline a copyright infrastructure,

- Launch of the Intellectual Property Action Plan by the European Commission,
- International developments around work identification and metadata standards such as the International Standard Content Code (ISCC), the Photo Metadata Standard or the compatibility between the International Standard Audiovisual Number (ISAN) and the Entertainment ID Registry (EIDR),
- European research projects such as the Copyright Hub (providing copyright information and access to simpler licensing), the Linked Content Coalition (standardisation of identifiers, metadata and messaging), ARDITO (publishing, audio-visuals, and images), and Music 2025 (United Kingdom),
- The report "Towards European Media Sovereignty, An Industrial Media Strategy to leverage Data, Algorithms and Artificial Intelligence", a report by Guillaume Klossa, special adviser to European Commission Vice-President Andrus Ansip⁶,
- Worldwide efforts to interconnect music metadata repositories including projects by Phonographic Performance Limited (PPL), Repertoire Data Exchange (RDx) project, International Copyright Enterprise (ICE) Services, Cube project, and the Society of Authors, Composers and Publishers of Music (Société des auteurs, compositeurs et éditeurs de musique / SACEM), URights project,
- Commercial developments such as Civil (journalism), The Creative Passport (music), PicaPro (photography) and the News Provenance Project (journalism),
- Launch of an attribution ledger initiative to form a publishing industry consortium to define rules and protocols required for verified attribution by Access Copyright (Canada),
- Passage of the Music Modernization Act (United States) to improve music licensing and royalties for streaming media services that led to the creation of the Mechanical Licensing Collective (MLC) and a new rights database.

The Intellectual Property Action Plan⁷ provides the main background document at EU level that anticipates this study. Through this action plan, the European Commission aims at helping Europe's creation and innovation industries remain world leaders and accelerate their digital transformation. In particular, it includes steps to improve the protection of intellectual property (IP), promote the use of IP by small and medium-sized enterprises (SMEs), facilitate IP sharing to increase technological diffusion in the industry, fight counterfeiting and better enforce IP rights, and promote a global level playing field.

The IP Action Plan aims to ensure that the **IP framework is fit for the digital age**. It refers explicitly to a **copyright infrastructure** defined as:

⁶ Available at https://ec.europa.eu/info/publications/towards-european-media-sovereignty_en, NOTE: ALL URLs AND HYPERLINKS REFERRED TO OR USED IN THE FIRST PART OF THIS REPORT HAVE BEEN ACCESSED AND CHECKED IN JUNE OR JULY 2021.

⁷ See *An intellectual property action plan to support the EU's recovery and resilience*, European Commission, COM(2020)760

"The **set** of rules, technologies and institutions that **frame** data management practices in the creative industries (e.g., improve authoritative and updated information on rightholders, terms and conditions and licensing opportunities)⁸".

In this study, we follow this broad definition of "copyright infrastructure".

When referring to "copyright", we include the rights related to copyright (including neighbouring rights, in particular those of performers and phonogram producers, as well as the new related rights in favour of press publishers), although the main focus here is on the data applied to copyrighted works and protected performances/fixations. As an alternative wording to designate this, we use the "management of rights data".

2.1.2. Scope and approach of the study

The study investigates the potential use of digital technologies, such as cloud-native applications, digital watermarking and fingerprinting, distributed ledgers or AI, to improve the management of data related to copyright-protected content by European creative industries.

Accordingly, the study:

- Provides a technical assessment of the identified problems related to copyright data management in creative industries,
- Maps and analyses ongoing industry-led initiatives addressing the identified problems,
- Quantifies the problems – in terms of both costs and missed opportunities, and their drivers and impacts,
- Indicates potential avenues to solve or reduce the identified problems.

The study takes a dual **geographical and sectoral focus**. Specifically, it covers developments in:

- The EU and its Member States, and
- Canada, the United Kingdom, and the United States,

where it considers:

- The film & TV industry,
- The music industry (publishing and recording, taking into account the different distribution channels of the music industry), and

⁸ See *An intellectual property action plan to support the EU's recovery and resilience*, European Commission, COM(2020)760, page 12.

- The publishing industry (books, press, journals, and images used therein⁹).

2.1.3. Methodological remarks

In terms of (empirical) methodology, this report is based on three main pillars: interviews conducted with industry experts (semi-structured interviews based on pre-defined interview guidelines); secondary research (identifying existing studies covering the research questions of the study); and online surveys in various creative sectors, with a broad outreach strategy towards associations, companies and individuals in the different sectors.

2.1.3.1. Interviews

We conducted **more than 80 interviews** covering three creative industries (film and TV, music, and publishing)¹⁰ in the 27 EU Member States, Canada, the United Kingdom, and the United States. The interviewees were stakeholders of the **digital** value network in these creative sectors – for example, authors, performers, rightholders, producers, publishers, management organisations, broadcasters, online platforms, IP offices, and technology providers. They were represented by individuals, companies, or trade bodies. The interviews were conducted under the Chatham House Rule¹¹ between January and May 2021, comprised up to three sessions each, and were often followed by the exchange of documents. The interviews were also instrumental to discuss ongoing initiatives (see section 2.4 and corresponding annex) and avenues for future action (see section 2.7).

2.1.3.2. Literature review and secondary research

For the literature review, we thoroughly analysed more than 20 core documents, including recent research and working papers, studies, position articles, and communications from the European Commission. In the selection process, we chose recent documents that covered different industries, technologies, and perspectives. For the structured evaluation, **we applied a scheme that differentiated between the identification of the problem** (e.g., interoperability of metadata), **its impact** (on the management, licensing, enforcement or remuneration of rights, and on innovation) **and the use of digital technologies** (blockchain, cloud-native applications, digital fingerprinting and watermarking, machine learning, and standardisation).

For the secondary research, our goal was to quantify the problems in more detail by collecting data points. In doing so, **we distinguished between the problems** (metadata quality, interoperability etc) **and the different domains** (rights management, licensing, enforcement, or remuneration).

⁹ For this study we have considered all images (photography, graphics, etc.) as part of publishing as their monetisation happens when they are “published” (not only in the press, but also in books, journals or online)

¹⁰ The creative industries also include sectors such as advertising, software and video games, which are outside the scope of this study part but to which many of its insights would apply.

¹¹ “When a meeting, or part thereof, is held under the Chatham House Rule, participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s), nor that of any other participant, may be revealed”. Chatham House, ‘Chatham House Rule’, 2021 (<https://www.chathamhouse.org/about-us/chatham-house-rule>).

2.1.3.3. Online surveys and primary research

In order to collect further data to support the quantification of metadata-related challenges, we designed, programmed and rolled out four online surveys for the music, publishing, film, and TV broadcasting sectors. The design of the survey questionnaires was based on insights from interviews and expertise of our study team as well as on our modelling of the causal relationships between (potential) challenges regarding rights metadata and (potential) resulting impacts (impact modelling).

In their final versions, the surveys consisted of two main modules: one general part, with selected, mostly open questions on the respondents' perspective on the topic (e.g., if respondents were aware of metadata challenges in their industry and – if so – what the impacts are), and one more detailed part, targeted at topical experts (e.g., estimating the extent of the lack of rights metadata and the related economic impacts).

The surveys were distributed through the main trade associations of the respective creative industries as well as across the networks of the study team. The target groups were the associations themselves and their members (the associations were asked to distribute the survey among their members).

The surveys ran from mid-April to the end of May 2021.

Owing to the high levels of complexity and intricacies of the topic, but also to the different levels of relevance of the topic in different industries, the number of respondents accessing (some of) the industry surveys differed greatly (between 7 for broadcasting and 124 for publishing). Furthermore, the non-completion of the surveys further reduced the number of responses to some questions (especially in the expert module) to – in the extreme case – only one response. The survey results are therefore rather indicative and taken only as one piece of evidence supplementing the insights from other empirical work (the interviews and literature review).

2.2. A technical overview of the relevant legal framework

Several legal sources or provisions, at international, EU or national level, refer to, or at least might have some implications, for the management of copyright data¹². In the following section, we **outline the regulatory framework** for copyright data management, **classify metadata**, analyse explicit and implicit **technical specifications of data management**, and mention **ongoing initiatives** recently proposed by the European Commission and the American Music Modernization Act, and discuss **technical challenges**.

The **sole purpose** of this technical analysis is to identify explicit or implicit requirements pertaining to data, metadata, and data or metadata management. It is by no means a legal analysis of copyright policies.

2.2.1. International framework

2.2.1.1. The formalities under the Berne Convention

As summarised by WIPO, “the prohibition of formalities for copyright protection is the result of a historical process. Before the 1886 Berne Convention, each country had its own rules for recognition of copyright in a work. Consequently, authors had to comply with formalities on a country-by-country basis. The Berne Convention introduced the principle that authors in Union countries need only comply with the formalities imposed by the country of origin of a work. This rule was replaced in the 1908 Berlin revision of the Convention by the current principle of formality-free protection, reflected in Article 5(2)¹³ of the present Paris Act 1971, according to which the enjoyment and the exercise of copyright shall not be subject to any formality”¹⁴.

Copyright exists and is protected (even) without formality. This means that copyright vests with the author independently from any declaration, registration, deposit of a copy or of the data regarding a work, etc. **Copyright metadata** (and their collection, disclosure, etc.) are not required to enjoy copyright protection; the **exercise of copyright** however in fact often requires that some data concerning the works and the rightholders are exchanged and confirmed.

2.2.1.2. WIPO treaties (1996)

The World Intellectual Property Organisation (WIPO) Copyright Treaty and the WIPO Performances and Phonograms Treaty¹⁵ define and protect a class of metadata called

¹² “Data: information, especially facts or numbers, collected to be examined and considered and used to help decision-making, or information in an electronic form that can be stored and used by a computer”, source: Cambridge Dictionary. In the digital world, every content is data. The file formats such as .jpeg, .mp3, .mp4, .pdf, .epub, all of them are data.

¹³ “The enjoyment and the exercise of these rights shall not be subject to any formality; such enjoyment and such exercise shall be independent of the existence of protection in the country of origin of the work. Consequently, apart from the provisions of this Convention, the extent of protection, as well as the means of redress afforded to the author to protect his rights, shall be governed exclusively by the laws of the country where protection is claimed” in *Paris Act relating to the Berne Convention for the Protection of Literary and Artistic Works* available at: <https://treaties.un.org/doc/Publication/UNTS/Volume%201161/volume-1161-I-18338-English.pdf>

¹⁴ From https://www.wipo.int/copyright/en/activities/copyright_registration/index.html

¹⁵ WIPO, WIPO Copyright Treaty (WCT), Geneva, 20 December 1996; WIPO, WIPO Performances and Phonograms Treaty (WPPT), Geneva, 20 December 1996.

Rights Management Information (RMI) defined as follows: “*information which identifies the work, the author of the work, the owner of any right in the work, or information about the terms and conditions of use of the work, and any numbers or codes that represent such information, when any of these items of information is attached to a copy of a work or appears in connection with the communication of a work to the public*” (Art. 12 (2) WIPO Copyright Treaty; comp with Art. 19 (2) WIPO Performances and Phonograms Treaty).

Countries which ratify the WIPO Treaties must provide “adequate and effective” legal protection against those who without authorisation remove or alter, or are otherwise trafficking with, the electronic RMI. There is however no positive obligation in the WIPO Treaties to provide a sound system regarding RMI and rights data management.

2.2.2. Acquis communautaire

2.2.2.1. Definition of Rights Management Information (RMI)

The European Directive 2001/29/EC¹⁶ on copyright in the information society expands the WIPO definition: “*Rights Management Information means any information provided by rightholders which identifies the work or other subject matter [], the author or any other rightholder, or information about the terms and conditions of use of the work or other subject-matter, and any numbers or codes that represent such information. [This] shall apply when any of these items of information is associated with a copy of, or appears in connection with the communication to the public of, a work or other subject matter []*”.

For this report and the purpose of data management, the term of **rights management information (RMI)** will continue to cover two distinct types of metadata:

- i. **Identification metadata.** These metadata make up the smallest set of data that uniquely defines an identified entity – works, related subject matters, and stakeholders¹⁷. Although different sets of data can be identified, there is usually community consensus on the appropriate set to use in particular circumstances.
- ii. **Rights metadata.** This metadata or attribution metadata are the information relating to the authorship and ownership, information about the terms and conditions of use and reuse of a protected subject matter. It links the answers to *who did what, who owns what, and what we can do with that*.

Copyright data management handles more metadata than the two types included under the RMI definition:

- iii. **Descriptive metadata.** These metadata or rich metadata or discovery metadata, can include information that is helpful in searching for a particular entity (the genre of a novel or musical composition), or aiding its enjoyment or understanding (the influences on the creator or group of creators) or for general interest (the location of the first performance of a concerto),

¹⁶ Article 7(2) of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, OJ L 167, 22.6.2001, p. 10.

¹⁷ “Stakeholder: a person such as an employee, customer, or citizen who is involved with an organisation, society, etc. and therefore has responsibilities towards it and an interest in its success, source: Cambridge Dictionary. For this study, “stakeholders” include creators, authors, rightholders, and any natural or legal person who must be **identified** for the purpose of rights management.

- iv. **Usage metadata.** This is a category of metadata that is becoming more and more important because of the emergence of new business models, for example advertisement- or subscription-based digital platforms, mashups and prosumerism. This category comprises information on where, when, and how frequently content has been used, and how it has been monetized: it is essential to monitor usage of copyright-protected content and remuneration for its use.
- v. **Administrative metadata:** The metadata in all the previous categories have provenance – who asserted the information and when. This fifth category of metadata, often called administrative metadata, is needed to assess the reliability, value, and trustworthiness of the other metadata.

Rights management information is defined in *acquis communautaire*

- **Identification metadata** to answer the questions:
what is what, who is who
- **Rights metadata** to answer the questions:
who did what, who owns what, what can we do with that

Additional metadata is defined by *industry practices*

- **Descriptive metadata** for search and enjoyment of content
- **Usage metadata** for a fair trade of content
- **Administrative metadata** to trust all other metadata

Table 1: Categories of metadata

Source: Philippe Rixhon Associates

Copyright data management handles RMI, indexing metadata necessary to incentivise rights declarations and rights licences, administrative metadata necessary to trust data, and usage metadata necessary to maintain the transparency of uses and remuneration.

As under the WIPO Treaties, RMI under the 2001/29 Directive are only protected against their removal or alteration or additional acts of trafficking in RMIs (Art. 7 § 1)¹⁸.

The **protection of RMI** is however difficult to separate from the protection of (effective) technological measures (under Art. 6)¹⁹, often designated (in practice and literature) as Technological Protection Measures (TPMs), such as Digital Rights Management Systems (DRMs) including password solutions, encryptions, digital rights management tools (DRMTs), or bindings that attach work identifiers, rightholder identities and RMI. Those technological measures are only protected insofar they are “effective” (meaning they can achieve the protection objective (see Art. 6 §3)).

¹⁸ Article 7(2) of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, OJ L 167, 22.6.2001, p. 10.

¹⁹ Article 6(1) of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, OJ L 167, 22.6.2001, p. 10.

As already mentioned, although there is no positive obligation under the *acquis communautaire* to use TPMs/DRMs, the legal framework provides legal protection to TPMs when attached and helps fight new forms of piracy (involving RMI or TPMs/DRMs) and directly encourages the adoption of common standards concerning rights data management²⁰ (or technological protection measures).

2.2.2.2. Some examples of application of Rights Management Information coming from the EU copyright acquis

As other IP rights, copyright and the related rights balance different societal interests. The harmonisation of copyright brought about by the *acquis communautaire* applies to material interests, but not to moral rights²¹. Some moral rights protect the attribution of authorship or performance (the right to be recognised as creator and credited for that) and the integrity of the protected works or subject matter (as alterations or other modifications might impact the reputation of the creator). The right management system also can contribute to the objectives pursued by the legal protection of the moral interests of creators. We will come back to this consideration when discussing pain points and options.

Technically speaking, the protection of copyright-protected works, possibly under licensing contracts that can segment and distribute those rights, require the **availability** of metadata related to the work or related subject matter and these licensing terms and conditions. This requires the inclusion of information on the rightholders and authorised uses in RMI. This information should be actionable, allowing prospective licensors and licencees first to finalise and then to execute a licensing agreement. The type of rights availability defined in the RMI could refer or link to detailed terms and conditions stored elsewhere.

On a separate topic, more recently, the CDSM Directive imposes that the remuneration for authors and performers must be **appropriate, proportionate, and transparent**²². The application of this principle requires the provision of usage data along the licensing chains, which can be many, e.g., when a writer assigns the right of publication to a publisher limited to one language and negotiate translation rights or the rights for film adaptation etc. to other parties. This is also stated in Article 19 of the Directive on CDSM:

"Member States shall ensure that authors and performers receive on a regular basis, at least once a year, and taking into account the specificities of each sector, up to date, relevant and comprehensive information on the exploitation of their works and performances from the parties to whom they have licensed or transferred their rights, or their successors in title, in particular as regards modes of exploitation, all revenues generated and remuneration due".

Besides the above, the CDSM Directive provides four cases where the need for right information management (in a broad sense) is clear:

²⁰ See Recital 54 of the Directive EC/2001/29: "Compatibility and interoperability of the different systems should be encouraged. It would be highly desirable to encourage the development of global systems", and the Directive EU/2014/26 that encourages voluntary standards.

²¹ Recital 21 of Council Directive 93/98/EEC of October 1993 harmonizing the term of protection of copyright and certain related rights, OJ L 290, 24.11.1993, p. 9.

²² Articles 18–20 of Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the digital single market, and amending Directives 96/9/EC and 2001/29/EC, OJ L 130, 17.5.2019, p. 92.

- Art. 4, on text and data mining (TDM), introduces an exception for purposes different than research, where rightholders may reserve the rights. To do so, they must communicate this information to the miners, including through “machine readable means” for online content,
- According to Art. 5, § 2, Member States may provide that the exception for illustration for teaching does not apply when a licence for the same use is available. In this case, information on the licences must be accessible, which means that RMI must be available and easy to reach by the educational establishments,
- Art. 8-11 provide for a specific data infrastructure to manage the new licensing mechanism for out of commerce works: rightholders and Cultural Heritage Institutions need to speak to each other, through the EUIPO portal, to manage possible opt-outs,
- Art. 17: Licensing or asking for the prevention of unauthorised content by online content-sharing service providers requires rightholders to provide them with relevant and necessary information and in case of take down, a “sufficiently substantiated notice”.

Finally, the panoply of neighbouring or related rights recognised in the acquis communautaire has recently been completed by the new related right for press publishers (Art. 15 of the CDSM Directive). The application of this provision may give rise to exchanges of metadata between press publishers and information society service providers (e.g., rights metadata needed by the press publishers to license in an affordable manner significant excerpts of their articles, and usage metadata gathered by the information society service providers).

Metadata requirements are mostly implicit in the preceding paragraphs. They can also be explicit as in the following example. In the case of music, the Directive EU/2014/26 on collective management of copyright and related rights and multi-territorial licensing of rights in musical works for online use in the internal market imposes several obligations, in particular on the Collective Management Organisations (CMOs), that require some processing of rights information, and even provide principles of data governance by establishing data responsibilities on CMOs and users to ensure accuracy, comprehensiveness, and interoperability of the data across the value chain:

- Article 5. CMOs must exchange electronically **lists of rightholders, their authorised rights and withdrawals of rights** with their rightholders and users.
- Article 11. CMOs must **collect rights revenue** and accordingly exchange data with their users.
- Article 13. CMOs must distribute revenues accurately and in a timely manner based on usage reporting, i.e., on usage metadata as per section 2.2.1.1.v. Then, they must also **identify and locate the respective rightholders**.
- Article 16. CMOs and users must provide each other with **all the RMI necessary for licensing of rights**.
- Article 17. Users must provide a CMO, in an agreed format, with the **relevant RMI** at their disposal on the use of the rights represented by the CMO as is necessary for the **collection of rights revenues and for the distribution and payment** of amounts due to the rightholders. When agreeing on the format, CMOs and users must consider voluntary **industry standards**.

- Article 18. CMOs must share with their members/rightholders (i) **contact details** to identify and locate the rightholder, (ii) the **rights revenue** attributed to the rightholder, (iii) the **amounts** paid by the CMO to the rightholder **per category** of rights managed and **per type of use**, (iv) the **period** during which the use took place, (v) **deductions** of management fees or other fees required by national law, and (vi) any rights revenue that is **outstanding** for any period.
- **Article 24.** CMOs must have the ability to **identify** accurately (i) the musical **works**, wholly or in part, and (ii) wholly or in part, with respect to each relevant territory, the **rights** and their corresponding **rightholders** for each musical work or share therein. They must also **make use** of (i) **unique identifiers** in order to identify rightholders and musical works, taking into account, as far as possible, voluntary industry standards and practices developed at international or EU level, and (ii) adequate **means in order to identify and resolve** in a timely and effective manner **inconsistencies** in data held by other CMOs granting multiterritorial licences for online rights in musical works. This article explicitly requires CMOs to assure the functions of **licence management**, **unique granular identification** of works, assets and rightholders, **consumption monitoring**, invoicing, collecting and distributing revenue, and timely and effective **cross-border metadata management**. It implicitly requires reconciling RMI and matching RMI with consumption.
- **Article 25.** CMOs must **share** the abovementioned RMI with rightholders, online service providers and other CMOs.
- **Article 26.** At the request of rightholders, online service providers and other CMOs, CMOs must **correct erroneous RMI** without delay. They must also provide their members/rightholders with the means of **submitting their RMI in electronic form**.
- **Article 27.** CMOs must **monitor the actual use** of online rights in musical works that they represent and offer online service providers the possibility of reporting the actual use by electronic means. Then, online service providers must accurately report the actual use. Once they have done this, CMOs must invoice the online service providers by electronic means.

Exemplary technical challenges

This section highlights the technical challenges related to the implementation of two provisions of the *acquis communautaire* on rights metadata. We seized the opportunity to study their transpositions in two Member States. The texts selected for this highlight are gathered in Annex 5.1 for the reader's convenience. We looked at the English, French and German formulations of the articles of the *acquis communautaire* and then at their transpositions into French and German laws.

Copyright in the information society (EC/2001/29) – Article 7 §2²³

²³ Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, O.J. L 167, 22 June 2001, p. 10–19

Article 7 §2 of the Directive EC/2001/29 refers to obligations **concerning rights-management information** as follows:

*Rights Management Information means any information provided by rightholders which identifies the work or **other subject matter** [], **the author or any other rightholder**, or information about the terms and conditions of use of the work or other subject-matter, and any numbers or codes that represent such information. [This] shall apply when any of these items of information is associated with a copy of, or appears in connection with the communication to the public of, a work or other subject matter [].*

The European Directive speaks about the identification of **the author or any other rightholder**, the French legislation about the identification of **a rightholder**. Under copyright law, the term *rightholder* usually covers the *author* (initial rightholder) and *rightholder* (which could hold the rights by way of contract, inheritance, or deed). For the management of data, to keep track of the various ways the rights might be transferred to another holder generates a lot of complications: data managers will have to know if they should identify only *who owns what* or identify *who owns what* and *who did what*, and need to qualify the ownership: i.e. the kind of relation between the person (rightholder) and the work, e.g. if the person is an author, translator, illustrator, etc.²⁴

The management of RMI is also complex because of the need to provide granular information. For example, a news publication on which a press publisher enjoys the new related right (provided by the CDSM Directive) might contain several components, such as a picture, a table or another graphic illustration, and the rights on those granular elements do not necessarily belong to the same rightholders, so that the rights data management requires additional layers of information to take into account all the rights involved with a subject matter that comprises other protected items.

This very brief analysis highlights a few challenges of data management. **Metadata requirements** are **explicitly or implicitly, similarly or differently**, formulated in European directives and legislations of the Member States. The high level of harmonisation in the Union, as far as the customary management of copyright is concerned, does not reduce the complexity of metadata management, which depends on the diversity of copyright policy of individuals – creators, publishers, producers, performers, etc. – and is a value since it stimulates competition and innovation.

Dealing with this **complexity** requires legal expertise as well as **sophisticated IT systems** which can be out of reach or too expensive for most European small and medium creative enterprises in absence of a standard language to express individual rights policies in an easily formulated and understood way by SMEs (or to be more precise, by SMEs' machines).

²⁴ For example, the ONIX code list 17 (<http://www.onix-codelists.io/codelist/17>) counts 114 items for contributors.

2.2.3. Preliminary conclusions

Directive 2001/29/EC and Directive 2014/26/EU together provide a solid foundation to copyright-related data governance.

Acknowledging the emerging role of the internet, the 2001 Directive provided a common European definition of what metadata is and established legal protection of technological measures, when attached to the content that is being made available or distributed on the networks.

Later, in 2014, the Directive on collective rights management established sound principles of data governance, and relying on the capacity of CMOs, in particular in the area of music and multi-territorial licensing, and establishing the ground for electronic exchanges between rightholders, CMOS, and users. In parallel of these legal developments, the industry also worked on strengthening data governance with initiatives such as CISAC, DDEX and many others (see section 2.4).

We further analyse the functional and metadata requirements of the regulatory framework in section 2.2, but it appears that the *acquis communautaire* would explicitly or implicitly involve in practice to –

- **Identify** works (and other subject matters), rights and stakeholders (creators, rightholders, and actors on the content value network²⁵),
- **Link** works and rights, works and stakeholders, and rights and stakeholders, i.e., manage the relationships between the three types of entity - works, rights, and stakeholders,
- **Monitor** usages of works and remunerations for uses. In other words, monitor the flow of content from creators to consumers and the counterflow of remuneration from consumers to creators.
- **Fulfil** human- and machine-readable contracts to handle licences, distribution channels, and flows of royalty. *Human-readable* contracts are for ensuring the transparency and enforceability of agreements, and *machine-readable contracts are for coping with* the exponential volume of transactions.
- **Protect** personal data and privacy, as well as business confidentiality.

on the whole complex, dynamic, fluid, transitional and liminal content value network.

²⁵ The creative ecosystem must not only identify rightholders but also any stakeholder who must be ascertained to manage regulatory exceptions or limitations. See also section 2.7.1.1 on the prevalence of the content value network.

2.3. Literature review on metadata issues and digital technologies

2.3.1. Issues related to copyright metadata

Intangible assets play a crucial role in the data economy of the future. In the EU, the volume of annual investment in *intellectual property products* has increased by 87% in the past 20 years, whereas the volume of tangible investments has increased by only 30%.²⁶ Therefore, a functioning IP protection is crucial to achieving the EU's ambition "to acquire a leading role in the data economy"²⁷ as stated in the European strategy for data by the European Commission.

Although digital technology offers unprecedented opportunities for commercializing and publishing IP products – among them outputs from the cultural and creative industries – several challenges still prevent the realization of its full potential. Among these challenges are problems connected to metadata of creative works.²⁸ The following sections give an overview of the main insights from an **overview on the current literature** (studies, policy papers etc) on this topic. To scope the literature review and ensure the feasibility of a literature review in the context of this study, the study team focused primarily on the review of a pre-defined list of publications, including scientific research papers, policy papers and other content. The selected publications are listed in the Annex of this report.

2.3.1.1. Cases of imperfect metadata attached to content

Typically, metadata relate to the ownership of a creation and/or the identity of parties involved in the creation or entitled to trade with it.²⁹ They are crucial for functioning copyright data management.³⁰ However, not all creators use metadata systematically right from the start of the creative process or are completely aware of their importance and benefits. Specifically for the music industry, this is a recurring theme³¹. Although the importance of metadata is recognised among "established" players in the market, this is particularly a problem among artists and songwriters, and especially self-releasing, self-publishing artists and songwriters without formal music industry business partners. This can in its extreme lead to cases of at least elements of metadata being absent from content, despite large efforts to avoid this. Other studies also point this out for areas such as photography³², whereby 96% of images in news services lack metadata, and cultural heritage³³. For other sectors, in the literature reviewed for this study similar findings were not identified.

Some studies also suggest that metadata continues to be removed from published content – for various reasons.³⁴ This intentional removal of metadata on platforms is called *data-stripping*. For example, some digital platforms remove metadata, arguably to save data

²⁶ European Commission, 25 November 2020, "Making the most of the EU's innovative potential".

²⁷ European Commission, 19 February 2020, "A European Strategy for Data"

²⁸ Martin Senftleben et al., 12 February 2021, "Ensuring the Visibility and Accessibility of European Creative Content on the World Market"

²⁹ Paul Jessop, "Functional Requirements for the International Music Registry Analysis"

³⁰ Council of the European Union, 20 December 2019, "Developing the Copyright Infrastructure"

³¹ This observation has led to specific awareness raising campaigns in the music industry (#metadatapays) concerning the use of metadata see [here](#).

³² Council of the European Union, 20 December 2019, "Developing the Copyright Infrastructure" citing a study by Imatag available at: <https://blog.imatag.com/state-of-image-metadata-in-news-sites-2019-update>

³³ <https://pro.europeana.eu/post/62-of-rights-statements-are-accurate-together-we-can-improve-on-that>

³⁴ See for example: <https://www.cmswire.com/digital-asset-management/embedded-metadata-debate-to-strip-or-not-to-strip/>

capacity or to speed up loading of webpages with image content. While a larger file size because of metadata is insignificant for longer videos (comparing the file size with and without metadata), it is not negligible for photos. However, increasingly faster Internet connections should no longer justify this reason for removing metadata.³⁵

In this context, it can be mentioned that – although following the Berne convention there is no obligation of registration in order to be entitled to a copyright– some scholars see registration systems as a possible method for facilitating the exercise of copyright by providing right owners with a simple and effective means to clearly establish authorship or ownership of rights.³⁶ Some countries have established voluntary national registration systems or other formalities³⁷. These systems could also help to address the issue of metadata quality.^{38,39}

While there is no clear and widely accepted evidence in the literature yet on how large the dimension of the problem is (i.e. how many creative works lack relevant metadata), it is undebated that missing or erroneous metadata can lead to negative impacts as works and rightholders cannot be properly identified (see also below).⁴⁰

2.3.1.2. Interoperability of metadata

In addition to the problem of lack of metadata, interoperability between data management systems and related data libraries is according to existing studies not always ensured.⁴¹ Despite ongoing efforts in different sectors on standardising metadata, existing databases of rights still seem to often lack common technical standards and formats used are not completely interoperable.

This problem has been covered in debates and studies as early as 2001 when recital 54 of Directive 2001/29/EC announced that compatibility and interoperability of the different technical systems of identification of works in digital format should be encouraged.

Since then, certainly much has developed in terms of interoperability of metadata (for example through standardisation efforts). However, some publications still report on challenges in this regard⁴². This seems to arise from different sets of metadata and (despite existing standards) different approaches to data identification and verification.⁴³ Distributors or users are responsible for the formats used for metadata submission and exchange. Authors can rarely change them. The format is therefore an important element for the interoperability of metadata for identifying works and their rights holders.⁴⁴ A further reason for remaining challenges seems to be the work flow in the supply chain ("often different parts or "fractions" of metadata are kept at different locations by different

³⁵ French Ministry of Culture, 29 January 2020, "Towards more effectiveness of copyright law on online content sharing platforms"

³⁶ World Intellectual Property Organization (WIPO)

³⁷ Note: copyright registration is necessary in the USA to get full protection under the law, e.g., to get statutory damages, so domestic creators and rightholders routinely register their works with the Copyright Office. It does not seem to be done systematically by DIY artists/songwriters, and it does not appear to help much with getting publishers/songwriters paid.

³⁸ World Intellectual Property Organization (WIPO)

³⁹ This might be specifically relevant when authors are not cooperating with CMOs; when authors go to CMOs, their metadata is there.

⁴⁰ Council of the European Union; Martin Senftleben et al.

⁴¹ Martin Senftleben et al.

⁴² Intellectual Property Office and Ulster University, 18 June 2019, "Music 2025 Summary Paper"

⁴³ Intellectual Property Office and Ulster University, 18 June 2019, "Music 2025 Summary Paper"

⁴⁴ Council of the European Union

entities along the music supply chain⁴⁵) There, however, are important best practices and standards for the interoperability of this data within sectors, for example DDEX in the music sector⁴⁶, and ONIX in the publishing industry.

During the IPTC Photo Metadata Conference 2020⁴⁷, Martin Seidl highlighted a need of (1) a guide for synchronising metadata across standards⁴⁸, (2) to encourage software builders, publishers and producers to adopt metadata standards.

There have been attempts to solve the problem of interoperability through identification schemes and the adoption of common standards within one sector. However, identification metadata, sets of rights metadata, and exchanges of metadata differ from one industry to the next, which has negative consequences for the cross-sector interoperability.⁴⁹ In addition, there is currently no complete list of identifiers, as sources list only those identifiers that are relevant to their own domain.⁵⁰

2.3.1.3. Authoritative sources⁵¹

According to the reviewed literature, a general challenge of copyright data management in creative industries is the fact that databases containing information on rights and licensing terms and conditions can be – to different degrees – fragmented (several CMOs, platforms, etc.) or not publicly accessible. In the literature analysed for this study, this is discussed specifically for the music industry: co-authorship and copyright splits are common and may lead each co-author to declare their rights to different CMOs. Rights metadata are then difficult to reconcile⁵². Market participants may therefore not always have access to **reliable authoritative data** to obtain licences for the use of content (the term “authoritative data” refers in this report to information for which correctness is widely acknowledged and not disputed by any actor/market participant). To manage copyright efficiently and securely the verification of metadata against an authoritative source is crucial⁵³.

There are manifold activities which are trying to address the challenge of authoritative metadata – not only in the music industry, but different sectors. In section 2.4 of this report we cover these activities in detail. At this point, it suffices to mention that ongoing initiatives are tackling the issue on authoritative rights metadata: CISAC led a project to harmonise music cue sheets, an initiative that will bring significant benefits to everyone involved in music productions for audio-visual works. The project has been implemented in a collaboration between CISAC, representing authors’ societies worldwide, and music publishers and producers through the Society Publisher Forum⁵⁴. Along similar lines, CISAC

⁴⁵ Hardjono, T., Howard, G., Scace, E., Chowdury, M., Novak, L., Gaudet, M. et al., ‘Towards an open and scalable music metadata layer’, *arXiv*, 2019.

⁴⁶ French Ministry of Culture

⁴⁷ M. Steidl, *About IPTC Photo Metadata*, available at: <https://www.youtube.com/watch?v=X8-mYihITZM>

⁴⁸ Synchronise between IIM, XMP, EXIF, whereby IIM and XMP may contain licensing/copyright metadata

⁴⁹ Intellectual Property Office and Ulster University

⁵⁰ Council of the European Union

⁵¹ This section covers mainly findings from the music industry. The reason for this is that in the literature reviewed for this report, challenges were mainly discussed for this context. For other sectors the available literature base seems to be much more scarce.

⁵² Intellectual Property Office and Ulster University. This is mainly a problem with song rights (where co-ownership is very common) rather than recording rights (where co-ownership is rarer). Another problem is that artists and labels log recordings with the record industry’s CMO databases, and songwriters and publishers with the music publishing sector’s CMO databases, but recordings contain songs, and the matching generally happens after usage.

⁵³ Council of the European Union

⁵⁴ See <https://www.cisac.org/Newsroom/articles/cisac-and-publishers-come-together-launch-harmonised-music-cue-sheets>

has launched in September 2020 a major upgrade of the ISWC system to increase its efficiency, speed and accuracy. The underlying motivation here is, again, to improve the quality and therefore implicitly also the “authoritativeness” of the ISWC system.⁵⁵ In this sense, CISAC products are highly important to market participants and can therefore be regarded to a certain degree as authoritative. However, as CISAC itself acknowledges, data inaccuracies persist in their databases⁵⁶. An important problem is that, although CMOs have very comprehensive databases, they are often not exhaustive: not all (independent) rights owners are registered or not all works and related subject matter are declared. Additionally, as mentioned earlier, it is possible that different registrations of one and the same work exist at different CMOs. In these cases, the CMO must then reconcile these different registrations when analysing the usage data from the user reports.⁵⁷

Alternatively, in some cases payments may stay non-distributed due to lack of registration.⁵⁸ This lack of reliable authoritative data may have a negative impact on rightholders who may miss revenue streams linked to their unidentified content (see section 2.4 of this report).⁵⁹

2.3.2. Impacts of the metadata issues

The abovementioned challenges lead to different impact dimensions discussed in the literature. The following sections briefly cover the most prominent impact dimensions discussed in the literature.

2.3.2.1. Impacts on the management of rights

The identified problem discussed above may cause stakeholders to deal with inaccurate, missing, non-interoperable and fragmented metadata. In combination with the rapid growth of data exchange volume, this has resulted in high levels of complexity in the management of rights⁶⁰. Incomplete metadata require time-consuming manual reconciliation, for example through cue sheets in the music industry. Non-interoperable metadata and therefore fragmented datasets lead to duplication of solutions of the same type developed by individual organizations. Ultimately, this results in higher costs and inefficient management of rights⁶¹. The literature lacks a precise indication of how large the problem is (i.e. efficiency losses due to complex rights management).

⁵⁵ CISAC was reported to state that “the new system will end the current practice of having ISWC codes allocated by individual societies, a practice that leads to many data integrity challenges”. This confirms that authoritativeness of the ISWC system is – although widely accepted as an industry standard – still a problem. See <http://legrandnetwork.blogspot.com/2020/09/cisacs-iswc-system-gets-major-overhaul.html>

⁵⁶ See <http://legrandnetwork.blogspot.com/2020/09/cisacs-iswc-system-gets-major-overhaul.html>

⁵⁷ European Commission, *Remuneration of authors and performers for the use of their works and the fixations of their performances*, European Commission, Brussels, 2015.

⁵⁸ Council of the European Union, ‘Developing the copyright infrastructure – stocktaking of work and progress under the Finnish Presidency’, Brussels, 20 December 2019 (<https://data.consilium.europa.eu/doc/document/ST-15016-2019-INIT/en/pdf>).

⁵⁹ Council of the European Union, ‘Developing the copyright infrastructure – stocktaking of work and progress under the Finnish Presidency’, Brussels, 20 December 2019 (<https://data.consilium.europa.eu/doc/document/ST-15016-2019-INIT/en/pdf>).

⁶⁰ Intellectual Property Office and Ulster University, ‘Music 2025 summary paper’, 18 June 2019.

⁶¹ Senftleben, M., Margoini, T., Antal, D., Bodo, B., van Gompel, S., Handke, C. et al., ‘Ensuring the visibility and accessibility of European creative content on the world market: the need for copyright data improvement in the light of new technologies’, *SSRN*, 2021.

2.3.2.2. Impacts on the licensing of rights

Incorrect metadata may result in payment of licensing royalties for incorrect creations and/or payments to the wrong party. This may ultimately lead to the use of a creation whose exploitation was not authorized. Some errors in metadata are simple mistakes that can usually be corrected by clerical or administrative work. In other cases, ambiguities may arise between two parties with similar names⁶². In any case, the implications of insufficient metadata making licensing transactions inefficient is a regularly discussed topic in the literature⁶³.

Besides negative impacts on the disbursement of royalties, inaccurate or incomplete metadata make the content also hard to find and therefore difficult to license. This might increase the search costs for users to such an extent that users are incentivised to make unauthorized use of content. Therefore, incomplete metadata can contribute to digital piracy⁶⁴. Alternatively, search costs may lead to content not being used. Since search costs will often be higher for less known works or artists, this may lead ultimately to an unlevelled playing field between more and less prominent market participants and possibly to impacts on cultural diversity and innovation.

Another challenge is the fragmentation of copyright. Sometimes the copyright of a creation is shared by co-authors and divided contractually by territory, language, means of distribution, etc. This means that, for a single use of copyrighted work, a user might need several authorisations⁶⁵.

2.3.3. Use of new technologies

Throughout copyright history, technological change has enabled new ways of enforcing rights and new ways of using works. There are past and ongoing initiatives taken by the creative industries to use digital technologies to address issues of copyright data management⁶⁶.

2.3.3.1. Blockchain

Blockchain technology can be defined as a distributed, append-only database that - without a central trusted intermediary - enables transactions between human or software agents⁶⁷. In the field of copyright, metadata could be represented by cryptographic tokens⁶⁸. The idea is that, based on the registers of such tokenised elements, smart contracts can automate and standardise a variety of copyright-related transactions.

⁶² Jessop, P., 'Functional requirements for the International Music Registry analysis'.

⁶³ Council of the European Union, 'Developing the copyright infrastructure – stocktaking of work and progress under the Finnish Presidency', Brussels, 20 December 2019 (<https://data.consilium.europa.eu/doc/document/ST-15016-2019-INIT/en/pdf>); Jessop, P., 'Functional requirements for the International Music Registry analysis'.

⁶⁴ Senftleben, M., Margoini, T., Antal, D., Bodo, B., van Gompel, S., Handke, C. et al., 'Ensuring the visibility and accessibility of European creative content on the world market: the need for copyright data improvement in the light of new technologies', *SSRN*, 2021.

⁶⁵ European Commission, *Remuneration of authors and performers for the use of their works and the fixations of their performances*, European Commission, Brussels, 2015.

⁶⁶ Rendas, T., 'Copyright, technology and the CJEU: an empirical study', *International Review of Intellectual Property and Competition Law*, Vol. 49, No 2, 2018.

⁶⁷ Bacon, J., Michels, J. D., Millard, C. and Singh, J., 'Blockchain demystified', *Queen Mary School of Law Legal Studies Research Paper Series*, No 268/2017, 2017.

⁶⁸ "Cryptographic tokens represent programmable assets or access rights, managed by a smart contract and an underlying distributed ledger. They are accessible only by the person who has the private key for that address and can only be signed using this private key", from <https://blockchainhub.net/tokens/>.

The use of blockchain in copyright data management is often associated with the hope of facilitating the protection of copyrighted works, improving transparency, enabling smoother distribution of royalties and combating piracy more effectively⁶⁹. In addition, blockchain can be used to find orphan works⁷⁰. However, the use of blockchain also creates difficulties and entails structural incompatibilities. The highly fragmented nature of metadata conflicts with the impersonal, borderless, standardized, and automated regulatory solution that blockchain technology offers⁷¹.

For example, there is no legal consensus on how the blockchain fits into the traditional concepts of contract law. It is unclear how dispute resolution will take place, as the parties to a blockchain-based smart contract are anonymous. In the foreseeable future, this uncertainty about the legal status of smart contracts is likely to limit the emergence of copyright licences in blockchains⁷².

Another challenge for copyright is that blockchain transactions are immutable. Therefore, problems will arise in the event of misidentified artists, contract changes and dispute resolution outcomes. Thus, the blockchain, with its irreversible chain of information transfers, can lead to problems that do not occur with a conventional data storage technology⁷³.

In an initiative of the music industry, a blockchain-based solution was being built to link and manage two identification standards of copyright-protected content used for music recordings. The idea of the initiative is to use blockchain as a technological architecture to create and operate a shared database of rights management information⁷⁴.

The potential of blockchain for copyright data management is unclear and may vary among industries⁷⁵.

2.3.3.2. Artificial intelligence

Machine Learning, a subset of AI, can be used to improve metadata and detect piracy. After being “taught” millions of references, AI is able to identify creations and add or complement metadata very effectively⁷⁶. One downside of the technology is that AI is not effective at learning from small samples. For this reason, very old or cultural minority works are likely to be disadvantaged. In addition, AI requires access to extensive computing resources.

⁶⁹ European Commission

⁷⁰ Bodó, B., Gervais D. and Quintais, J. P., ‘Blockchain and smart contracts: the missing link in copyright licensing?’, *International Journal of Law and Information Technology*, Vol. 26, No 4, 2020, pp. 311–336.

⁷¹ Bodó, B., Gervais D. and Quintais, J. P., ‘Blockchain and smart contracts: the missing link in copyright licensing?’, *International Journal of Law and Information Technology*, Vol. 26, No 4, 2020, pp. 311–336.

⁷² Bodó, B., Gervais D. and Quintais, J. P., ‘Blockchain and smart contracts: the missing link in copyright licensing?’, *International Journal of Law and Information Technology*, Vol. 26, No 4, 2020, pp. 311–336.

⁷³ Intellectual Property Office and Ulster University, 18 June 2019 “Music 2025”

⁷⁴ Senftleben, M., Margoini, T., Antal, D., Bodo, B., van Gompel, S., Handke, C. et al., ‘Ensuring the visibility and accessibility of European creative content on the world market: the need for copyright data improvement in the light of new technologies’, *SSRN*, 2021.

⁷⁵ Intellectual Property Office and Ulster University, ‘Music 2025 summary paper’, 18 June 2019.

⁷⁶ French Ministry of Culture, *Mission Report – Towards more effectiveness of copyright law on online content sharing platforms: Overview of content recognition tools and possible ways forward*, French Ministry of Culture, Paris, 2020.

Different from other technologies, the results produced by AI are not always entirely reproducible and controllable. This means that sometimes unexpected errors can arise⁷⁷.

Nevertheless, AI offers opportunities to monetize creative content in the European Union. However, this will be possible only if the AI has access to metadata in a harmonized, accurate and interoperable format⁷⁸. Therefore, AI alone cannot just help to improve copyright data management; accurate and up-to-date metadata are also necessary to develop competitive AI systems in the EU.

2.3.3.3. Standardisation

EU content markets are benefitting from various common standards for rightholders to declare their rights to creations in an authorised, machine-readable and accessible form⁷⁹. Standardisation is key to ensure interoperability⁸⁰. It provides shared cost reductions between partners throughout the digital supply chain, improve the efficiency of licensing and counter duplication of work by homogenising the required data feeds. In addition, standardisation of metadata facilitates the use of new technology and allow automated processes for distribution of revenue streams⁸¹.

The European Directives encourage the use of voluntary standards. In particular, the CRM Directive 2014/26/EC encourages the use of standards across the value chain: for identification (Article 24(2)), for the exchange of information (Article 26), for usage reporting (Articles 17 and 27(2)), and for invoicing (Article 27(3)).

On the industry side, there are manifold standardisation activities in different sectors ongoing, often at the global level (ISO or similar). Such standards are most often open and developed on a consensus-based between various actors of the industry. These standards not only refer to data but also to data management, including practices to ensure the uniqueness of identifiers and resolve data conflicts. Further down in this report and in the annex to this report we give an extensive account of these and give overview of ongoing initiatives.

However, a standardisation does not ensure that all rightholders provide the data necessary to maintain data accuracy and completeness. The reason is that the cost of data entry may outweigh the revenue accruing from visibility and “findability” in a comprehensive database⁸². This can be especially the case for small repertoire holders in the light of economies of scale of data entry.

Certain sectors, such as the music industry, are more advanced than other sectors concerning established standards for metadata file formats. However, there are still voices

⁷⁷ French Ministry of Culture, *Mission Report – Towards more effectiveness of copyright law on online content sharing platforms: Overview of content recognition tools and possible ways forward*, French Ministry of Culture, Paris, 2020.

⁷⁸ Senftleben, M., Margoini, T., Antal, D., Bodo, B., van Gompel, S., Handke, C. et al., ‘Ensuring the visibility and accessibility of European creative content on the world market: the need for copyright data improvement in the light of new technologies’, *SSRN*, 2021.

⁷⁹ Council of the European Union, ‘Developing the copyright infrastructure – stocktaking of work and progress under the Finnish Presidency’, Brussels, 20 December 2019 (<https://data.consilium.europa.eu/doc/document/ST-15016-2019-INIT/en/pdf>).

⁸⁰ Intellectual Property Office and Ulster University, ‘Music 2025 summary paper’, 18 June 2019.

⁸¹ Council of the European Union, ‘Developing the copyright infrastructure – stocktaking of work and progress under the Finnish Presidency’, Brussels, 20 December 2019 (<https://data.consilium.europa.eu/doc/document/ST-15016-2019-INIT/en/pdf>).

⁸² Senftleben, M., Margoini, T., Antal, D., Bodo, B., van Gompel, S., Handke, C. et al., ‘Ensuring the visibility and accessibility of European creative content on the world market: the need for copyright data improvement in the light of new technologies’, *SSRN*, 2021.

that call for improvements for the processes or workflows by which metadata is collected, displayed and validated⁸³.

2.3.3.4. Fingerprinting

Digital fingerprinting consists of embedding an identification sequence in content in such a way that each recipient receives a slightly different copy that can be traced back to them⁸⁴. The data has a unique digital identifier, analogous to human fingerprints⁸⁵. In this way, an illegal redistributor can be identified, and legal action can be taken against them. So instead of preventing content from being copied or redistributed, digital fingerprinting focuses on identifying users responsible for the illegal redistribution⁸⁶. It also guarantees that the identity of the buyer is not known during distribution, as required by the GDPR. However, if a buyer is found to be involved in copyright infringement, his or her privacy can be revoked.

Fingerprinting is the most widespread solution for identifying protected content and is used ever more frequently⁸⁷. It is widely applied to audio and video content, in particular by YouTube and Facebook⁸⁸. Fingerprint identification with currently available tools misses only a very small portion of the protected content (false negatives) and identifies an equally small portion of false positives⁸⁹.

At the same time, fingerprinting is a challenge for both rights holders and platforms. Both groups must be able to maintain a reference base that is broad enough to enable content detection, given a large volume of uploaded content. Both extensive storage capacity and rapid content analysis capabilities are required for fast and accurate response⁹⁰.

Fingerprinting solutions can also be used to identify works that are not correctly reported by licensees. This requires that rightholders and collecting societies work to ensure that fingerprinting solution providers have the files in their databases to enable pre-identification of works and reduce the amount of royalties that fall into the pool of unclaimed royalties⁹¹.

⁸³ Hardjono, T., Howard, G., Scace, E., Chowdury, M., Novak, L., Gaudet, M. et al., 'Towards an open and scalable music metadata layer', *arXiv*, 2019.

⁸⁴ Megías, D., Kuribayashi, M. and Qureshi, A., 'Survey on decentralized fingerprinting solutions: copyright protection through piracy tracing', *Computers*, Vol. 9, No 2, 2020.

⁸⁵ French Ministry of Culture on online content sharing platforms".

⁸⁶ Megías, D., Kuribayashi, M. and Qureshi, A., 'Survey on decentralized fingerprinting solutions: copyright protection through piracy tracing', *Computers*, Vol. 9, No 2, 2020.

⁸⁷ Council of the European Union, 'Developing the copyright infrastructure – stocktaking of work and progress under the Finnish Presidency', Brussels, 20 December 2019 (<https://data.consilium.europa.eu/doc/document/ST-15016-2019-INIT/en/pdf>).

⁸⁸ Note: they use fingerprinting more here for blocking and monetising content, rather than in the anti-piracy process discussed in the previous paragraphs.

⁸⁹ French Ministry of Culture, *Mission Report – Towards more effectiveness of copyright law on online content sharing platforms: Overview of content recognition tools and possible ways forward*, French Ministry of Culture, Paris, 2020.

⁹⁰ French Ministry of Culture, *Mission Report – Towards more effectiveness of copyright law on online content sharing platforms: Overview of content recognition tools and possible ways forward*, French Ministry of Culture, Paris, 2020.

⁹¹ Butler, S. P., *Collective Rights Management Practices Around The World – A survey of CMO practices to reduce the occurrence of unclaimed royalties in musical works*, US Copyright Office, Washington DC, 2020

2.3.3.5. Data infrastructures

The idea of a *copyright infrastructure* was already debated at WIPO 10 years ago. The meeting 'Enabling creativity in the digital environment: copyright documentation and infrastructure' was convened to raise Member States' awareness of the important role that the copyright data infrastructure can play, especially in the digital environment⁹². It covered topics ranging from public registration and legal deposit systems to balanced and effective dissemination of creativity over the internet. A few days before, a WIPO working paper asked, 'What copyright infrastructure is needed to facilitate the licensing of copyrighted works in the digital age: the international music registry?'⁹³.

In parallel, a group of music stakeholders⁹⁴ attempted to build a Global Repertoire Database (GRD) to provide a single, comprehensive and authoritative representation of the global ownership and control of musical works. Once deployed, the GRD would have saved extensive costs, lost to duplication in data processing⁹⁵. The project was abandoned in 2014.

Blockchain emerged and led to new attempts to develop *copyright data infrastructures*. More than 30 of these were listed in a 2018 study⁹⁶. During the German Presidency of the Council of the European Union, the presentation of the conception of a European decentralised copyright platform for music⁹⁷ was made. The authors argued why a decentralised approach would make sense. According to the study authors, a comprehensive database is needed but centralised approaches have been unsuccessful, controversy over who would have control over the data and who would manage the catalogue would be circumvented, power relations would no longer hinder participation in a common platform, and the sole platform purposes would be data exchange and data quality improvement.

⁹² Marusak Hermann, R., 'Copyright infrastructure in the digital age: raising awareness at WIPO', Intellectual Property Watch, 19 October 2011 (<https://www.ip-watch.org/2011/10/19/copyright-infrastructure-in-the-digital-age-raising-awareness-at-wipo/>).

⁹³ WIPO, 'What copyright infrastructure is needed to facilitate the licensing of copyrighted works in the digital age: the international music registry', WIPO, Geneva, 2011 (https://www.wipo.int/meetings/en/doc_details.jsp?doc_id=183060).

⁹⁴ With the involvement of CMOs from North America, Australia and Brazil.

⁹⁵ PRS for Music, 'Global Repertoire Database makes strong progress with plans for 2013 well underway', 24 January 2013 (<https://www.prsformusic.com/press/2013/global-repertoire-database-makes-strong-progress-with-plans-for-2013-well-underway>); PRS for Music, 'Global Repertoire Database Working Group', 10 December 2010 (<https://www.prsformusic.com/press/2010/global-repertoire-database-working-group>).

⁹⁶ PwC, *Nach dem Streaming kommt die Blockchain, Hype oder echte Chance für die Musikindustrie?*, 2018 (<https://www.pwc.de/de/technologie-medien-und-telekommunikation/pwc-studie-nach-dem-streaming-kommt-die-blockchain-hype-oder-echte-chance-fuer-die-musikindustrie.pdf>).

⁹⁷ Gronau, N., 'Conception of a European decentralised copyright platform for music', presentation, University of Potsdam, Potsdam, Germany.

2.4. An overview of initiatives addressing rights metadata issues

With the aim of recognising the manifold activities that many organisations in different cultural industries currently are undertaking to improve copyright metadata management processes, this section summarises a **non-exhaustive list of initiatives** that currently play a role, or propose to play a role, in copyright metadata issues. While the topic may include issues related to data governance, data management, standards and interoperability⁹⁸, and others, we take the decision in this section to focus on interoperability. We suggest that any serious analysis and initiative must take these initiatives into account from the angle of interoperability. A first reason is that these initiatives – for instance the standards of identification and rights metadata and many systems of data access/exchange – already play an **essential and established role** and, therefore, any new initiative should integrate with the existing landscape as appropriate. Another reason for taking them into due consideration is because they – for instance frameworks, working groups and reports – provide guidance, tools or potential tools for solving aspects of the interoperability challenge.

The majority of the initiatives described come from specific content or cultural sectors (music, text publishing, audio-visual, images, libraries, museums) but because our focus is on **interoperability** across as well as within these domains, they are grouped here according to the functions they fulfil rather than the domains they inhabit.

Some of the initiatives are “digital natives” (that is, born in and for the digital age); others, in particular some of the major identifier and metadata standards, straddle the domains of digital, physical and abstract (“works”) content use. This is because our focus is not on interoperability in the supply chain for digital **content** but in the supply chain for digital **data**, which intermixes rights in digital, physical, and abstract (“works”) content. Creators, rightsholders, publishers, collective organisations and platforms need methods which support data for all these together. It is also the case that many rights transactions are for uses which cross the abstract-digital-physical divides, such as adapting, copying, and printing text documents.

Those listed are collective, open, or shared initiatives. There are of course a huge number of established or hopeful proprietary or collective systems and services operating in the digital rights network which currently play data interoperability roles: for example, collection society and library databases; proprietary identifiers such as Amazon Standard Identification Numbers (ASINs) or Dun & Bradstreet’s Data Universal Numbering System (DUNS Numbers); any number of proprietary content publishing, licensing, and tracking services and systems. These need to be understood in general, but the history of content supply chains tells us that so long as none of these has a monopoly position, they will adapt their data and processes to accommodate well-designed standards or exchanges which provide more opportunity, so the detail of proprietary systems is not a dominant factor in designing for interoperability.

The table (accessible in the Annex to this report) lists initiatives for convenience under these headings, recognising that there is some overlap in scope:

⁹⁸ Here, interoperability is meant to include exhaustiveness and authority, that can be considered as conditions of interoperability. Furthermore, interoperability refers explicitly to a distributed framework.

- A. Working groups and reports
- B. Data frameworks
- C. Identifier standards
- D. Identifier standards with metadata
- E. Metadata standards (content and rights)
- F. Metadata standards (rights)
- G. Data access/exchange
- H. Datasets
- I. Rights platforms
- J. Authentication
- K. Policing

The list of identified initiatives can be accessed in Annex 5.3 of this report.

These initiatives are very important and must be considered as the starting point for an analysis of the current situation regarding the state of metadata in the copyright industries. At the same time, it is clear that gaps still exist as well as room for improvement. In this respect, the following **main observations** based on an analysis of the list of initiatives emerge:

In terms of the main building blocks for data interoperability – **identifiers and metadata schemas** – the following primary sectoral gaps seem to exist:

- The image sector has not established shared identifiers for content, parties or rights.
- In the publishing sector, the failure of the International Standard Text Code⁹⁹ means that (unlike musical works) many textual works are not identified in any standard way except in relation to some specific manifestation identifier. In addition, many smaller or fragmentary textual items such as blog posts or poems have no standard identification at all. DOI provides work identifiers for journal articles and many related types of content and is starting to be used to cover more of the gaps such as book chapters,

Most evidently, for the focus of this study, there are no standard identifiers for rights (as opposed to standards to identify manifestations of works and standards to identify people (authors, performers, rightsholders)), and only proprietary identifiers for licences in different sectors.

There are other types of initiative not listed in the overview in the annex. 21st century technical innovations such as blockchain bring some new tools to the network. In the music sector, there has been a first proliferation of blockchain-based tools that intended to help musicians to self-publish and self-license, without publishers or CMOs¹⁰⁰. Now, we can already observe a higher level of maturity in initiatives and companies such as **Fuga** (www.fuga.com) or **Unison** (www.unisonrights.es), a full-fledged Independent Management Entity (IME). However, while distributed blockchain ledgers offer new options for trading and security, they do not add anything new to most of the data interoperability challenges and seem likely to present some new data-related challenges of their own.

⁹⁹ See <http://www.istc-international.org/>

¹⁰⁰ PwC, *Nach dem Streaming kommt die Blockchain, Hype oder echte Chance für die Musikindustrie?*, 2018 available at: <https://www.pwc.de/de/technologie-medien-und-telekommunikation/pwc-studie-nach-dem-streaming-kommt-die-blockchain-hype-oder-echte-chance-fuer-die-musikindustrie.pdf>

Initiatives on signing, authentication and certification of parties are a quite rapid recent growth area and should provide some of the essential trust mechanisms needed for the rights data framework.

Initiatives on information/education also tend to be sector-based and are likely generally to remain so as this is where they are most effective.

2.5. Summary of sectoral interviews on pain points and priorities related to copyright metadata

Data management is – for many reasons¹⁰¹ – challenging for all creative industries. However, there are large differences between industries (and stakeholders within one industry) on how prevalent these matters are perceived to be and how high they are (or should) ranked on the industry or policy agenda.

In this section we present our insights and analysis on the main “**pain points**” related to rights management in the cultural industries addressed in this study. With the term pain points we refer to the main **challenges** or issues which were identified during the study.

The analysis relies on a literature review on issues related to copyright metadata (see previous section) as well as more than 80 interviews covering three creative industries; film & television (28 interviews), music (31 interviews), and publishing (books, journals, photography, and press¹⁰²; 24 interviews) in the 27 Member States of the European Union, Canada, the United Kingdom, and the United States. It is important to keep in mind that the interviewees exemplified the stakeholders of the digital value network; authors, performers, rightholders, producers, publishers, management organisations, broadcasters, online platforms, intellectual property offices, and technology providers. They were represented by individuals, companies, or trade bodies. The interviews were conducted under the Chatham House Rule¹⁰³ between January and May 2021, comprised up to 3 sessions each, and were often followed by the exchange of documents. The interviews were also instrumental to discuss ongoing initiatives (see section 2.4) and necessities for future avenues (see section 2.7).

The study team made large efforts to design an interview programme covering exchanges with a large variety of stakeholders (and therefore also a variety of perspectives and needs). The limitation to the number of interviews which could be carried out in this project implies though that in some cases a “complete” representativeness of every nuanced view in all sectors and sub-sectors might not have been reached. This should be taken into account in interpreting the following findings.

Based on interviews and literature review, it can be summarised that the analysed creative sectors are primarily facing data-related challenges in the following four areas:

- **Costs**, in the area of *rights management*
- **Efficiency issues**, in the domain of *licensing*
- **Challenges concerning payment processes**, in the field of *rights remuneration*
- **Risks of misappropriation and other rights infringements**, in the sphere of *rights enforcement*

¹⁰¹ See for example our analysis of different metadata requirements formulated in European Directives in section 2.2.

¹⁰² The creative industries include also sectors such as advertising, software and video games, out of scope of this study part but for which many of its insights would apply.

¹⁰³ “When a meeting, or part thereof, is held under the Chatham House Rule, participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s), nor that of any other participant, may be revealed” on [chathamhouse.org](https://www.chathamhouse.org), May 2021.

In the following sections, these four pain points and their drivers are characterised in detail. We will also refer to sector-specific differences, as not all pain points are equally relevant in all cultural industries.

2.5.1. Costs related to rights management

2.5.1.1. Characterisation

"Managing assets is a crucial aspect of any organisation, and digital assets are no exception. With the massive supercharge of digital media, companies are creating more digital than physical assets. These include but are not limited to images, videos, blogs, and digital products"¹⁰⁴. From this statement alone, the importance of digital asset management and the management of the digital value chain becomes immediately clear.

A **generic digital media supply and value chain** can be defined by the following 5 phases¹⁰⁵:

- **Create**, whereby assets (i.e., pre-existing content and intellectual property) and talents (i.e., creators, performers and other contributors) are acquired, products or events are planned, portfolios are managed, and digital media is mastered, formatted and localised,
- **Manage**, that includes the generation, protection, storage and maintenance of identification, rights, descriptive, and administrative metadata coming from the creation phase and necessary for the delivery, storage and retrieval phases,
- **Deliver**, which can in turn consist of 4 sub-phases: (1) licensing which requires first business modelling, programming, promoting, and selling, and then licensing itself and contract management, (2) distribution, which includes ordering, transferring, fan engagement, and residual distribution, (3) eventually tracking, usage metering and reconciliation, and (4) payments, i.e., invoicing, transfers, and accounting,
- **Store**, whereby masters¹⁰⁶ are managed, protected, and stored,
- **Retrieve**, whereby the digital media (and related metadata) are retrieved to be enjoyed for itself or embedded in a next cycle of creation.

At each of these steps, **metadata accompanies the digital media asset**. This metadata requires human resources and systems that are costly. Moreover, at each of these little steps the creative industries are facing **challenges**, varying from sector to sector, jurisdiction to jurisdiction, or size of organisation to size of organisation.

Examples of these challenges include the following: how can one acquire copyright-protected pre-existing content to be aggregated in new products or adapted for new creations, what data about creators and contributors is necessary to assure their fair remuneration across the value chain, how can (descriptive) metadata help streamline the creative supply chain by reducing the number or complexity of manual administrative

¹⁰⁴ See <https://filecamp.com/blog/how-to-create-a-digital-asset-management-strategy/>, retrieved in May 2021

¹⁰⁵ Based on "5 Reasons Why Digital Asset Management for Photographers Is a Must", retrieved at <https://getprostorage.com/blog/digital-asset-management-for-photographers/> in May 2021

¹⁰⁶ In the media business, a **master** recording is the official original recording of a song, sound, performance, image, film, text, etc.

tasks, how can digital assets be identified and tracked through the creative process and beyond, what masters or metadata must be captured and maintained, how to organise a hub to manage, protect and store digital assets and related metadata, how to ease licensing through data taxonomy, contract templates and standard processes, how to integrate intra or inter-company dataflows, how to process transactions to ensure timely payments to all contributors and rightholders, and so forth.

2.5.1.2. In the film & television industry

The film & television industry is – at least from a metadata management point of view – dealing with **exclusive** content, compared to the music or news sectors, that – from the same point of view – handle content **commodities**. In the film & television industry¹⁰⁷, a much smaller number of producers are negotiating mainly exclusive licences with a much smaller number of distributors or retailers. Based on interviews, we propose that the **metadata priorities** of the film and television industry can be ranked as follows: (1) the exclusive content must be found to reach a maximum audience (requires mainly descriptive metadata), (2) the exclusive content must be protected against misappropriation (requires mainly administrative metadata), and (3) contributors must be remunerated (requires mainly rights metadata)¹⁰⁸.

The priorities vary also within the film & TV industry, primarily between commercial producers who must recoup their investments (high priority for licensing-relevant metadata) and public producers/broadcasters who receive public funds to cover production costs (low priority for licensing-relevant metadata)¹⁰⁹.

Descriptive metadata are used to make films and TV programmes discoverable and to make them accessible, typically by subtitling¹¹⁰. This requires high-skilled professions and is the domain of promising research, using artificial intelligence either to produce indexing metadata or subtitles, two subsets of descriptive metadata¹¹¹.

Despite existing and ongoing initiatives (for example the joint work of EIDR and ISAN to create Interoperable Audiovisual Standard Identifiers) the interviewees¹¹² tended to perceive a lack of rights metadata **standardisation** (more precisely a lack of standards and identifiers existing in the AV sector) and **system interoperability** as well as data quality as major pain points when it comes to production, licensing-relevant, customer-discovery and rights metadata. It starts at financing and production levels and impacts the whole value chain in the European audiovisual sector. An exemplary, non-representative

¹⁰⁷ We refer here to the mainstream or bulk of the business. Situations varies between major studios and independent producers, and between Europe and the USA, for example.

¹⁰⁸ Based on 15 interviews in the film and television sector covering both public and commercial producers and distributors.

¹⁰⁹ For descriptions of the film value chain, see E. Madudova, *Creative industries value chain: The value chain logic in supply chain relationships*, July 2017, P. Bloore, *Re-defining the Independent Film Value Chain*, BFI, 2011, or A. Finney, *The film value chain model and current restructuring*, Routledge, 2014

¹¹⁰ See the Greta & Starks app that enables people with sight or hearing loss to experience fully accessible cinema and includes foreign language subtitles and audio versions for an international audience at <https://www.gretaundstarks.de/>

¹¹¹ See <https://ai4mediadata.com/> and <https://largo.ai/>

¹¹² As stated above, this section is largely based on interviews with industry stakeholders, sector experts involved in the study team as well as pre-existing metadata expertise by the study authors. Out of a total of 83 interviews for this project, 28 were carried out in the film/television industry. Interviewees include organisations affiliated to major public and private film producers, large European broadcasters, distributors, but also independent industry experts.

analysis of metadata coming from 6 interviewed producers/broadcasters within the same film niche (music and dance films) showed that none of them were using the same structure of metadata, that metadata sets were missing metadata and/or included erroneous metadata, that none of them were using standard identifiers such as EIDR or ISAN, and that all of them were using labour-intensive data entry procedures. While this example is not necessarily generalisable to all industry players, it gives an indication of the challenges that seem to remain for interoperability in parts of this sector.

The section 2.6 will cover interoperability in more detail and provide further evidence and analyses.

Limecraft Flow¹¹³ is a good example of the most recent developments for the capture of metadata during production and BBC Silvermouse¹¹⁴ a good example of state-of-the-art post-production recording of metadata.

2.5.1.3. In the music industry

The music rights industry has two **distinct strands** each focused on a different set of rights – the record industry focused on recordings (or the “master rights”), and the music publishing sector focused on song rights for lyrics and compositions (or the “author rights”, see Figure 1). Key stakeholders in the record industry include recording artists, record labels, music distributors and the collective management organisations (CMOs) representing performers and labels. Key stakeholders in the music publishing sector include songwriters, composers, music publishers, and the CMOs representing authors and publishers (for example those represented by the leading network of CISAC).

Each of the strands has their own **data standards** (e.g., in the record industry ISRC to identify recordings and IPN/ the International Performers Database for performers; in the publishing sector ISWC to identify musical works and IPI to identify interested parties like songwriters and publishers), as well as their own databases and data initiatives. This of course implies challenges for the **interoperability** of data. The main industry-wide databases of recordings and musical works are managed by the CMOs.

There are also important industry-wide initiatives like DDEX initiative and the RDx project to better connect CMO databases and to improve metadata and especially metadata interoperability in the sector.

In principle, however, each of the above-mentioned strands has different data challenges, with the greater challenges arguably on the publishing side, partly because co-ownership of rights is common when it comes to musical works (see also below). Nevertheless, one **key issue** – identifying what song is contained in what recording (so **matching ISWCs to ISRCs**) – is a **challenge that unites the two strands**, even if the impact of that issue mainly affects songwriters and music publishers.

¹¹³ See <https://www.limecraft.com/features/>. Limecraft Flow allows an intuitive and efficient media organisation and metadata entry and the exportation of metadata to multiple document formats.

¹¹⁴ See <https://www.bbc.co.uk/academy/silvermouse-for-independent-productions/>. Silvermouse is a web-based system which issues forms to production teams to record the post-production paperwork for their TV programmes and non-linear audio-visual content.

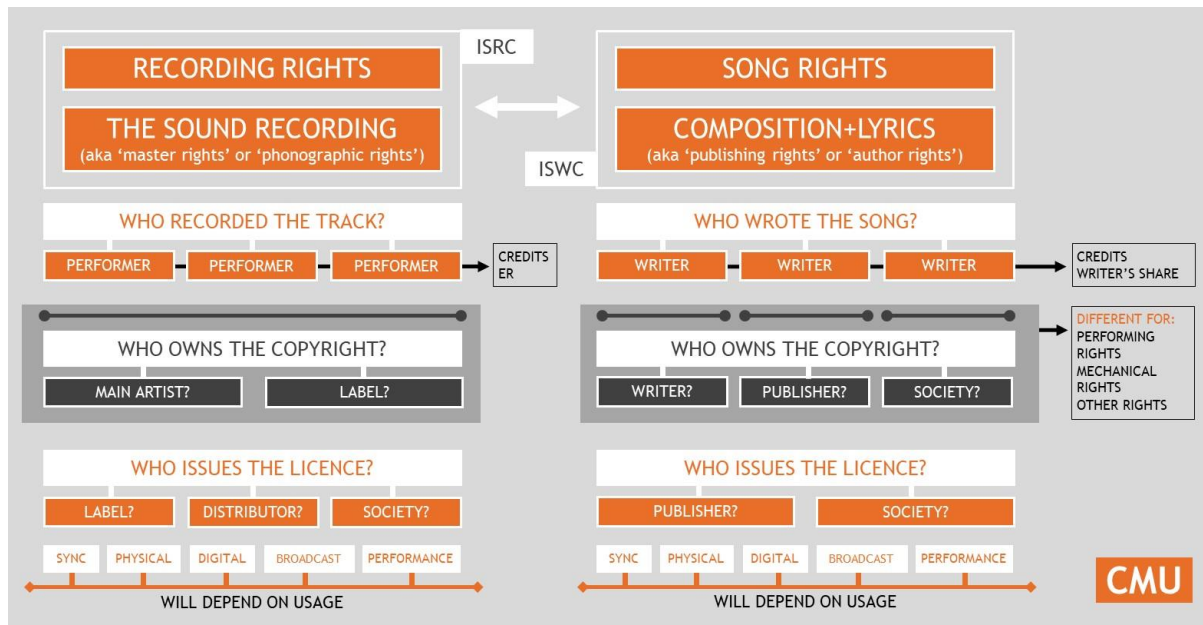


Figure 1: The rights framework in the music industry and implications for music licensing

Source: CMU Insights

Licensing in the music industry is very complex. That complexity can be traced back to different drivers¹¹⁵:

- Collaboration has long been common, especially with pop music, but the number of stepwise co-authorships has definitely increased among songwriters and composers.
- The number of performances and recordings of each song by the same original performers, in studio, live, best of, etc. can be very large, the number of licensed covers even larger.
- Songs can be embedded, as a whole or in part, in many audio-visual products, from a 30-second advert to a feature film.
- The two copyrights, one for the work (the song) and one for the recording, need to be matched.

Against this background, in the interviews¹¹⁶, three main metadata **challenges** were mentioned, with their **drivers** leading to **pain points**. These challenges were repeatedly brought up by the interviewees, although the concerns were not shared to the same extent nor given the same importance or priority by all interview partners: for example interview partners from recording industry associations pointed to the huge efforts currently

¹¹⁵ Based on the interviews, cross-checked with the industry expert advising the study team and with Stage Enterprises, technology provider developing the RDx and Cube systems.

¹¹⁶ For this section, 17 interviews were carried out in the music industry. Interviewees included 3CM Unlimited, Bitfury/Surround, Sunbeam Productions, Nielsen/Gracenote, Teosto, IFPI, Pex, County Analytics, Stage Enterprises, Rightscom/DDEX, Hipgnosis Songs Trust, Impala, ICMP, Google/YouTube, SACEM, Music Finland, ICE Services

undertaken in standardisation and rights data management processes. As can be seen – for example – by the high number of ongoing initiatives listed in the annex to this report and summarised in section 2.4, there is certainly a lot of attention given to this topic by various stakeholders.

However, remaining challenges which came up in the research for this study included the following:

- **Matching metadata: *Driver*:** the identification of a specific song, or (all) the songwriters and music publishers, with an interest in the song, or of the revenue splits between these songwriters and music publishers might not always be known at the point in time when the distribution/the streaming starts (or can start). ***Pain point*:** necessity to match *ex post* the rights management information pertaining to the works contained in the recordings¹¹⁷. This is a costly process that may only be sustainable above a certain number of streams per month, and even then not 100% complete or accurate.
- **Duplication of metadata: *Driver*:** some registration records actually refer to the same recording: (1) Recording titles are not unique and therefore not sufficient to cluster, e.g., multiple Karajan Beethoven cycles, (2) Studio and live versions will have the same title and artist, (3) Same recording can have different language metadata, e.g., Blue Danube Waltz/An der schönen blauen Donau, (4) Timing issues when multiple entry points exist. There can be different mix ups – some single recordings have multiple ISRCs (usually because a new ISRC was issued after a change in distributor or label) and multiple recordings can have the same ISRCs (where a different version of track is released with the same ISRC). ***Pain points*:** cases of duplicate registration of the same recording may cause confusion and delay, cost to fix, revenue misattribution, and loss of trust in the system.
- **Capture and curation of metadata: *Driver*:** progress is being made to capture rights management information as close as possible to the moment of creation¹¹⁸ and alleviate the issues mentioned earlier. However, (re)capturing and cleaning rights metadata pertaining to catalogues is fastidious. A rightholder, who recently bought catalogues of hits, estimates that only 3% of the songs have exhaustive reliable and up-to-date metadata, i.e., have all data needed for exploitation. ***Pain points*:** data curation costs relying on very qualified personnel and sophisticated systems, in the meantime loss of potential revenue; delayed capture of rights management information also increases the likelihood of data conflicts and disputes down the line, which can delay and halt payments, and could ultimately result in costly litigation.

¹¹⁷ See also, for example, a quote by Robert Ashcroft, PRS for Music chief executive: 'Today we spend an inordinate amount of time correcting errors and resolving disputes which hold up payments to music creators (<https://www.prsformusic.com/m-magazine/news/auddly-launches-metadatapays-campaign/>); see also a quote by Director of music publishing, YouTube, saying that 'solving the lack of ownership information in the music publishing industry remains an ongoing challenge' (<https://www.prsformusic.com/m-magazine/news/tool-launched-by-abbas-bjorn-ulvaeus-to-receive-1-million-grant/>)

¹¹⁸ See for example Session at <https://session.id/>. Session proposes self-serve mobile and web applications to support song collaborations, simplifying and improving the early identification of accurate and authoritative music registration data, capturing detailed creator contributions including the early identification and agreement of copyright ownership between creators, publishers, and labels.

2.5.1.4. In the publishing industry (books, press, journals, and images)

When it comes to rights metadata, the interviews¹¹⁹ showed that there are major **differences between publishing sectors**. In general, rights management is well organised for printed publications and no pain points are reported by the various actors of the industry. Problems start with the digitisation of content, whereby eBooks and journals have invested heavily in developments of standards and exchanges (e.g., ISBN, DOI, ONIX) which are widely used. Although photography (e.g., IPTC) and press (e.g., ISSN) made also substantial efforts, their standards are less used.

New rights or exceptions to exclusive rights trigger the necessity of deploying more efforts, specifically in the field of metadata, to exploit them.

- **First example: Reservations** – Article 4 §3 of the Directive EU/2019/790 on copyright states that the exception or limitation provided for text and data mining shall apply on condition that the use of works and other subject matter has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.
- **Second example: Granularity** – Article 15 of the same Directive EU/2019/790 defines the related right of press publishers on press publications (except for very short excerpts). The economic value of this right is potentially important, but its granular exploitation (e.g., for one excerpt of one article of one issue of one newspaper) will be challenging.

Pain point 1: Costs

- The production process of digital media in creative industries brings along manifold challenges related to rights data management – from acquiring content for re-use in new works to capturing and maintaining metadata. They cause substantial administrative efforts and costs.
- These costs seem to be specifically critical in the music industry, mainly because of higher complexities in the stakeholder setup and the copyright and licensing system itself. There are without doubt important activities ongoing to optimise IT systems and processes for rights data exchange (like DDEX in the music industry). This certainly helps to make processes more cost-effective. However, challenges remain. For example, it seems that rights management information is at times not taken care of *ex ante* in the early stages of the music value chain and must be dealt with *ex post*.
- In the publishing sector, no substantial pain points are reported in the subsector of printed publications.

¹¹⁹ 24 interviews were carried out for this report in the publishing industry. The interview partners came largely from publishing associations (books, journals, newspapers and magazines). Other interviewees included online platforms, technology providers and standardization bodies.

2.5.2. Efficiency issues in the context of licensing

2.5.2.1. In the film & television industry

The film and TV industry, like other creative sectors, is a heterogeneous cluster of very different subgroups. Interviews with the representatives of the large film studios revealed that, for them, the market is operating efficiently and relies on the festivals and well-established networks of large players on the production and distribution sides. The situation is similar among live broadcasters of major sport events. In this sense, these interview partners did not report “pain points” with respect to markets efficiencies.

The situation is very different among the independent sector and niche players who form the majority of European audio-visual producers – independent film producers, documentaries, music films, etc.. Interviews of film producers and distributors – outside the blockbuster circuit – have helped list pain points from their separate points of view.

Film producers that have been interviewed¹²⁰ mentioned instances of scattered, complex, diverse and at times incomplete metadata and a lack of standardised identification. This was linked in the interviews to, for example, barriers across market segments preventing cross-sectoral collaboration and distribution and fragmented markets where it is difficult to find the accurate and authoritative metadata.

Film distributors who have been interviewed mentioned: the difficulty to get all rights one needs for a broadcast into one contract due to the number of rightholders of one asset as well as (1) poorly stocked, illiquid, and fragmented market where it is difficult to know what is available, (2) suboptimal business and remuneration models, (3) difficulty to reach new audiences, and (4) lack of scalability.

The independent film producers and distributors expressed clearly what they expect from efficient markets, (1) the possibility to find content or for their content to be found by distributors, retailers and audiences (requires descriptive metadata), then (2) the possibility to find each other (requires producer and distributor identification and rights metadata), later (3) systems helping sellers and buyers draft licence agreements (require rights metadata and embedded knowledge of industry practices and copyright regulations), finally (4) systems automating licensing for example systems generating and managing micro-licences (require granular identification of content and trusted rights management information).

Notable European development efforts to meet the industry expectations and build efficient markets for independent players include:

- **Cinando¹²¹**, the premier online network for film professionals offering tools to navigate the film industry: contacts, films, projects in development, market screening schedules, market attendees, screeners, etc.
- **The decentralised applications of Cascade8¹²²** including Archipel, a marketplace for TV, VoD and ancillary rights, a film market platform, aiming at

¹²⁰ As mentioned above, the interviewees include around 15 organisations including major public and private film producers, large European broadcasters, distributors, but also independent industry experts.

¹²¹ See <https://cinando.com/>

¹²² See <https://cascade8.com/dapps/>

facilitating screenings and interactions between sales agents and buyers, and a marketplace for content investors

- **Cinemarket**¹²³, the platform that aims at connecting sellers and buyers on a 24/7 digital film market
- **Content.Agent**¹²⁴, the B2B online marketplace and search engine for music + dance films developed by IMZ Members for IMZ Members.
- **Vuulr**, a new player in the market, <https://www.vuulr.com>
- **FilmarketHub**, <https://www.filmarkethub.com>
- **Slated**, a US-based platform with major European talent involved, <https://welcome.slated.com>
- **White Rabbit**, just launched in Cannes¹²⁵

2.5.2.2. In the music industry

On the works/songs side – metadata issues make licensing a service a **major task** because basically each song needs to be licenced, and the service provider does not necessarily know what songs are in the recordings the label/distributor upload, nor who owns those rights.¹²⁶ This makes it harder for services to come to market, contributing to a marketplace dominated by a small number of well-funded major players. It also pushes quite a lot of the royalty processing costs onto the industry, which are incurred by the publishers and songwriters. Interestingly, in the USA under the MMA the services pay for the MLC, so are covering these costs. They agreed to this because it was better than the status quo of being sued for not knowing who to pay.

Discussing market efficiency with rightholders in the music industry leads immediately to considerations of fairness (linked to the percentage of retail revenue distributed to rightholders) and transparency (linked to the size of retail revenue), overtaking in some ways the issues related to the licensing mechanisms themselves.

This study does **not** assess the appropriateness of a specific royalty percentage for a specific stakeholder of the music value chain. It does show the relationship between this topic and metadata, especially usage metadata. Indeed, the remuneration of a rightholder is not only a matter of percentage (the relative size of the slice of the pie), but also a matter of revenue or total monetisation (the size of the pie). Usage metadata can inform on revenue and total monetisation. Sharing usage metadata between retailers and rightholders can contribute to restore a level-playing field; however, sharing usage metadata requires a precise allocation of the retailers' revenues and the respect of business confidentiality.

¹²³ See <https://www.cinemarket.io/en>

¹²⁴ See <https://www.imz.at/imz-online-services/>

¹²⁵ See <https://www.filmneweurope.com/fne-innovation/item/122004-fne-innovation-white-rabbit-embrace-ze-fans-a-rabbit-chasing-pirates>

¹²⁶ Again, the situation has improved compared to the past due to projects like DDEX and others.

The topics of market efficiency and fairness were also brought together when the authors of reports¹²⁷ commissioned by the UK Intellectual Property Offices or House of Commons spoke about playlists and recommendation algorithms. They are in turn fed by data: descriptive and usage metadata, not to mention user metadata: their perceived preferences or the perceived preferences of other – somehow related – users.

The interviews showed that some links between high-quality metadata and efficient markets are not always clearly understood by rightholders or the bodies representing them.

2.5.2.3. In the publishing industry (books, press, journals, and images)

The interviewed stakeholders from the publishing industry commented very differently on the efficiency of the markets. We can summarise their positions under three categories:

- Our management of rights metadata allow us to market our content as we want,
- Increasing the efficiency of the market is not our priority, and
- Inefficient markets do not allow us to fully exploit the digitisation of our content and we fear that we are missing out on a lot of revenue.

If licensing appears to be relatively well mastered by trade book publishers, it seems that there are still a lot of non-exploited micro-licensing opportunities.

Interestingly, there are already proven ground-breaking mini or micro-licensing solutions and services. For example:

- **Pay-Per-Use Services**¹²⁸, providing a quick-and-easy way for faculty, students and staff at colleges and universities to secure permissions to use and share content from leading titles in science, technology, medicine, humanities, news, business, finance and more, typically in course packs and classroom handouts or to republish an article, book excerpt or other content in your own books, journals, newsletters and other materials.
- **RightsLink® for Scientific Communications**¹²⁹ to automate the collection of open access publication charges as well as submission, page, colour, custom cover, reprint, and other value-added services, and to give publishing programmes the flexibility to diversify and test, build, and support sustainable business models.

Further conversations with a European press association and one of their national members confirmed that micro-licensing could quickly become a priority if publishers would see an efficient way to do it – which presupposes granular identification of content and automated micro-licensing, as indeed the costs of manual micro-licensing would exceed the value of the related micro-licence. In turn, automated micro-licensing requires reliable, exhaustive, and up-to-date rights metadata – “exhaustive” meaning all the metadata necessary to licence the content and remunerate the authors and rightholders.

¹²⁷ Music 2025 (<https://www.gov.uk/government/publications/music-2025-the-music-data-dilemma>), Economics of Music Streaming (<https://committees.parliament.uk/work/646/economics-of-music-streaming/publications/>), and UK music creators' earnings in the digital era (<https://www.gov.uk/government/publications/music-creators-earnings-in-the-digital-era>)

¹²⁸ See <https://www.copyright.com/academia/pay-per-use/>

¹²⁹ See <https://www.copyright.com/publishers/manage-publication-charges/>

Similarly, a major book publisher association mentioned the still untapped opportunity of micro-licensing in the education sector. Nowadays, teachers and professors tend not to follow one printed book to cover the matter of a multi-lecture course, but to curate and distribute among pupils and students the supporting content, lecture by lecture, from various digital sources. They are ready to pay for the limited use, but the costs to administer the limited licence is a multiple of the royalty.

As mentioned in the literature review, 96% of images in news lack metadata. This figure has been confirmed by a start-up developing solutions for image protection. One of the drivers of this situation is the lack of understanding of what metadata are and how they are used. Many photographers and graphic designers still think that writing their names on or underneath the image would be sufficient. That prevents them from seizing the market opportunities opened by years of collaboration between the International Press Telecommunications Council (IPTC)¹³⁰, the Coordination of European Picture Agencies Stock, Press and Heritage (CEPIC)¹³¹, and Google that led to the "Licensable Images" feature¹³² enabling users to be directed to rights data and to a page where the image can be licensed from.

Pain point 2: Efficiency issues in the area of licensing

- Digitisation has led to much lower barriers to entry in the creative industries and concretely to a rise of creations and products. Digitisation of rights licensing, let alone automated licensing, has so far not kept up with these developments.
- Despite important improvements related to developed standards, there stills seems to exist some potential for further efficiency gains with respect to licensing processes in the music industry and the related rights metadata needed for this. Specifically, a transparent exchange of usage metadata is currently not always a given. Niche players in the film and television industry (independent film producers), but not the major studios we spoke to, raised concerns as well. In film production, interoperability issues in rights management systems are again mentioned as pain points and drivers for inefficient licensing.
- In the publishing industries, licensing appears to be relatively unproblematic, although stakeholders' comments diverge. Current licensing processes and systems may, however, lead to an under-exploitation of micro-licensing opportunities.

2.5.3. Challenges concerning payment processes

In essence, the following drivers and related pain points concerning metadata and payment issues emerge from our interviews and subsequent analysis:

¹³⁰ See <https://iptc.org/standards/photo-metadata/quick-guide-to-iptc-photo-metadata-and-google-images/>

¹³¹ See <https://ceplic.org/>

¹³² See <https://iptc.org/news/googles-licensable-images-feature-is-live/>

- **Metadata-related inaccuracy: Driver:** instances of a lack of reliable, exhaustive, and up-to-date metadata in the upstream journey of the remuneration. **Pain point:** inaccurate payments – “inaccurate” in the sense of “not as contractually agreed”, not in the sense of “fair, appropriate, proportional, and transparent” – was the first issue mentioned by songwriters, the authors of the British reports mentioned in section 2.5.2.2, and the technology provider developing systems to solve that issue. Another example is the inability to identify the rightholder, and consequently to pay the rightholder directly¹³³.
- **Metadata-related slowness: Driver:** instances of low interoperability or automation between payment systems at any point of the upstream sequence of many transfers. **Pain point:** slow payments and high costs of remuneration.

For the sake of completeness, and again looking exclusively at the payment mechanisms, the transparency obligation introduced by article 19 of the DSM Directive on copyright¹³⁴ is introducing requirements on **usage metadata** and data management:

*“Member States shall ensure that authors and performers receive on a regular basis, at least once a year, and taking into account the specificities of each sector, **up to date, relevant and comprehensive information** on the exploitation of their works and performances from the parties to whom they have licensed or transferred their rights, or their successors in title, in particular **as regards modes of exploitation, all revenues generated and remuneration due**”.*

From a data management point of view, it means that **more metadata** will need to be reliable, exhaustive, and up to date, and that **more reconciliations** will need to be performed between business models (modes of exploitation), monetisation (revenues generated) and remuneration at the top of the value chain.

2.5.3.1. In the film & television industry

Several pain points were addressed in our interviews with industry stakeholders in the film and television industry. We list the following issues for which rights metadata can (potentially) play a role:

- Film & television stakeholders – producers, financiers, sales agents¹³⁵, talent – may not know what rights are available, where the content is licensed, how much they are owed and when; participation management is complex.
- Payments may be delayed for years or never reach the right owners, especially at the beginning of the value chain.
- Companies may struggle with manual, opaque, error-prone reporting and payment systems, and high overheads, a situation which could potentially be improved by better rights metadata

¹³³ This is sometimes referred to as the royalty black box.

¹³⁴ Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, O.J. L 130, 17 May 2019, p. 121

¹³⁵ Sales agents are in principle curators and gatekeepers who drive up the value of the creative goods, but they are bottlenecks in effective rights management as curation and data management functions do not map, especially in the digital age. See R. Smits, *Gatekeeping in the Evolving Business of Independent Film Distribution*, Palgrave, 2019.

- Participants are sometimes paid in cheques and cannot verify their entitlements, as mentioned during the workshop organised by the study team on 24 June 2021. Again, improving the link between rights metadata and payments could help make payment issues more transparent for all stakeholders.

2.5.3.2. In the music industry

Deals between the labels and the services do not normally provide any licences for the separate song rights contained in the tracks that the labels provide to the service. These licences are issued by CMOs and/or music publishers (depending on repertoire), who need to be separately reported to and paid. This is where metadata issues arise, increasing the complexity of licensing, the administrative burden of the CMOs and publishers, and the efficiency and accuracy of getting authors paid.

The recent UK parliamentary enquiry on the economics of music streaming¹³⁶ analysed remunerations on the whole value chain, highlighted two major clusters of issues and pointed out potential consequences. A few edited excerpts from the report are listed below (focusing on “issues”, therefore somewhat biased towards where improvements would be important in the music industry as compared to what actually already is “working” well):

Metadata issues: (1) cases of lacking metadata for the underlying song when record labels license a recording to streaming services, (2) lack of a minimum viable data standard to ensure that services provide data in a way that is usable and comparable across all services, (3) lack of a comprehensive musical works database and a registration portal so that rightholders can provide accurate copyright data to necessary stakeholders easily.

Related to metadata issues are the following **royalty chain issues:** (1) the licensing and royalty chains of song rights causes complexity to the system, (2) lack of royalty chain information to provide transparency to creators about how much money is flowing through the system and where problems are arising, (3) lack of practical alert systems to inform creators and representatives about data conflicts.

The above issues are as many **drivers** leading to the following **pain points**:

- **Challenges related to payment processes:** potentially, song rightholders are not always efficiently or accurately remunerated for their work, which can result in lower payments overall that can affect the commercial viability of a songwriter’s career or independent publisher’s business; notwithstanding any consideration about their fair share of revenue between recording and song rights, that consideration being outside the scope of this study, moreover
- **Slow payments:** Some payments may be slower than expected by rightholders or even not reach them. Mismatched, incomplete or missing metadata can result in delays to creator royalties. Even worse, this can result in payments being misallocated or otherwise consigned as unclaimed or non-attributable royalties to *black boxes*¹³⁷. Total black boxes on the song side – not just streaming black box – consisted of \$2.5 billion in unallocated income in 2019 alone (see the \$424 million

¹³⁶ House of Commons, Digital, Culture, Media and Sport Committee, Economics of music streaming, Second Report of Session 2021–22, available at:

<https://publications.parliament.uk/pa/cm5802/cmselect/cmcmuds/50/50.pdf>

¹³⁷ After a period of time, black boxes are then assigned pro-rata to streams that have been correctly identified, which is established in standard publishing agreements.

in historical unmatched royalties received by the US Mechanical Licensing Collective from Digital Service Providers in 2021)¹³⁸.

2.5.3.3. In the publishing industry (books, press, journals, and images)

The interviewees from the publishing industry did not report pain points related to payments, except representatives from the image sector who mentioned the frequent absence of any payment due to the reasons stated in the sections hereabove or hereunder.

Pain point 3: Challenges concerning payment processes

- Complex and confidential licensing arrangements, relying on error-prone metadata, seem to drive sometimes inaccurate payments (“inaccurate” in the sense of “not as contractually agreed”, not in the sense of “fair, appropriate, proportional, and transparent”). This was noted specifically in the music and film industries.

2.5.4. Issues related to misappropriation and other rights infringements

2.5.4.1. Characterisation

In the discussions on this challenge, different topics or layers of the problem are frequently grouped together, even though the discussions would often benefit from a more nuanced view. Interestingly, interviewed rightholders in this study unanimously categorise misappropriation as follows:

- **Plagiarism**, typically presenting someone else's work or ideas as one's own, with or without their consent, by incorporating them into one's work without full attribution;
- **Sharing uncleared¹³⁹ third-party content**, typically communicating to the public (a) user generated content without clearing the rights of copyright-protected materials recycled in the user's own production, or (b) third-party content on a content sharing platform without appropriate licence,
- **Piracy** - illegally reproducing or disseminating copyrighted material - typically distributing commercially large quantities of content stolen from leakages or recaptured¹⁴⁰,
- **Upcoming issues**, such as deep fakes using a form of artificial intelligence called deep learning to make images of fake events.

¹³⁸ “The Mechanical Licensing Collective Receives \$424 Million in Historical Unmatched Royalties from Digital Service Providers”, *The MLC*, 2021, <https://www.themlc.com/press/mechanical-licensing-collective-receives-424-million-historical-unmatched-royalties-digital>

¹³⁹ “Clearing” rights means verifying that permission or licence is agreed upon for all included material before sharing the content.

¹⁴⁰ See R. Lobato's work on piracy as distribution, a.o. *The six faces of piracy: global media distribution from below*, Greenwood Publishing Group, 2008

In any case, a consequence of all four types of misappropriation in the cultural industries is lost revenue. Misappropriation prevents the creative industries from recouping their investments and generating the free cash flow necessary to support their operations and maintain their capital assets.

Metadata aspects play a **crucial role** in the context of **misappropriation and copyright infringements**. Implicated metadata includes **identification metadata, RMI, administrative metadata**, and additional metadata necessary to **track and trace the provenance of the infringing content**, starting from the original distribution channel to the final display or stream.

Interestingly, most metadata and techniques used to track and trace infringing content can be used to fight misinformation and add a layer of verifiable trust to all types of digital content through provenance and attribution solutions. Starting with photos and videos, the content authenticity initiative is creating for example a secure end-to-end system for digital content provenance through open-source development, cross-industry collaboration, and interoperability of tools¹⁴¹.

Misappropriation and related metadata are **explicitly or implicitly addressed** in the **acquis communautaire**¹⁴² (see section 2.2.2).

2.5.4.2. In the film and television industry

In the film and television industry, piracy often relates to (i) sharing uncleared third-party content on so called clip platforms (OCSSPs), (ii) piracy linked to downloadable, streamed or torrented¹⁴³, or scripted or unscripted, content, and (iii) piracy related to linear live streaming, typically sport competitions.

Piracy in the film and TV industry is not a new trend. However, there are reports that film piracy has been skyrocketing as people stayed home during the recent Covid19-pandemic. New data released by Muso¹⁴⁴ reveals significant increases in film, TV, and publishing piracy during 2020.¹⁴⁵ Unfortunately, it therefore seems that the declining trend in piracy that Muso revealed in the EUIPO Observatory report on *Online Copyright Infringement in the European Union - Music, Films and TV (2017-2018), Trends and drivers*¹⁴⁶ has been reversed.

¹⁴¹ See Content Authenticity Initiative, contentauthenticity.org, retrieved in May 2021.

¹⁴² Here, specifically, the Directives on rental and lending rights (Council Directive 92/100/EEC), broadcasting (Council Directive 93/83/EEC), the information society (Directive 2001/29/EC), collective management (Directive 2014/26/EU), online transmissions (Directive (EU) 2019/789) and the digital single market (Directive (EU) 2019/790).

¹⁴³ 'Torrenting is the act of downloading and uploading files through the BitTorrent network. Instead of downloading files to a central server, torrenting involves downloading files from other users' devices on the network. Conversely, users upload files from their own devices for other users to download'; Bischoff, P. 'What is torrenting? Is it safe? Is it illegal? Are you likely to be caught?', 8 April 2021 (<https://www.comparitech.com/blog/vpn-privacy/is-torrenting-safe-illegal-will-you-be-caught/>).

¹⁴⁴ See Muso, 'Piracy in 2020' (<https://www.muso.com/magazine/piracy-in-2020-a-snapshot-view>; accessed May 2021).

¹⁴⁵ There were 130.5 billion visits to piracy sites, 39 visits per person connected to the internet and 65.5 billion visits for TV content; there was a 13 % increase in publishing piracy sites between the first and the second semester; and illegal streaming made up 93 % of TV piracy and 52 % of film piracy.

¹⁴⁶ UIPO Observatory, *Online Copyright Infringement in the European Union - Music, films and TV (2017-*

It is interesting to note that – despite the high prominence of copyright infringement issues – the challenges are of varying relevance to different market participants. Addressing this pain point is a high priority especially for commercial film and television producers and broadcasters.

Practical issues when reacting to infringement also play a role. In the case of linear live streaming and pre-releases, infringement requires an immediate blocking order, which must be a blanket¹⁴⁷ mandate followed by an automated response. Generally, procedures of **notice and takedown**¹⁴⁸ or **notice takedown and stay down**¹⁴⁹ are applied. New approaches are tested; **licence and monetise**¹⁵⁰ and **knock and talk**¹⁵¹. The latter approach is used by the investigators of a commercial broadcaster and involves knocking on doors and talking with offenders for the purpose of gathering evidence, getting their cooperation, and keeping down their nodes, and more nodes, of infringing networks. The broadcaster launched the '**knock and talk**' procedure after having noticed that commercial confrontations with platforms or internet services providers or judicial processes against infringers brought only meagre results¹⁵².

Estimating the lost revenue is complex and not always possible. Subscription-based channels can compute reliable estimates by multiplying the number of infringers (the result of a thorough professional investigation) by the lost average revenue per user (ARPU). An internal Goldmedia analysis conducted on behalf of Vaunet in 2018 concluded that the commercial and fiscal losses caused by piracy amounted to EUR 700 million per annum – for the Germany TV broadcasters alone.

2.5.4.3. In the music industry

Similar to the film and TV industry, protecting works from being illegally distributed is a key priority for the music industry¹⁵³. Industry stakeholders are taking various actions to react to this industry pain point. These action – which illustrate the high priority of this pain point for the music industry – include the following:

- **Taking action on illegal music uploads.** Rightholders identify where music has been illegally uploaded online and notify the host to remove the infringing content (notice and takedown). The International Federation of the Phonographic Industry (IFPI) works in collaboration with record companies and its national groups around the world to implement a system for identifying where unlicensed music has been

2018) – *Trends and drivers*, EUIPO, Alicante, 2019 (https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/quantification-of-ipr-infringement/online-copyright-infringement-in-eu/online_copyright_infringement_in_eu_en.pdf).

¹⁴⁷ "A blanket license is a license that gives the licensee the right to perform all of the works in the repertoire for a single stated fee that does not vary depending on how much [content] from the repertoire the licensee actually uses", from https://itlaw.wikia.org/wiki/Blanket_license

¹⁴⁸ Content is removed by the host following notice – a process operated by online hosts in response to court orders or allegations that content is illegal.

¹⁴⁹ In addition, this requires that a service, after it has received a request to take down a certain copyrighted work, must also prevent the same work from becoming available on the service again in the future.

¹⁵⁰ Fully or partially automating a licensing and remuneration process for the rightholders' benefit

¹⁵¹ Asking the infringer to cooperate and not only to take down the infringing content but also to dismantle the infringing network.

¹⁵² See also the White Rabbit initiative supported by the European Commission at <https://www.whiterabbit.one/>

¹⁵³ These paragraphs are essentially excerpts from International Federation of the Phonographic Industry (IFPI), *Global Music Report 2017 – Annual state of the industry*, IFPI, London, 2017.

distributed without permission and calling for it to be removed¹⁵⁴ (or licensed and monetised by advertising if the service offers this ability). There has been a particular focus on illegal content stored on user upload services, and in 2016 alone 1.6 million videos and streams were reviewed, with more than 500,000 infringing pieces of content removed.

- **Addressing the threat of stream ripping.** Stream ripping is the process of ripping or creating a downloadable file from music that is available to stream online. It is typically done by users to produce an MP3 from a streamed music video, creating a file that can then be kept and listened to offline or on other devices. The process has become the most common way of illegally downloading music, with research conducted by Ipsos in 2016 finding that 30% of all internet users (and 49% of 16- to 24-year-olds) had engaged in the practice in the previous six months¹⁵⁵. Stream ripping sites compete unfairly with licensed music services, enabling users to permanently download music licensed only for ad-supported streaming and then listen to it offline without advertisements and without paying. Unlicensed stream ripping companies profit from the advertising space they sell on their sites and do not return any revenue to those who create or invest in the music they make available. In 2017, the most heavily used stream-ripping site, YouTube-MP3.org, estimated to have had more than 60 million unique users per month, was shut down.

In May 2021, websites were still promoting 'the list and comparison of the best YouTube to MP3 converter tools that will let you convert YouTube videos to MP3 by just entering the YouTube video URL and clicking a button as the command to convert the video. In a simple 2-3 steps, you will get the MP3 file'.

- **Protecting music releases.** Building marketing and publicity around a music release is a crucial part of the promotional work record labels deliver in collaboration with their artists. If music is leaked online ahead of its release date, the entire marketing campaign can be put in jeopardy and an artist's work can be completely undermined. Therefore, the IFPI runs a 24/7 site checking, alert and takedown service for members and national groups, monitoring all heavily used copyright-infringing sites and social networks. In 2016, 19.2 million URLs were identified as hosting infringing content and actioned for removal by the IFPI and national group programmes, and 339 million requests were sent to Google requiring it to *delist* infringing sites.

The amount of revenue lost by the music industry due to misappropriation can only be roughly **estimated**. Existing studies on this topic are often done by the music industry itself or on behalf of it and are all subject to substantial methodological challenges.

2.5.4.4. In books, journals, photography and press publishing

Books. Piracy in the publishing world is growing due to the use of digital means. In Italy, the loss due to book piracy amount to € 526 m, according to an IPSOS study promoted by the Italian Publishers Association¹⁵⁶. The Spanish "Observatorio de la piratería y hábitos de

¹⁵⁴ Some independent labels also use private companies for this kind of work, such as MUSO, see <https://www.muso.com/>

¹⁵⁵ PRS published more recent data – albeit only for the UK – demonstrating that this is still a problem, see: <https://completemusicupdate.com/article/stream-ripping-grew-1390-in-three-years-says-prs-for-music/>

¹⁵⁶ See *La pirateria del mondo del libro vale 528 milioni, danno per il sistema Paese di 1,3 miliardi*, *Giornale della Libreria*, 22 Jan 2020. <https://cutt.ly/hQyY3o7>

consumo de contenidos digitales” estimates 597 m illegal access to books on the Internet, with a market value of € 5.3 billion¹⁵⁷.

eBooks. The illegal consumption of e-books ranges from 21% of all e-book readers in Germany to 92% of the e-readers in Russia and China¹⁵⁸. EUR 250 m are lost annually in author income because of pirated book sales in the United States (The Authors Guild, 2019). According to Christina M. Rau¹⁵⁹, many legitimate websites in the book-selling business have a lot of problems with piracy. eBay has an increasing number of pirated PDFs. Pirates will download books, create PDFs, and then resell the PDFs on their own. Amazon also experiences issues. “eBook thieves” buy eBooks, strip the digital rights management code, return the book to Amazon, and then upload the file on their own to sell. The process can be complicated and tricky, but still, people with savvy and know-how are managing to do so to the detriment of authors. Additionally, there are websites essentially devoted to selling pirated works. Some sites will simply sell as many pirated books as they can. One such site claims to be a copyist church, a religion with the aim of copying content. A major problem is that many piracy sites are in jurisdictions in which litigation is difficult. Still, lawyers can send out takedown orders with a bit of hope attached.

Children’s books. Children books are read and shared on online content-sharing service platforms without always considering the rights of authors, illustrators, or publishers.

Cartoons and manga. The largest legitimate streamer of cartoons and manga listed misappropriation as the main pain point by far. Single illustrations, pages, or books are not only posted online as stand-alone content or embedded in user-generated content without proper attribution or licensing, but they are also captured from the legitimate distribution networks to be offered/sold on illegitimate websites.

Academic publishing. Again, one of the three major publishers of scientific, technical, and medical (STM) journals cited misappropriation as the number one issue. They were seconded by the other two major publishers and their trade body. Scientific, technical and medical publishers **reconquer** the moral ground, which illicit sites allege would have lost, by adapting their licensing agreements, opening parts of their catalogues, and facilitating citations. Nevertheless, piracy and plagiarism continue to flourish.

Press. The digital press suffers principally from their content being captured and republished as is or in condensed forms on websites from which they do not perceive royalties. Article 15 of the EU Directive on CDSM codifies (or confirms) their neighbouring right, but they are still very poorly equipped to exploit it, as their granular content severely lacks identification, attribution, and licensing conditions.

Pain point 4: Issues related to misappropriation and other rights infringements

- Protecting works from being illegitimately distributed is a key priority for all three industries covered in the study.

¹⁵⁷ See <http://lacoalicion.es/observatorio-de-la-pirateria/observatorio-de-la-pirateria-2019>.

¹⁵⁸ *Counterfeit and Piracy Watch List*, Commission Staff Working Document, Brussels, 14.12.2020 SWD(2020) 360 final, available at https://trade.ec.europa.eu/doclib/docs/2020/december/tradoc_159183.pdf

¹⁵⁹ Rau, C. M., ‘How to help combat piracy in publishing’, Book Riot, 25 July 2019 (<https://bookriot.com/piracy-in-publishing/>).

- Methods like stream-ripping (in the music industry) or converting eBooks to pdfs without DRM features are flourishing.
- Improved use of identification metadata, rights management information, administrative metadata and additional metadata necessary to track and trace the provenance of the infringing content could be part of an approach to address copyright infringement problems.

2.6. Empirical findings on challenges related to copyright metadata

One of the objectives of this study was to analyse the order of magnitude/quantify the (economic) impacts associated with challenges connected to copyright metadata. This quantification builds on the literature review (including secondary data collection) and the expert interviews (see sections 2.3 and 2.5). With those steps, we identified and discussed (potential) challenges in this area. This chapter subsequently aims at providing further empirical insights about the challenges (e.g., a potential lack of metadata) and quantify associated costs and missed opportunities as much as possible. Depending on the availability of secondary data as well as the response rates for our primary data collection (especially stakeholder surveys, see methodological remarks in section 2.1.4) this was possible to greater (or smaller) extent. Limitations due to data gaps are clearly flagged in the following in order to give transparency about the statements made and the empirical basis for them.

In the following sections, the key empirical results are presented industry by industry, taking up the categories used in section 2.5 on industry pain points.

2.6.1. Film & television industry

In this section, we present our main empirical findings in the areas of film production and distribution. We will also refer to specifics regarding the TV production and broadcasting sector. There are **commonalities** as regards the management of rights metadata between the film sector and the broadcasting sector, considering that the latter is one of the distribution channels for the former. However, there are also **different challenges** associated with rights metadata, for example when considering misappropriation, including piracy, which is a specifically large problem for the film industry. It is surely also relevant to TV productions but – owing to the type of content produced – to a somewhat smaller extent¹⁶⁰. We therefore clearly highlight below which findings refer to which sector while pointing out the findings that are common to the different sectors.

2.6.1.1. Overarching empirical findings

An overall contextual aspect to consider when analysing the situation regarding metadata in the film and broadcasting sectors is the fact that for different kinds of metadata (not necessarily rights metadata) there are important standards such as the Entertainment Identifier Registry Association (EIDR) or the European Broadcast Union's (EBU's) initiative EBUCore. These standards are tools for the industry to describe multimedia content in a rich and extensive way. For example, in EBUCore, aspects that can be described range from the basic title of the content to the camera type used to record the content (Tech 3349 – acquisition metadata).

Descriptive metadata (see 2.2.2.1 for a definition) are a highly important tool for the industry to identify content and especially specific details within a production. To realise this potential, works must be described in such a way that their production attributes are preserved and users can navigate to the content meeting their needs. A simple, hands-on exemplary use case for the need to efficiently identify a specific piece of content would be the task of programming a broadcaster's schedule. If there was no descriptive metadata at all, a person trying to assess the relevance of a specific video for its programme would

¹⁶⁰ For example, TV shows or sports streams are illegally made available on piracy platforms, but the problem seems to be larger for main blockbuster Hollywood productions.

need to screen this video in its entirety. On the other hand, descriptive metadata can help to reduce these efforts by providing detailed information about the content and its features. This simple example already shows the high value of and need for an adequate description of the content through descriptive metadata.

The main challenge with descriptive metadata in the film and broadcasting sectors does **not**, however, seem to be a **lack of metadata**, but rather the abundance of available data and the **compatibility between data records** (see also the pain points in section 2.5). Although there are – as mentioned before – important standards available, a respondent to the survey stated that “there are as many formats as there are players” (for a concrete description of the challenges see footnote 161). Also, existing standards seem not always to be used consistently, as anecdotal evidence in section 2.5.1 has shown. In support of this view, in a study on metadata and rights interoperability for content interchange between TV programme producers, Delgado et al noted that “in more or less detail, every content producer uses a self-defined set of fields to catalogue (non-standardised metadata)”.¹⁶¹ In addition and related to this, in an interview with a large private broadcasting company for this study it was said that “we surely need common metadata standards and common metadata taxonomy” (referring to descriptive metadata).

This pain point of the industry is already being addressed in different ways. Apart from the examples on standardisation efforts listed in Annex 5.3¹⁶² and here-above, there are entire Horizon 2020 projects such as the EUscreenXL project aimed at finding ways to make audiovisual content from broadcasters and audiovisual archives findable and accessible. The main objective of the project is to act as the pan-European aggregator for Europeana in the audiovisual domain, bringing together major European broadcasters, and to actively reach out to all European audiovisual archives. However, the example of EUscreenXL shows that – even though there are metadata collected – the focus is clearly on **descriptive metadata**; **rights metadata** seem to be somewhat of secondary importance. A finding from a publication related to EUscreenXL highlights this. This study¹⁶³ analysed which ¹⁶⁴metadata elements¹⁶⁵ were routinely used by the cataloguing systems of 20 TV content providers, with a particularly strong link to the public service broadcasting community.

At least three aspects are interesting to note.

- First, when asked which metadata standards were used for cataloguing audiovisual data, 11 out of 20 respondents stated that they use an in-house schema. This supports the quote above regarding the high heterogeneity of systems used.

¹⁶¹ In more detail, Delgado et al put it his way: “In more or less detail, every content producer uses a self-defined set of fields to catalogue (non-standardised metadata), which depend a lot on where content is stored. Those who just use stickers on the tapes or notebooks, only register the title and the date, while those who use databases, like MS-Access or My-SQL, and dedicated software, can easily afford to have twenty (or more) fields to describe each content. In between, we found some local channel storing metadata in MS-Excel files or MS-Word files”. (Delgado et al., “Metadata and Rights Interoperability for Content Interchange between TV Programs Producers”, 2006, pp. 267-278.)

¹⁶² The work on the IMZ Metadata Standard is also worth mentioning (<https://standards.imz.at/>; accessed June 2021).

¹⁶³ Rendina and Evain, ‘Publication of metadata schema for the pan-European aggregator’, 2013.

¹⁶⁴ The term “findability” refers to possibility for a human or machine to actually learn about or discover the existence of a specific work. A higher standardisation of descriptive metadata helps both humans or machines to use the “correct” search terms. It, consequently, raises findability of a work.

¹⁶⁵ Rendina and Evain also report awareness problems with regard to the concept of metadata: ‘We visited more than twenty local televisions of Catalonia, and, by means of the responses received to a questionnaire, we realised that some of them did not even know what metadata means, and of course, they did not use any standardised metadata schema.’

- Second, when asked which elements were routinely provided by the cataloguing system, different content-related metadata elements were covered. However, when it comes to rights metadata (metadata elements related to rights, terms and conditions, or IP rights restrictions) only around half of the respondents said that these elements were included.
- Third, metadata standardisation is primarily an effort of the television sector, one sees only smaller efforts in the large film sector.

Similar findings emerged from the small (therefore non-representative, only providing anecdotal evidence, n=5) survey carried out for this study among stakeholders in the film industry. One survey participant from a film production company for TV, cinema and commercials claimed that “especially for commercial films the copyright issues [attributions] for director, artistic collaborators, technical crew, narrators, music composers and actors are practically non-existent.”¹⁶⁶ Even if this opinion might be an extreme one, the variety of different contributors mentioned in this quote already highlights one other major challenge: Similar to the situation in the music industry covered above, the high number of different contributors can lead to a **fragmentation of rightholders**. Different parties are involved and rights are consequently at least potentially split between various rightholders. In consequence, it is challenging to record and manage all different contributions and related rights.

Against the background of the statements made above, it needs to be acknowledged that developments have taken place in the recent past in the audio-visual sector. One of these developments is the launch of the dual registration service by the Entertainment ID Registry Association (EIDR) and the International Standard Audiovisual Number International Agency (ISAN-IA) (see also the Annex to section 2.3 for further initiatives). This cooperation aims at enabling the production and distribution of content, facilitate its exchange, and enhance its discoverability. The involved parties have worked together to implement a fully interoperable system that allows content creators, owners and distributors to obtain identifiers from both organisations with a single application. However, the registration focuses mainly on the identification of the work itself and other descriptive metadata. The rightholders are only covered in optional entry fields of the registration form (except for the director or the work, see Dual Registration Form at <http://standard-ids.org/>, version 9.2. accessed on 1 December 2021).

In our survey, we also asked about the main impacts perceived by those respondents (n=3 for this specific question) who reported that they experienced rights metadata quality issues. The answers received suggest that the current situation implies several consequences for daily business. None of the respondents reported that they saw ‘no impact’. The impacts that were mainly selected as predominant (from only 2 out of the 3 respondents answering this question, two from an association, one from a production company) focused on the issues of:

- missing payments to rightholders and remuneration issues at the beginning of the value chain

¹⁶⁶ Furthermore, one survey contributor was of the opinion that - particularly in the archive area - even if there are rights metadata available, the “metadata of the film industry are incomplete and no longer correspond to reality with regard to copyright metadata (e.g., there are no longer some countries where the films were made).”

- low bargaining power due to the lack of transparency on the usage/consumption of works¹⁶⁷
- complex licensing processes.

Other aspects such as missed business opportunities due to metadata quality issues were only chosen by one survey respondent.

Due to the small number of participants in the survey, these findings alone can obviously not be taken for granted from a methodological point of view. However, as the next paragraphs will show, there is additional empirical material supporting the assumption that these aspects are indeed relevant pain points, especially for film producers and distributors (for the broadcasting sector this cannot be confirmed based on the evidence available).

2.6.1.2. Costs of rights management due to metadata challenges

Summarising the previous section, we see two **main findings** regarding metadata in the film production, but also in the broadcasting sector: there is an **abundance**, but – despite important standards – also a **high heterogeneity** in the way description metadata are used. On the other hand, it seems that rights management information (rights metadata) per se seems to be lacking at least in some cases and seems to be somewhat of a topic of secondary importance in metadata standards or metadata collection efforts by individual organisations (see also the example of the standard ids Registration Form mentioned above).

These two observations already imply that rights metadata management can be complex because different systems cannot easily interoperate with each other¹⁶⁸. This, in turn, translates into a high(er) level of **administration efforts** and therefore also **costs**. We tested this hypothesis (“lack of rights metadata leading to higher administrative costs”) for both the film production and distribution sector as well as for the broadcasting sector. The collected evidence underlines that a differentiated analysis is needed.

From the survey targeted at film producers and distributors, we can only take an **indication of additional administrative costs** which are related to challenges with rights metadata. We asked this question directly in the survey, giving the option to choose between different brackets/ranges. The survey results suggest that a low to medium double-digit number for the percentage of additional administrative costs (compared to a situation of costs with “perfect” rights metadata) could be a reasonable estimate.

Broadcasters and associations from the broadcasting industry were not able to provide estimates like the ones above. Unfortunately, there is also no secondary data available that could be used to estimate this figure in a sufficiently sound and reliable way.

2.6.1.3. Efficiency issues/ slow and inaccurate payments due to metadata challenges

One possible benefit of a high-quality rights metadata infrastructure is the potential to radically increase the efficiency of licensing processes: if rightholders can be identified in a transparent way, licensing flows and remuneration to the correct rightholders will be

¹⁶⁷ See also above the points raised by some music industry interviewees who stressed that transparent usage or consumption data is indeed a challenge.

¹⁶⁸ Again the cooperation between EIDR and ISAN-IA can be mentioned here. It exactly addresses the point of interoperability by setting up a single registration process used by both organisations. This will mean lower administration efforts and therefore also costs to stakeholders.

significantly improved. In fact, **challenges to identify the correct rightholders** are indeed often mentioned as a current pain point (see also the answers to overall challenges related to metadata from the survey among film industry stakeholders).

A working hypothesis of this study is that a sub-optimal rights metadata situation could lead to **inefficiencies** in the market, which in turn would affect the volume of licensing in a given industry and therefore also result in missed business opportunities. Since secondary data on this aspect is – to our knowledge – currently not available, we tested this using one of our questions in the survey to industry stakeholders. Specifically, we asked: “To what extent is the volume of licensing lower because of poor effectiveness or efficiency of licensing processes due to the lack of attached, interoperable or authoritative copyright metadata?” However, only one respondent was able to respond to this question, making it impossible to report conclusive estimates.

With regards to the **broadcasting** sector, the following quote illustrates the efficiency issue in a qualitative way: “Nowadays, the control of rights for selling digital content between TV content producers is done by paper contracts signed by the different parties. When the receiver TV or producer gets the content (either digital or a physical videotape), it is very difficult to know if the use they make of the content is the one defined in the contract. For instance, one could sell a programme to broadcast it once, and then broadcast it some more times.”¹⁶⁹ This again demonstrates that efficiency gains could be realised if rights metadata (and in fact usage metadata) were easily identifiable and could be used for licensing and remuneration processes. They could additionally help to prevent, as noted in the quote above, illegitimate use of content – an issue that will be covered further down in this section.

Furthermore, **transparency** is important when it comes to the distribution of revenues between different contributors or stakeholders of a film production. There are different dimensions to this challenge. One is the **speed** of the remuneration process; another is the accurate distribution of income. Currently, the majority of producers seem to have little data-backed understanding about viewership performance and how to improve the circulation or accessibility of their content.

Regarding the **speed of royalty payments**, it needs to be considered that payments to contributors or rightholders are highly dependent on figures generated at the ultimate point of sale. Collection agencies typically calculate a film’s revenues from numbers based on box office admissions and “calculate distributions to its creators using old-school spreadsheets. (...) Performance reports that show how well a film is doing are generated only every six months to a year”¹⁷⁰ Accordingly, payments to rightholders further up in the value chain have to face even longer delays until royalty payments can be expected. Especially during times when the liquidity of film producers might be strained (for example during a time of economic downturn with a large impact on the creative industries such as during the COVID-19 pandemic), this can be highly problematic from a business operations point of view.¹⁷¹

¹⁶⁹ Delgado et al., “Metadata and Rights Interoperability for Content Interchange between TV Programs Producers”, 2006, pp. 267-278.

¹⁷⁰ F. Khaliq, “Tackling the dark side of the movie business: the blockchain startup ensuring film workers get paid”, *The Guardian*, 2020, <https://www.theguardian.com/careers/2020/may/26/tackling-the-dark-side-of-the-movie-business-the-blockchain-startup-ensuring-film-workers-get-paid>

¹⁷¹ See for example a related quote from the German Film Producers Association: “Producers and financiers are the first to get on board a project but the last to see its revenues, because money takes a lot of time to travel from where it is generated to their pockets.”

2.6.1.4. Issues related to misappropriation and other rights infringements

Misappropriation – defined in this report as plagiarism, minor offences such as sharing third-party content on a platform without an appropriate licence and piracy (i.e. illegally reproducing or disseminating copyrighted material) – is a well-known phenomenon in both the film industry and the broadcasting industry. It happens in both sectors but through different channels and mechanisms – from peer-to-peer downloading of blockbusters in the film industry to illegal linear live streaming of football matches in the broadcasting sector.

This report does not repeat the vast number **impact estimations** from the many studies which have tried to quantify the extent and economic impact of piracy, for example in works by the Organisation for Economic Co-operation and Development (OECD). However, it seems clear that the economic impact is significant. A European Commission study by Herz and Kiljański suggests that “unpaid movie viewings” reduce movie sales in Europe by about 4.4%. Lost sales differ substantially by country: they are in the range of 1.65% in Germany and 10.4% in Spain.¹⁷² A US study reports higher values: according to Ma and Montgomery of Carnegie Mellon University, the elimination of piracy would boost box-office revenues by 14% per year.¹⁷³ The same order of magnitude was shown in a study by NERA in the US context. The study suggested that digital video piracy causes lost revenues for the US industry of at least \$29.2 billion and up to \$71.0 billion annually, representing a revenue reduction of between 11% and 24%.

As stated above, the broadcasters are facing different **piracy methods** because of to their different business model and even for different distribution channels within one company. For example, one large private European broadcaster that was interviewed for this study stressed that copyright management worked in the pay-tv ecosystem because the set-top boxes protect the content, but it did not work in the free-to-air domain.

A working hypothesis of our study was that there was an indirect link and (causal) chain between copyright metadata/rights management and misappropriation/piracy.¹⁷⁴ A lack of effective rights management can make licensing processes inefficient and therefore increase non-authorised use of copyrighted material. What is probably more important is the fact that rights enforcement is made more difficult in a situation of untransparent rights holdership information.¹⁷⁵

We directly asked respondents to our survey about their estimation of the impact of poor effectiveness and efficiency of licensing processes as well as about difficult rights enforcement and the level of piracy in the film industry. It proved, however, impossible to obtain sufficient number of responses to generate robust, quantified estimations. However, the answers we have suggest that piracy levels could indeed be lower if copyright enforcement was supported by a high-quality rights metadata situation. It seems reasonable to argue that piracy would not disappear with an improved copyright data framework, but that the detrimental effect of piracy could be at least reduced.

¹⁷² Herz and Kiljański, *Movie Piracy and Displaced Sales in Europe: Evidence from Six Countries*, 2016, see <http://dx.doi.org/10.2139/ssrn.2844167>

¹⁷³ L. Ma. et al., “The Dual Impact of Movie Piracy on Box-Office Revenue: Cannibalization and Promotion”, Carnegie Mellon University, 2016.

¹⁷⁴ See also the OECD report “Piracy of Digital Content” on the role of technology for piracy – both as a driver (distribution of pirated material over the Internet) as well as a barrier (Digital Rights Management)

¹⁷⁵ See the OECD report “Piracy of Digital Content” on the relationship between the (effectiveness of the) copyright regime and piracy levels.

There are therefore at least some indications (although no conclusive evidence per se) that a high-quality rights metadata infrastructure could contribute to lowering the levels of piracy in the film and TV broadcasting industry. In this context, a lot of attention is given to digital watermarking as well as content identification technologies to detect illegitimate use of copyright-protected work. However, it was also mentioned in several of the interviews conducted among broadcasters that digital watermarking is (i) still in its infancy (for example because it can be circumvented by re-encoding digital content) and (ii) needs to be supplemented with solutions that do not only satisfy OCSSPs' specifications, but also rightholders' requirements. An improved rights metadata infrastructure could be part of a solution.

2.6.2. Music industry

2.6.2.1. Overarching empirical findings

Copyright data management in the music industry faces a fundamental challenge: due to the nature of the creative process, the number of contributors to a single piece of music can be very high. When a new song is written and recorded, composers, lyricists, featured artists, session musicians, studio producers, sound engineers, record labels and music publishers may all contribute to a smaller or larger extent in the creation of the product, and may therefore have rights in or connected to the resulting track. This track – as mentioned above – may contain three distinct copyrights – in the lyrics and composition (the author rights) and the recording itself (the master rights).

If anything, this challenge has only increased as the number of people involved in the creation of an average track has increased over the years. For example, the average number of credited songwriters in the US market's top 10 streaming hits of 2018, per-track, was reported to be 9.1 – with a reported 21 people involved in the writing and production of one single track by Drake.¹⁷⁶ Considering the huge numbers of releases every week, and the potentially large number of recordings of the same song, the high complexity of the rightholder landscape becomes immediately clear.

Despite – or maybe because of – the fragmented situation of copyright in the music industry and the complex match between author and master rights, it appears that there is still a low level of **awareness of the importance of rights metadata** among individual creatives and artists in the industry. There are several organisations and initiatives working on changing this situation, trying to raise awareness for IP, with many CMOs and industry organisations running education programmes of one form or another.

Initiatives working on awareness raising among individual creators regarding IP in general and rights metadata issues in specific include the Music Rights Awareness Foundation, on its own and in cooperation with WIPO.¹⁷⁷ The United Kingdom's Intellectual Property Office also recently published a "beginner's guide to music copyright".¹⁷⁸ Additionally, there are manifold educational activities on authors' rights (including metadata aspects) by organisations such as GESAC or CISAC on the songwriter side, and the national CMOs, trade bodies and unions, and IMPALA on the

¹⁷⁶ T. Ingham, "How to have a streaming hit in the USA: Hire 9.1 songwriters (and a rap artist)", *Music Business Worldwide*, 2019, <https://www.musicbusinessworldwide.com/how-to-have-a-streaming-hit-in-the-us-hire-9-1-songwriters-and-a-rap-artist/>

¹⁷⁷ Music Rights Awareness Foundation, "WIPO for Creators", <http://musicrightsawareness.org/projects/wipo-for-creators/>

¹⁷⁸ Music Copyright Explained, "Access the Guide", <http://musiccopyrightexplained.com>

performer side. Awareness-raising activities are also important within the context of the DDEX initiative.

The objectives of the mentioned initiatives and similar endeavours are to raise awareness, and to improve concretely the management of metadata. One step forward, for example, would be if rights metadata properly were properly captured and recorded straight from the start in the music production process¹⁷⁹. According to interview partners for this study, this is still not a given, partly due to the growing number of smaller actors/producers/labels active in the field. The following quote illustrates this issue:

"As the main aim of the studio work is to create the best possible atmosphere for the best possible performances, producers do not always give their full attention to administrative tasks. In case of DIY indie labels the activities are often more hobby-like than professional by nature, and they might lack skills and knowledge on the music industry and its practises. For example, all are not even aware of the ISRC code".¹⁸⁰

As a logical consequence from this assessment, interview partners and the survey respondents repeatedly stressed the importance of "metadata education" in the music sector. One survey respondent from an association representing music writers put it this way:

"It is critical that we begin to rectify the industry's data issues by first engaging and educating our writers. We must ensure our songwriters and composers understand the concepts of data authority, the importance of globally recognised standards and identifiers."

It is worth noting that the importance of metadata education is THE single issue which was raised as highly important by all stakeholders in the music industry. While for some of the findings reported below, interpretations and assessments of their current relevance differed between stakeholders and interview partners, there is a very broad agreement on the importance of awareness and education related to rights metadata.

Apart from challenges regarding awareness of the importance of metadata, the following empirical findings hint at **existing challenges related to missing rights metadata** in the music industry.

Similarly to rights awareness, industry actors have acknowledged the importance of complete, high-quality rights metadata for a long time. There are several industry or standardisation initiatives, some selected examples are listed in the text box below (see also section 2.4 and specifically Annex 5.3 for a detailed overview)

The DDEX initiative's main mission is to improve data interoperability within the music industry. Its operating agreement also explicitly mentions a focus on metadata creation and standardisation, which are prerequisites for interoperability.

In addition, standardisation efforts in the different strands of the music industry (e.g., in the record industry ISRC and IPN, in the publishing sector ISWC and IPI) have contributed to raise the availability of rights metadata of different kinds by providing

¹⁷⁹ There are emerging solutions to simplify this process, see section 2.4.1.3 for the example of session.id

¹⁸⁰ J. Muikku, Finnish Music Publishers' Association et al., "Metadata of Digital Music Files: Summary", 2017.

clear and standardised formats and therefore providing orientation on how to define and store rights metadata.

However, there are indications of **ongoing challenges** with respect to the completeness of metadata. One piece of evidence is a study commissioned by the Finnish Music Publishers' Association, the Finnish Musicians' Union and other organisations, which found that for the "top 10% of reported tracks [in the statistics of NMP, a joint company owned by PRS for Music and the Nordic Copyright Bureau], which represent more than 90% of commercial value, composer/author data is still missing from more than 1/3 of the reported tracks."¹⁸¹

The information gathered through a small survey in the music industry for this study suggest a similar situation.¹⁸² We have asked respondents to what extent – in their opinion – rights metadata are missing in the music industry – using mostly qualitative scales, but also asking for quantitative estimations.¹⁸³ Qualitative results of the survey – in line with suggestions in interviews with industry experts as well as secondary sources – indicate that metadata could be missing¹⁸⁴ in the music sector to a "small to medium extent". The issue seems to be more pressing for catalogue works and songwriters rather than for new creations: for catalogue works respondents stated that rights metadata are missing to a "significant extent".¹⁸⁵

To provide a balanced and neutral picture of the empirical findings, it must be noted at this point that not all our interview partners in the music industry saw a lack of rights metadata as a predominant industry issue¹⁸⁶. For example, one association representing music labels stressed its view that it did not see large metadata issues and additionally believes that the music sector itself would be well set to solve remaining issues itself.

¹⁸¹ J. Muikku, Finnish Music Publishers' Association et al., "Metadata of Digital Music Files: Summary", 2017, p. 9.

¹⁸² It must be noted that the number of respondents to the survey carried out by the study team was low (for specific questions only 3 (!) responses). Results thus need to be interpreted with caution. However, the responses do include the views from different groups such as authors/songwriters, publishers and labels. They thus give indications of the extent of the issues, especially when findings are combined with secondary data from other sources. Furthermore, this data is – to the best of our knowledge – the first attempt to collect data on these issues at all.

¹⁸³ The decision to use qualitative scales was made based on survey technique considerations. Quantitative scales would have made it even more difficult for respondents to choose the "right" response. Indeed, we empirically saw high levels of missing data due to non-response in our surveys and will therefore not be able to report quantitative estimations for all of our survey questions. Qualitative answers have the downside of not delivering concrete numbers. However, they allow a "semi-quantification" of the issues which – in the research setting of this study – was deemed the most adequate approach.

¹⁸⁴ It is important to keep in mind that a respondent stating that some metadata for musical works might be missing, does not necessarily imply that she sees no metadata in use at all. It could well be the case that specific information is captured with relevant metadata, but that metadata is still not 100% complete and accurate.

¹⁸⁵ This statement is based on the indications from the survey, but is also consistent with the general notion that the rights metadata framework has seen substantial developments in the last years. It seems therefore clear that for older (catalogue) work, the metadata situation is different (and "worse") than the one for recent works. The same is mentioned in other studies like Music 2025 where it is stated that "...not all of the metadata on (...)back catalogues is accurate because the data was often lost in the creative process; sometimes, old tracks were recorded using now obsolete technologies making it almost impossible to retrieve all of the data."

¹⁸⁶ Indeed, the situation is worse for song rights than recording rights – mainly because co-ownership is so common with song rights, and rarer with recording rights. On the recordings side the issue is more likely to be performer data, which is a problem when it comes to Equitable Remuneration. Performer data is likely to be worse on catalogue because the industry only really started taking these revenue streams seriously in relatively recent history (since the 1990s).

The DDEX initiative and the RDx project to better connect CMO databases were cited in this context as good examples of this industry-driven approach to improve metadata and especially metadata interoperability in the sector. Furthermore, the focus – according to this stakeholder organisation – should be on making consistent use of the already existing metadata and implementing high-quality recognition systems to ensure effective tracking and monitoring of works (e.g., on platforms like TikTok and Facebook). Other stakeholder organisations representing authors underlined the need of higher transparency on usage of works by online services and stressed that in their view closer cooperation between online platforms and rightholders, wider use of industry standards and a more transparent verification/auditing process of their usage reports is needed.

We included a specific question in our survey, asking for quantitative estimations of the **percentage of new works¹⁸⁷ for which some metadata is missing or outdated**. However, due to the small number of respondents and – correspondingly – insufficient levels of certainty about concrete numbers, the results of this quantification attempts are not sufficiently reliable. They, however, broadly indicate, that the problem should not be neglected and are line with the qualitative statements mentioned above.

There are some – although also somewhat inconclusive – indications that the situation might be worse for catalogue works (see also section 2.5.1.3 for an example of a transaction of catalogue works for which only a very small share of songs had all metadata needed for exploitation). A reason for the potential difference between new works and catalogue works regarding the presence of metadata might be, that there has been some progress in awareness and use of rights metadata. Many important initiatives in this respect have been mentioned in this and previous sections of this report, some older, some more recent. It might be the case that metadata are more consistently taken care of for more recent works compared to the ones in the past. Catalogue works would, consequently, benefit from a retrospective update of their rights management information.

If it was indeed confirmed that metadata challenges in music are more pressing for catalogue works than new works, this would be notable for the following reason: despite the high growth in new releases in the music industry every week, approximately 80% of the music business still relies upon their catalogues, according to an interview with a major record label executive for the report Music 2025¹⁸⁸.

As described earlier in this report, there are different kinds of metadata – ranging from identification metadata to rights, descriptive and usage metadata. Therefore, one of the questions for the survey respondents was about the **kinds of metadata that are missing** in their view. Analysing the responses¹⁸⁹, it can be seen that data relating to creators and rightholders (including IPIs) or information on rightholders' splits seem to be missing more often. For work identifiers (ISWCs) and recording identifiers (ISRCs) this seems to be less the case. However, there is also contrasting evidence from an expert consulted during the study who had the opinion that "half of all musical works have no ISWC" – a somewhat counterintuitive finding given the fact that the ISWC is a leading music industry identifier

¹⁸⁷ Technically the term works relates to a song itself and therefore only relating to the "author rights" (see above). When we use the term "work" we mean the song (author) and recording (master) rights combined.

¹⁸⁸ F. Lyons et al., "Music 2025", Intellectual Property Office and Ulster University, 2019.

¹⁸⁹ Answers from 3 respondents from associations, including an association of indie labels and an association representing songwriters and composers.

and according to CISAC "tens of millions of ISWCs have been assigned worldwide since the code was created in 2002"¹⁹⁰

It is worth making a **distinction** in this context between **metadata** that is missing because it was **never issued or logged**, and **metadata** that is missing because it is **inaccessible to users of music**. For example, online music services will – according to interviewees – usually have all the metadata they will need about the recordings (ISRCs and other required data), but will not have metadata about musical works contained in each recording. That metadata (ISWCs and IPIs) exists, but it is a challenge for the online music service to connect each ISRC to an ISWC and the accompanying IPIs. This makes usage tracking of musical works more difficult.

Broadcasters may not have access to ISRC or ISWC information – again that metadata exists, but is not necessarily provided with music that is delivered to the broadcaster. The broadcaster will have track and song titles, but not the more accurate ISRCs and ISWCs.

After having assessed the general situation regarding the existence and availability of metadata, the following sections will go one step further and assess different potential **impacts**. As a first broad indication of impacts, responses to an overarching question on (potential) impacts can be cited. Respondents were asked qualitatively whether copyright metadata issues (accuracy, timeliness, interoperability etc..) affected their processes such as licensing, collection or distribution of revenues or other aspects. Only two participants from organisations representing independent record labels as well as music creators opted to answer this question, but both indicated that they see "**significant challenges**" (no other category was chosen). Most pressing mentioned issues relate – again – to the identification of the (correct) rightholders. The aspects "difficulties to identify the correct rightholder to be paid", "conflictive attribution / claims", "missed business opportunities" and "complex licensing" were most prominently mentioned. All these aspects translate into higher costs associated with rights management, which was already defined as one of the primary "pain points" in section 2.4. The next section analyses the cost aspect in more detail.

2.6.2.2. Costs of rights management due to metadata challenges

Based on the data just presented on the challenges regarding the availability of rights metadata in musical works, it can be argued that this is directly connected to **additional costs of rights management**. Concretely, this could mean that additional efforts on the side of songwriters, publishers, record studios and labels, and CMOs are needed to identify contributors to a work in order to arrange remuneration processes. High-quality metadata would facilitate this task, low quality metadata makes it more challenging and costly. In our survey, we tested the assumption that a sub-optimal rights metadata situation leads to higher administrative costs.

In essence, the survey results seem to imply that this problem is indeed relevant. Although concrete figures cannot be derived due to insufficient data availability, the additional costs seem to be non-neglectable.

What are the **drivers of these costs** apart from a (partial) lack of rights metadata attached to musical works? Part of the challenge is the (lack of) interoperability of different systems capturing rights metadata in the music industry. Of course, industry standards such as ISWC and ISRC, initiatives like DDEX and many other organisations are working

¹⁹⁰ See <https://www.cisac.org/Newsroom/news-releases/cisac-launches-major-project-upgrade-international-musical-work-identifier>

on standardisation from recording data and rights to sale, usage and reporting. However, research like Music 2025 points out that there is an “increasing fragmentation of datasets, along with increasing administration costs for data management and a duplication of data solutions built by individual organisations”.

The impressions collected through our survey support this view. For new creations, answers are somewhat mixed: one respondent answered that a lack of interoperability (e.g. between databases of rights) is an existing, but rather small issue for new works in the music industry. Two respondents saw it as a significant issue, however. When referring to catalogue works, the issue was uniformly seen as “significant”. The pattern seems to be recurring challenges related to rights metadata seem to be an issue and especially for catalogues and older works.

2.6.2.3. Efficiency issues (licensing) due to metadata challenges

Rights metadata are an important “fuel” for an efficient functioning of licensing processes: if rightholders are clearly and swiftly identifiable – and rights metadata support this process – **licensing processes can be faster, easier to process and less costly**. On the other hand, issues with metadata can affect the functioning of the market negatively and present a barrier to efficient markets. Strictly speaking, two different phenomena can be identified: (1) process inefficiencies in licensing arrangements which are currently actually taking place and (2) market inefficiencies caused by missed licensing opportunities as such (i.e. regarding licensing arrangements which would be facilitated by a better metadata situation but are currently not taking place – for example because of high costs, see previous section).

To explore the first issue, we directly asked the respondents to our survey to what extent they thought that a lack of attached, interoperable or authoritative rights metadata is currently causing *inefficient licensing processes*, i.e. making processes more complex, less accurate, slower, and more expensive. Answers from respondents were spread between the categories “to a small extent” and “to a significant extent”, suggesting that there is at least some concern regarding this and that the dimension of the problem might be relevant for various actors to a different degree. Indeed, this is a recurring scheme as – in contrast to the findings from the survey conducted during the course of the study – other stakeholders (associations representing labels and authors in the music industry) disagreed on the conclusion of efficiency losses due to missing metadata.

The (anecdotal) evidence on what efficiency losses mean in concrete terms in respect of – for example – lower volume of licensing transactions is extremely difficult to judge, even for the experts participating in our interviews and surveys. Anecdotal evidence suggests, however, that the problem exists at least to some extent.

Interestingly, there seems to be some agreement between respondents on the impact of missing rights metadata on **micro-licences**. With this term we referred to *small* licensing arrangements for example in an indie film or education context. Respondents fed back that micro-licensing was affected at least to a “medium-level extent”. The option “to a significant extent” was also chosen. This answering pattern makes sense, since for larger licensing transactions it might be economically feasible to invest large efforts in the licensing transactions. Micro-licences imply – by definition – lower volumes, making it even more important to keep the costs associated with the licensing deal low.

2.6.2.4. Challenges concerning payments processes

A major issue in the digital music industry is royalties that cannot be allocated directly to the correct rightholders. This is often referred to as “black box”. While data are not

available for the EU, the dimension of this issue seems substantial, at least considering the estimate for the US: in February 2021, it was announced by the Mechanical Licensing Collective that it had received \$424 million in historical **unmatched royalties** from Digital Service Providers¹⁹¹. Unmatched royalties are royalties earned that cannot be matched to a copyright owner by licensees or their licensing partners in the music industry (e.g. CMOs or music publishers). This happens when songs are not registered correctly or the contact information for the songwriter is unavailable. Additionally, even if the song was registered correctly, a licensee might still struggle to match the ISRC to the ISWC. In an ideal world, all this information would be available to industry players through rights metadata. In the current situation, this is obviously not the case – as can be seen from the unmatched royalties mentioned above.

Despite the case of the MLC described in the previous paragraph, in the interviews for this report the study team experienced dissenting views on this challenge. Music publishing associations were of the opinion that challenges related to royalty payment processes were only minor, especially thanks to the substantial developments in standards and IT systems facilitating these processes. One other interview partner for this study, an independent expert in the music sector, however estimated that 20 to 25% of music streaming revenue owed to songwriters might not be correctly allocated due to missing ISWCs, mismatch between ISRCs and ISWCs, or missing splits. According to him, inaccurate or incomplete data also matters when there are songs where only 80% of the copyright is claimed or, worse, 120% is claimed. In the first example 20% of the money ends up in “black box”. In the second, payments usually stop until the dispute is addressed.

There are several **implications** of challenges regarding matching works (or the corresponding royalties) to rightholders. One aspect is accuracy, another aspect is **speed**. In principle, it would be beneficial for individuals and the efficient functioning of the whole music market that a fast attribution and pay-out to rightholders would take place. In this context, it is important to see that the payment of unmatched royalties to the Mechanical Licensing Collective is only the beginning of a process to identify rightholders and to actually pay out royalties to them. The time needed to do so is substantial. This was also reported by survey participants. Respondents indicated that one effect of imperfect metadata leads to a “significantly slower” process of attributing authors’ or rightholders’ contributions to musical works.

Ultimately, issues of lacking and non-interoperable metadata result in making remuneration processes difficult. The previous example of unmatched royalties has already given an impression of the quantity of the challenge. The feedback from survey participants – although based on a small number of answers – supports the notion that the effect on remuneration processes is non-negligible.

Interesting indications can also be derived from a survey question on how **remuneration/compensation** is affected for different stakeholder groups in the music production process (authors/songwriters/composers; publishers; performers; record labels). The results, however based on a small number of data points, suggest that for the first group (authors/songwriters/composers) the effect is larger than for the other groups. Respondents agree that those stakeholders tend to experience a lower remuneration level

¹⁹¹ “The Mechanical Licensing Collective Receives \$424 Million in Historical Unmatched Royalties from Digital Service Providers”, *The MLC*, 2021, <https://www.themlc.com/press/mechanical-licensing-collective-receives-424-million-historical-unmatched-royalties-digital>

due to missing rights metadata. For the other groups (publishers, performers, labels), results are more ambiguous, but still suggest some effect.

An additional aspect was raised by one of the contributors to the survey. This point relates to the dynamics of the remuneration situation rather than an analysis of the status quo. The survey participant raised the point that “a lack of music writer recognition has led to a perception that the recording and performing artist and label carry more value. Hence music writers have a weak negotiation position when it comes to the value of per-stream rates for music writers.” According to this rationale, this could lead to a vicious circle for music writers and deteriorate their bargaining situation vis a vis other stakeholders.

2.6.2.5. Issues related to misappropriation and other rights infringements

Two final empirical findings can be highlighted when assessing the impacts of metadata challenges in the music industry. One is related to misappropriation, copyright infringement and enforcement; the other is related to the issue of diversity in the musical landscape.

From the theory-based impact model which was used as the foundation of the empirical work, we derived that missing metadata could also have an effect on copyright infringements and piracy of copyrighted works. The **rationale** behind this hypothesis was that (1) less transparent rights metadata are potentially a factor enabling the unauthorised use of copyrighted data¹⁹² and (2) less effective rightholder attribution could make it more difficult to enforce copyright and therefore increase the costs associated with enforcement actions.

Stakeholder answers tend to indicate that this might be indeed the case, although the evidence is mixed. On the one hand, in an initial question on the main perceived impacts of metadata quality issues, the item “piracy or difficulties to enforce copyrights” was not actively selected by respondents. On the other hand, when qualitatively exploring the question on whether rights enforcement is more complex and the number of disputes and litigations higher due to rights metadata challenges, respondents assessed this as an existing problem (one person stating that this is true to “a medium extent”, two persons stating “to a significant extent”). In concrete terms, this eventually translates into higher **enforcement costs**. If these additional costs are added to the figures on the loss of revenues reported by industry stakeholders due to misappropriation/copyright infringements, the potential benefits of better support of copyright enforcement by high-quality metadata become very clear.

Last but not least, there could be a link between metadata challenges and the diversity of the musical landscape. When remuneration of musicians or composers is biased (see above) or when attribution to the correct rightholder is not achieved, this could put smaller actors of the musical ecosystem at a disadvantage. The participants in the survey for this study agreed on this issue. All stated that – in their opinion – a lack of rights metadata would indeed lead to less diversity of creative works – even to a significant extent.

¹⁹² However, this needs not necessarily be the case as can be seen by the example of broadcasters’ content, which is clearly attributed to them, but still pirated.

2.6.3. Publishing industry (books, press, journals, and images)

2.6.3.1. Context of copyrights in publishing

Like the music industry, the publishing sector is **not a homogeneous sector**. It consists of a multitude of sub-sectors, with the most notable **distinction between books and news publishing**. These two subsectors differ considerably – as it becomes later also visible in the answers to the survey we conducted – in terms of the extent to which metadata challenges might be present. Typically, the book sector experiences less metadata challenges than the news publishing sector. However, even those major sub-sectors can be classified further into “sub-sub sectors” which again face meta-data challenges to different extents (with some sub-sectors sector claiming that they would not have metadata issues at all): there are, for example, considerable differences in the publishing sector between children’s books (which tend to have a lot of graphical illustrations requiring granular protection) or the practices in academic publishing (where metadata use is more pronounced than in other book publishing sectors).

We used, as for the music sector and the film sector, two types of questionnaires to gather empirical evidence: a shorter one with mostly open-ended questions, targeted at a broad industry audience (though it was also answered by specific copyright/metadata experts), and a more complex one targeted at an expert audience. The second survey also contained items aiming at quantifying the scope of possible metadata issues and their impacts in more detail. For both versions, it was made clear that all questions were to be seen in the digital context (e.g. online distribution, online news, etc.) of the publishing industry.

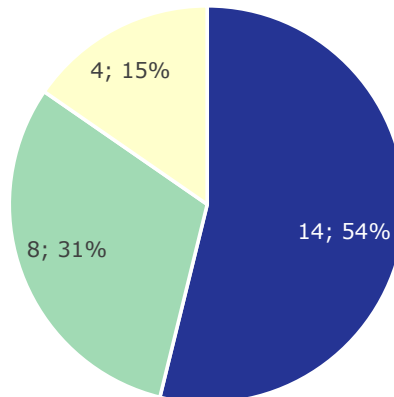
In comparison to the other industry sections (music, film/TV), **response rates** as well as absolute number of responses are higher. The shorter questionnaire had 30 respondents from the overall publishing sector. This allowed us – in contrast to the other sector analyses – to provide summary statistics for the survey results. In terms of sample characteristics, of the 30 responses, six (20%) were provided by individual creators, eight (around 27%) by associations, 12 (or 40%) by commercial companies (mostly SMEs) and four (13%) by “others”. At the level of lines of businesses, we obtained answers from authors (both self-publishing and using publishing houses), journalists, photographers and graphic designers (for book publishing as well as for news publishing); book publishing houses; news publishing houses; online platforms, and CMOs. This corresponds to a broad range of types of stakeholders active in the publishing sector.

Overall, and given the diversity of sub- and sub-sub-sectors in publishing, survey results should by no means considered statistically significant. Rather, the results are to be cautiously interpreted and explorative and indicative in nature.

2.6.3.2. Overarching empirical findings

The shorter questionnaire contained mostly open-ended questions, the first being if respondents were aware of metadata issues (see the following figure). Specifically, we asked whether respondents were aware of any copyright metadata quality issues (accuracy, completeness, interoperability, etc.) that affects processes such as acquiring rights on content, licensing (especially micro-licensing), collection or distribution of royalties/revenues or other activities. More than half of the respondents reported issues; 31% (or 8 respondents) reported no issues and for 4 respondents the answers were not clear as to whether they see or do not see issues.

In your creative sector, are you aware of any copyright metadata issues?



■ Copyright metadata issues observed ■ No Copyright metadata issues observed ■ Undetermined

Figure 2: Number and shares of respondents having identified metadata issues (publishing industry)

Source: Technopolis Group survey

Those who reported issues discussed them mostly in relation to **complex copyright environments**, but also commented on a series of specific issues that could arise in practice. Such specific issues are the lack of metadata standards for rights identification and **copyrights on images**. Those who reported negatively most frequently did not go into the details and just reported that they were not aware of any issues.

We asked the respondents a scaling question to qualitatively assess the extent to which the respondents believe metadata to be missing in the publishing industry. As can be seen in Figure 3, there is a peak of frequency of responses in the category "to a medium level extent" (11 responses) and "to a small extent" (six responses). Notable is also that eight respondents did not give an answer to the question or found the question not applicable.

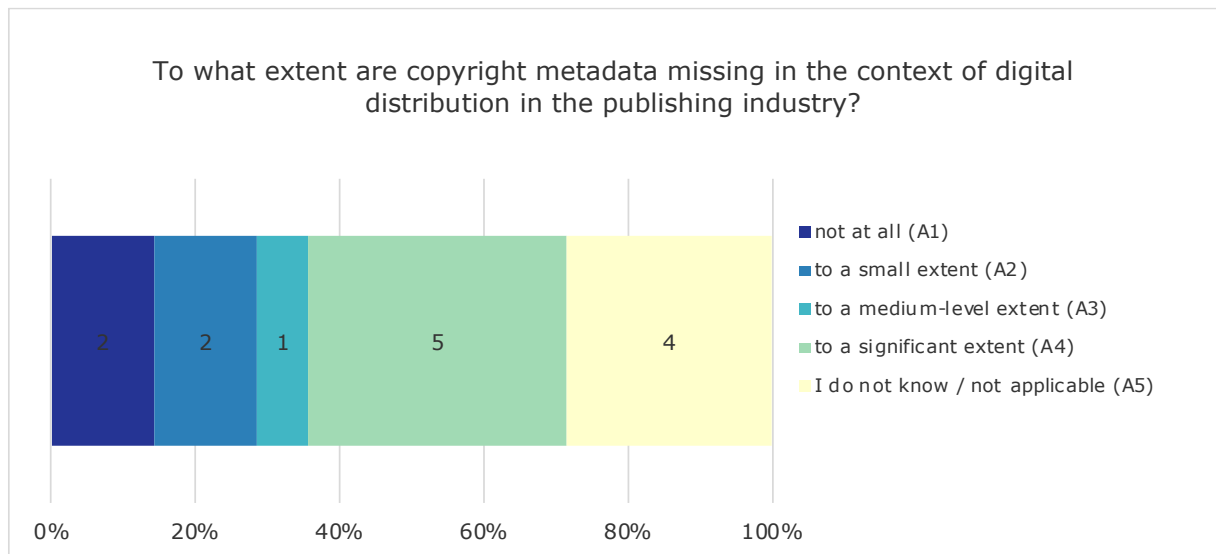


Figure 3: Estimated extent of missing metadata (publishing industry, qualitative estimations)

Source: Technopolis Group survey

The starting points for the long expert questionnaire were, similarly, assessments of the extent to which metadata is missing in the industry, whereby a distinction was made between books on the one side and newspapers and magazines on the other side (see Figure 4). As can be seen, the 14 respondents deemed metadata issues to occur a bit more likely with newspapers and magazines. The answers were more polarised in the book segment, hereby reflecting different experiences in sub-sub sectors.

The results here reflect feedback from stakeholders from these groups received by other means like interviews and/or statements during the expert workshop. Specifically, associations from the book publishing sector voiced the opinion to the study team that metadata issues are not problematic in their sector. However, the study team had the impression that these statements were specifically made for the traditional “physical” book publishing business and not the “digital world” which was explicitly the focus of this study.

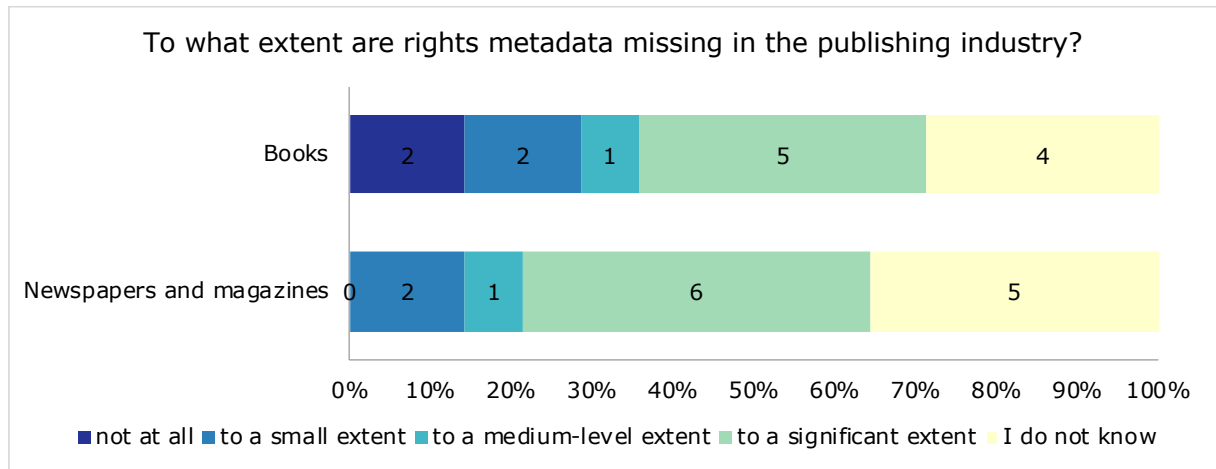


Figure 4: Estimated extent of missing metadata, by segments books and newspapers/magazines (qualitative estimations)

Source: Technopolis Group Survey, absolute number of responses

We asked respondents of the expert survey to quantify in more detail the extent of missing metadata (see Figure 5). From the figure, it can be read that one respondent opined that missing metadata can be observed for more than 10% to 50% of books, three respondents stated that metadata was missing for more than 50% of books. No respondent indicated that this would be the case for less than 10%.

The responses show that newspapers and magazines may face more problems in terms of missing metadata than books. However, the low number of responses has to be kept in mind.

For books, the results were more polarised in that there was a higher number of respondents who declared to not be able to provide an estimate.

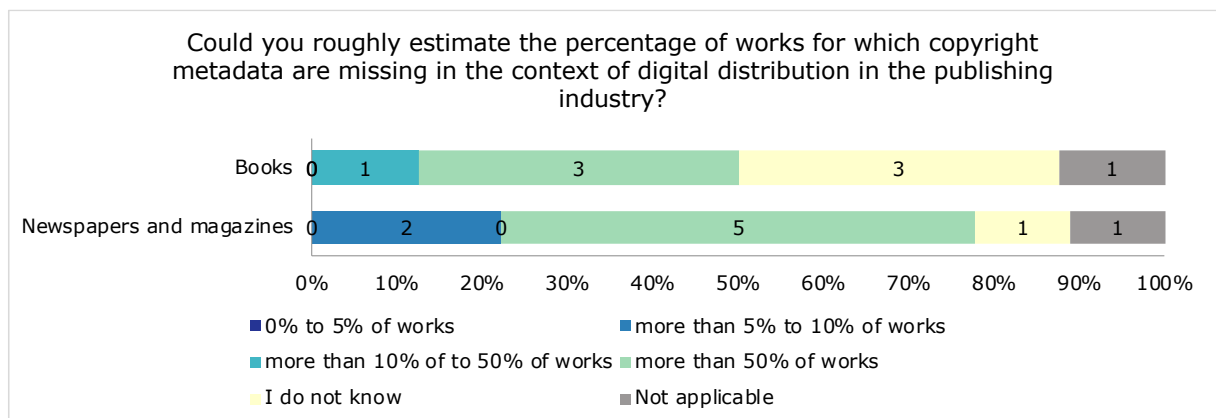


Figure 5: Estimated extent of missing metadata, by segments books and newspapers/magazines (percentages)

Source: Technopolis Group Survey, absolute number of responses

2.6.3.3. Costs of rights management due to metadata challenges

In line with the objectives of this study, we went one step further and asked the respondents of the expert questionnaire about quantified impacts (see Figure 6) – including

particularly different **types of costs** incurred due to metadata challenges. Only a few respondents were able to put a figure on higher costs due to more complex rights attribution, and those who did, had rather low estimates. By contrast, the aspects of loss of turnover and higher enforcement costs were assessed more rigidly, and with quite significant estimates. The majority of respondents saw turnover losses in the range of more than 10% to 50%.

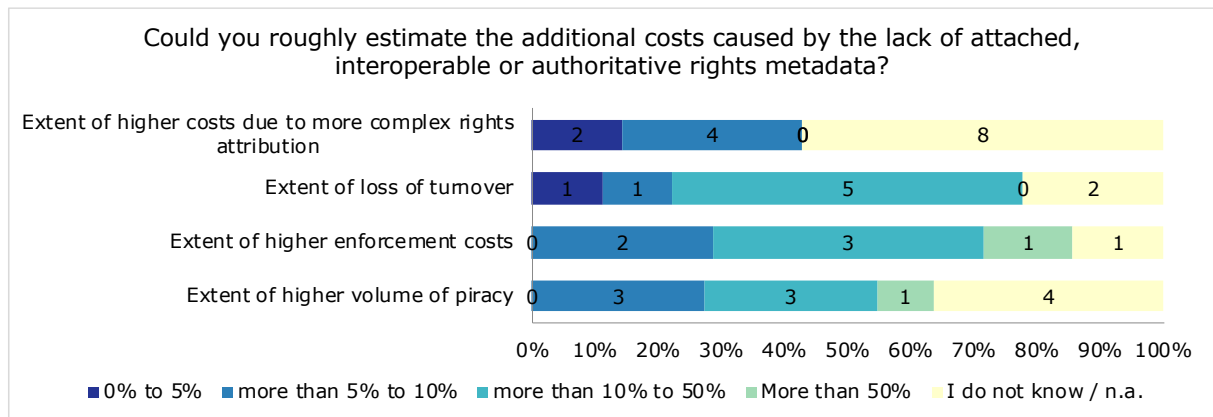


Figure 6: Quantitative estimations of impacts of missing metadata in the publishing sector

Source: Technopolis Group Survey, absolute number of responses

Unfortunately, we can contrast these findings only to some extent with **secondary data** as, to our knowledge, there are only very few other reports available on this topic. Most notable are the Nielsen reports, which did studies in the United Kingdom and the United States for books in 2012 and 2016.¹⁹³ Even they have the limitation that they focus on specific sets of metadata used for the trade of books and not rights metadata. They analyse very different types of metadata, but nevertheless give some interesting insights. The 2016 UK study combined bibliographic data (through its own book database) as well as sales and borrowing data (BookScan Total Consumer Market and LibScan library borrowing data) from 60,000 suppliers of bibliographic data, retail sales data (for over 90% of the UK book market) and library borrowing data from 70 UK public library authorities. The study looked at – among other aspects – the completeness of BIC-standard metadata per ISBN title and compared the sales per ISBN title at different levels of completeness of the information in the records. It found that around 15% of the ISBN title had incomplete BIC information. In addition, sales per ISBN number of titles with complete BIC and an image were considerably higher than sales where BIC information was incomplete (even if there was also an image present). Even though not focused on rights metadata, this shows that **missing metadata** can be an issue and a cause for significant cost (or foregone sales) in the publishing industry. It also suggest that the value of complete metadata and – to put it differently – the impact of missing metadata can be substantial.

Respondents of the short questionnaire were asked to provide an estimation of the extent of the impact. Around two thirds (20) of the respondents provided inputs, with several respondents pointing to the difficulties of giving estimates. Qualitative assessments were very diverse, ranging from indications on “minimal” to “enormous” losses due to metadata

¹⁹³ Walter, et al., “Nielsen Book UK Study: The Importance of Metadata for Discoverability and Sales”, 2016.

issues. Similarly, quantitative estimations were very heterogeneous and ranged from 1-2% loss of turnover to 10-15% loss of turnover.

Overall, the heterogeneity in answers relates to both (a) the lack of data respondents could rely on, and (b) the diverse nature of the publishing industry, e.g., its very different sub-sectors. It furthermore shows the overall methodological challenges in assessing metadata challenges as well the related impacts.

2.6.3.4. Efficiency issues due to metadata challenges

We asked the survey participants whether they experienced specific issues related to metadata issues pre-defined by the study team (see Figure 7). The feedback to this question can be interpreted as an indication of market inefficiencies – in case the issues are rated as challenging or problematic by market participants. Of 30 respondents, 11 said they did not experience any issues in relation to the specifically given copyright contexts. For the remaining 19, there was mostly not a clear ranking visible, as all aspects were experienced more or less to the same extent by the different respondents (between 6 and 7 answering choosing any of the categories listed, except for the category “antipiracy”, which was ticked off by nine respondents).

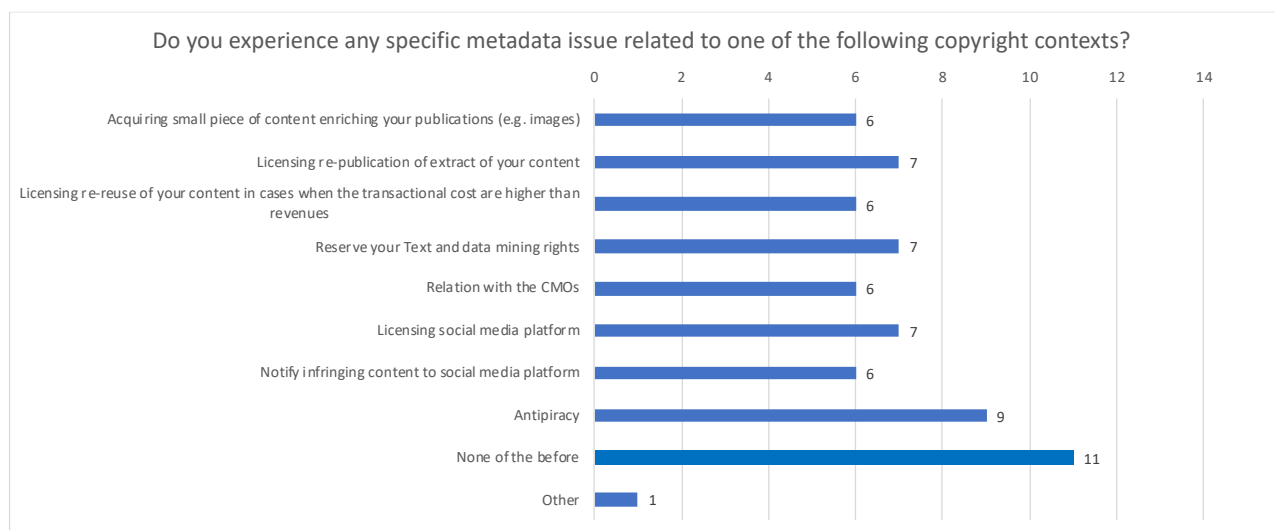


Figure 7: Metadata issues experienced by respondents (publishing industry)

Source: Technopolis Group survey, related to one of a selection of copyright contexts, multiple responses possible, absolute number of responses, n=30

The comments received for each of the answer categories of the above figure yielded a diverse picture:

- With respect to **antipiracy and enforcement**, it was mostly said that a lack of metadata would make it difficult to trace who is (rightfully) using copyrighted works. We will discuss the topic of antipiracy in a dedicated section further down.
- Regarding other aspects one respondent referred – in relation to **acquiring small pieces of content to enrich publications** – to a lack of consistent metadata that *“...makes it hard to pull together a picture of rights acquired in images for a whole work when images may be sourced from different vendors.”* In essence, this view supports the pain point of a lack of interoperability (or consistency) of metadata.

- For the aspect “**licensing re-publication of extract of content**”, one respondent said this would be almost impossible for complex products like videos (in news publishing), while another one indicted the cumbersome processes of manually checking the rights available, which would be prohibitively costly. Another one complained about the removal of metadata which would subsequently make right management tasks like identifying and contacting rightholders difficult.
- Regarding the aspect “**reserve your text and data mining rights**”, some respondents saw this as a potential new issue, raising new questions. Particularly with respect to this aspect, one could observe a call for more standards and standardisation activity, whereby one respondent called specifically for tech companies to observe such standards (beyond their own). One respondent specifically asked for a voluntary and distributed system that “...*should be separate from the actual rights management, to avoid lock-in effects, and business neutral, i.e. serving both direct and collective management, commercial and non-commercial licensing, subscription and pay-per-use models.*” For eBooks, it was mentioned: “*Authors do currently have the option to mark the eBooks metadata with an opt-out. There is also no register where authors can mark their works and ISBN, for instance, with opt-out (or opt-in). The rights reservation protocol is handled by publishers, but not harmonized and not always with the allowance to do so.*”
- In relation to the aspect “**Relation with CMOs**”, comments related to a lack of metadata regarding administering Reproduction Rights Organisations (RRO), to difficulties establishing opt-in and opt-out clauses in licensing, and to still manual entries of data for some CMOs, which could be automated using metadata.
- In relation to “**social media platforms**”, comments pointed again to the novelty of the issue which translates into a number of sub-issues: older contracts whose interpretation is difficult in the context of the new social media and to the generally too weak management of metadata in the specific context of social media licensing. These comments resonated also with those for the related category of “**notify infringing content to social media platform**”.

The following figure shows the extent to which lack of interoperability and lack of authoritative sources are deemed an issue in publishing (see Figure 8). It emerges that the aspect of “authoritative” sources was the one aspect where most respondents provided assessments. The results suggest that the lack of authoritative sources is a more pressing problem with out-of-print copies, while this issue is less visible with in-print copies.¹⁹⁴

The aspect of lack of interoperability showed a divided view, where a considerable number of respondents were not in a position to assess the scale of the issue; however, among those who did make an assessment, all (in books) were of the opinion that a lack of interoperability was a “significant” to “medium-level extent” issue. In newspapers and magazines, of those who had an opinion, five saw this also as “significant”, two as “medium-level” and only one as “small extent” issue.

¹⁹⁴ In qualitative interviews, the relevance of the concept of “authoritativeness of metadata” was questioned by some respondents from the book publishing industry. They argued that due to the nature of the business and often direct negotiations of publishers with authors, it would usually not be problematic to have reliable (authoritative) rights information or rights metadata. Again, this is likely to be true in traditional book publishing but might be very different in online environments and/or news publishing with much higher complexities in the content production process as well as much higher number of contributors at more granular levels (for example photographers providing a picture to an online text).

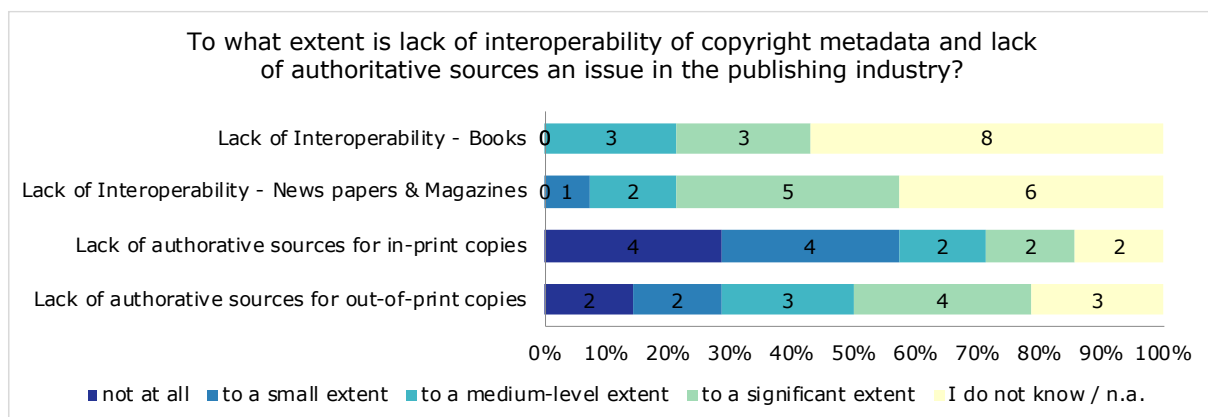


Figure 8: Extent to which lack of interoperability and lack of authoritative sources are perceived as an issue (qualitative estimations)

Source: Technopolis Group Survey, absolute number of responses

The polarised views on this question are also reflected in the comments to the questions. Some were quite articulate and detailed in the assessment of their issues. One comment referred to the plethora of licensing structures and conditions that would make licensing difficult – obviously, there is an expectation that better use of metadata would also improve this situation. The comments further identified three major issues: the lack of standards in relation to identifying rights; the use of proprietary standards that would make interoperability difficult; and, even if standards are available, the data is not entered or missing. As a result, licensing processes become overly complicated and require considerable manual work / intervention.

Then, we asked respondents to estimate in general terms, on a scale, the possible impacts of the perceived issues with metadata. The results are provided in the following Figure 9. One can see that the predominant impact concerns a lower volume of licensing, with 9 out of 14 respondents gauging the impact as happening to a “medium-level extent” or to a “significant extent”.

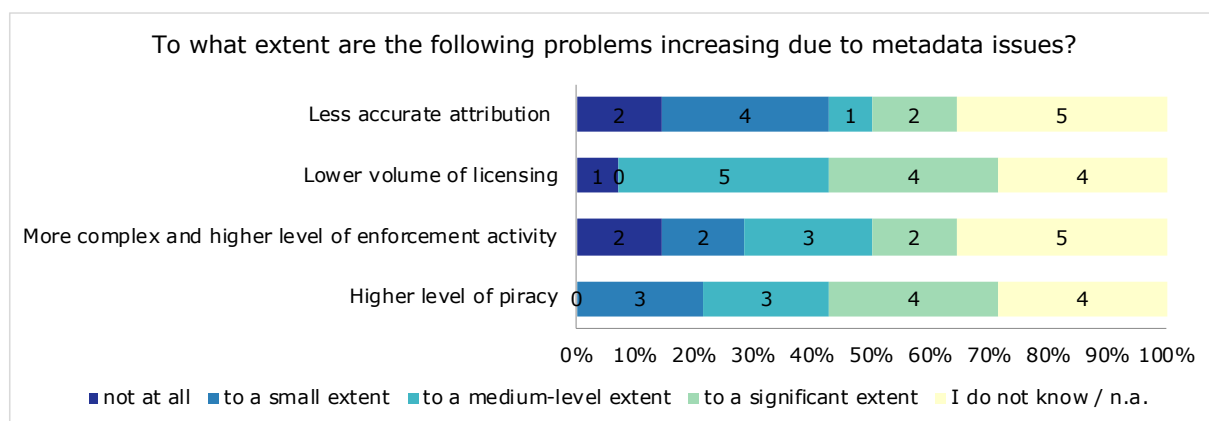


Figure 9: Estimation types impacts by respondents (qualitative estimations)

Source: Technopolis Group Survey, absolute number of responses

2.6.3.5. Challenges concerning payments processes

The expert survey enquired into different impact dimensions. For the respondents who experienced metadata quality issues, a comparatively prominent impact relates to **difficulties to enforce rights** (piracy), followed by **missing payments** to rightholders and the impacts of complex licensing, missed business opportunities and difficulties to identify the correct rightholders to be paid (see Figure 10). Conflicting attributions and delayed payments seem to be less of a concern. Following this reasoning, low-quality metadata seems to create fundamental problems with respect to identifying and paying rightholders at all, and to boost piracy activity.

The prominent role of piracy (as a form of misappropriation) echoes also the answers to the question in Figure 10, section 2.5.1 on metadata issues in relation to antipiracy. These answers therefore primarily refer to usage metering, which is relevant both in a context of authorised as well as unauthorised usage of copyrighted works. However, it was also said that piracy may not be an issue of metadata (as opposed to watermarking or fingerprinting¹⁹⁵), as wilful piracy acts would easily be able to remove metadata, making such data not a very effective tool in this context. On the other hand, though, the use of metadata might help in combatting acts of unauthorised use where metadata is not stripped of the content. It was said, for example, that *"...if there were full, accurate and protected metadata, obviously combating unauthorised use would be easier."*

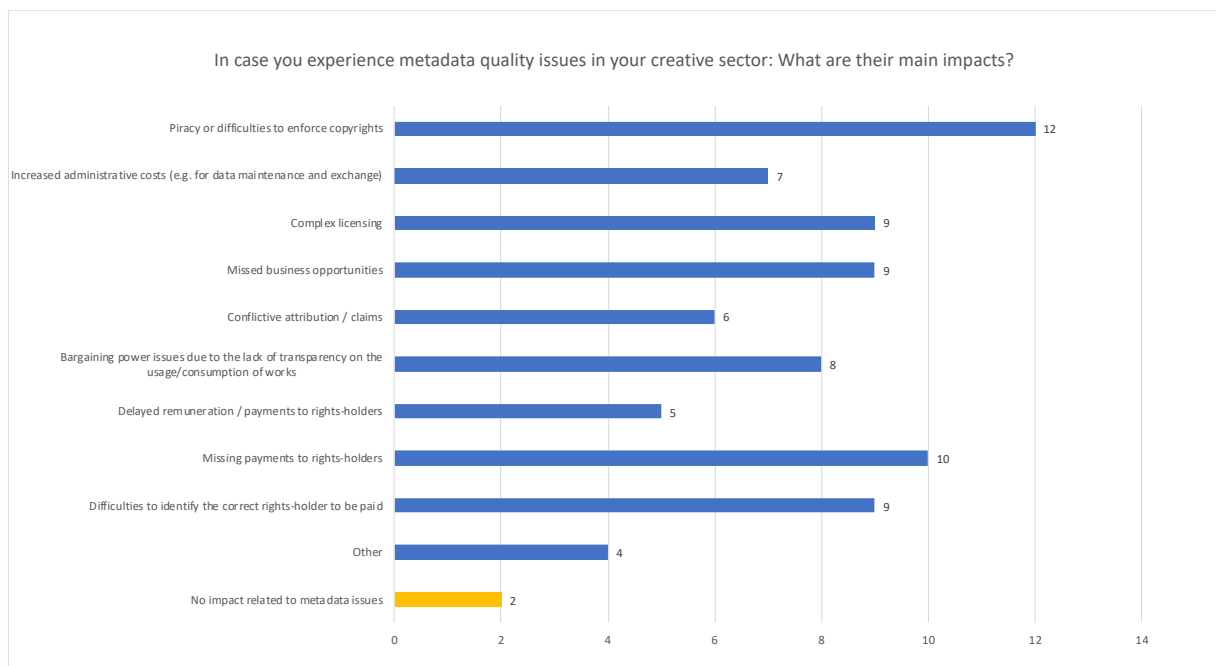


Figure 10: Impacts perceived due to metadata quality issues

Source: Technopolis Group survey, multiple responses possible, absolute number of responses

There were only a few comments made relating to the issues in Figure 10:

- One respondent, for example, saw no negative impact at the moment, but anticipated possible future problems as *"...future successful work in this area could depend on ensuring that enforcement tools on online platforms are interoperable*

¹⁹⁵ Not considered as metadata by the respondents.

with the standardised metadata publishers have been using for years." The same respondents warned against too much centralisation in a future copyright infrastructure. This *"...would increase cost and reduce efficiency, as it would create additional barriers, processes and complexity for publishers to manage the metadata of their own works, especially for the smallest publishers. It would also imply unnecessary duplication of information and increase the risk of inconsistencies among duplicated data sources, which in turn would create major issues for the publishing sector as a whole as information would become more likely to be unreliable."* Hence, this respondent saw mostly the need to work on standards and standardisation of metadata.

- A second comment was concerned with platforms, and public libraries and universities, offering flat rates for accessing content so *"...they avoid tracking, documentation and therefore fair remuneration."*

One can observe two opposing views, one being that metadata are unrelated to the impact dimensions as described, while the other taking the opposite view. The two following statements are in a sense "representative" of the heterogeneity of views.

"We consider that most of the issues raised here have little or nothing directly to do with the existence of rights metadata. Market failure (the cost of a transaction exceeding its value) is the whole raison d'être for CMOs. This is why CMOs' "permissions services" have been developed – because otherwise the costs to rightholders would often exceed the revenue (particularly as so many permissions are granted for nothing). The creation of such permissions services has been made possible by the development of new "rights metadata"."

"All of the categories above are significant. The lack of metadata and lack of good underlying contract storage – not just for main author agreements but for illustrators, contributors and permission licences means that many book publishers have no idea what rights they have in a complete work. (...) Publishing's long legacy and reluctance to digitise the back office means that there is a huge problem and gaps in contractual paperwork – even if they did digitise what they have and apply metadata to it – there would be huge holes where there is simply no information about rights ownership for various content elements that have been published."

Eventually, we asked respondents of both questionnaires to provide ideas for setting priorities in order to improve copyright metadata in the publishing industry. The largest cluster of answers (eight out of a total of 21 responses) centred around **improved standards and standardisation activities** in the metadata domain, whereby some added, for example, that one should *"...ensure that dominant platforms do not set the standards and rules"*. One respondent underlined the responsibility of some internet platforms in removing RMI and forcing the use of some standards more beneficial to their own purpose.

The second largest cluster was in relation to **raising awareness**. It was often said that awareness on this issue with many stakeholders is low as reflected for example, in the comment that *"...the topic is currently not widely known."* This may be also the reason why some respondents explicitly stated to not have any suggestions, such as reflected in the following comment: *"I do not have any suggestions to offer, we are still learning and adapting. We are in fact looking for more information and input into how we can manage better."*

Further comments were made in relation to the possibility to track publications and in compliance with existing laws (e.g., the territoriality principle).

2.6.3.6. Discussion of survey findings

All in all, the analysis of the survey results for the publishing sector shows a very diverse range of views as regards the extent of possible metadata issues. This reflects, on the one hand, the diversity of business models and market segments in the publishing industry. Hence, the aggregation of data to single figures or general industry-wide statements cannot be made.

On the other hand, and although the number of observations for the publishing was higher than for the two other sectors, the number of responses to the survey were still very low. To some extent, this problem can be offset for the cases when associations answered on behalf of their (many) members. Still, one immediate conclusion would be that for better data, the exercise should be tailored to the very different market segments of the publishing industry, using very specific questionnaires. This was outside the scope of the study.

Having said that, there are still considerable insights to be taken from the available data. One of these insights is that there is clearly a set of knowledgeable industry actors which do perceive rights metadata issues, at the very least for their specific market segments (and this is clearly more than one market segment). One other particular issue is the diversity of different stakeholders' awareness of the topic of rights metadata, which is clearly improvable.

In terms of actual issues, several aspects can be noted based on the responses to the survey. For example, there may be a problem of untapped potential to enrich the metadata. This means that metadata at overall product level – for trade purposes – seems to be often available. However, some comments from the periodicals indicate that this would only hold true “in principle”, i.e. the metadata structure is there, but actual data entries are missing or ignored by specific stakeholder groups. At a more granular level, i.e. copyright-protected works *within* a book/publication/online newspaper article, there seems to be much less metadata available (particularly as regards the actual rights held, i.e. metadata at rights level).

Certainly, an issue are images/photographs, which repeatedly come up when discussing missing metadata and/or identifiers.¹⁹⁶ A number of comments suggest to working more on standardisation and interoperable, open standards.

There is some hope that improved metadata can play an important role in fighting piracy. However, the question remains as to the extent to which (wilful) piracy is actually a metadata problem, since wilful pirates can remove any metadata easily. A related problem is the traceability of who uses the product, an issue exacerbated in publishing because it is rather easy to “cut, copy, and paste” content. Metadata can be seen here as one necessary, facilitating element but not as a sufficient means against wilful pirates. As long as (and if) piracy is done only because alternative ways of identifying rightholders and their rights is too cumbersome, metadata could play a role (although one could argue that this would be minor cases).

Given the few available data points based on various degrees of opinions, hard data and educated estimates, one issue that seems to come through is that there does not seem to be a precise common vocabulary of what metadata is and even further different perceptions of what a copyright infrastructure constitutes.

¹⁹⁶ See also a report by Imatag reporting that 97% images in news services lack metadata (<https://blog.imatag.com/state-of-image-metadata-in-news-sites-2019-update>)

2.6.4. Summary of the empirical findings in the different industries

Summarising and comparing the insights from the empirical analyses in the different creative industries covered in this study, the following main observations and conclusions on metadata pain points and challenges emerge:

- Firstly, the need for higher awareness as well as informational and training measures on the topic of metadata was mentioned in all of the creative industries covered. There is a broad consensus among different stakeholder groups on this aspect. Consequently, this aspect will likely need to receive particular attention in the future (see section 2.7 of this report on suggestions for future avenues).
- Additionally, the results indicate that a lack of quality of at least some rights metadata elements could be specifically relevant in the music industry. This is due to complex rights ownership structures, the nature of the creative process, (co-authorships, various contributors...), insufficient awareness of the importance of metadata among creators or small industry actors (see above) as well as other factors. The actual size of this problem is assessed differently by different stakeholder groups – with recording company associations assessing the challenge as less pressing than creators and independent experts interviewed for this study. There is some more consensus on the notion that at least some metadata is missing or could be outdated for back catalogues and older works. It seems – based on our results – also not negligible for the publishing, although this is most likely more problematic for (online) newspaper publishing rather than book publishing.
- Problems of compatibility of descriptive as well as rights management data was identified in the empirical work in the film industry – despite recent developments to increase compatibility through a harmonisation of the registration process between EIDR and ISAN. However, it seems still also problematic in the music industry (despite ongoing developments like standardisation activities), apparently not as much in the publishing industry.
- Concrete additional costs due to higher administrative and management efforts caused by suboptimal metadata are difficult to quantify, but based on qualitative evidence from interviews with recording labels, associations representing creators and independent experts obtained for this study seem to be non-neglectable, at least in the music industry. There are also indications that this might be problematic in the film industry. This correlates with the degree of fragmentation observed in the rightholder landscape in these two industries (e.g. co-authorship of songs or creative works in general).
- Another similarity between the music and film industry seems to be the pain point of a lack of speed of royalty payments. This aspect can be seen as one of the main inefficiencies with respect to licensing and royalty payment processes in these industries.
- Some music publishers and film producers also point to a lack of transparency on the usage/consumption of works. Again, in these two sectors this seems to be relevant, in the publishing industry this was not a primary concern – due to different industry mechanisms and processes as well as distribution and business models (e.g. streaming in music and film).

2.7. Avenues for future action - strengthening the momentum of ongoing initiatives addressing rights metadata issues

We have seen that many promising initiatives co-exist in the various analysed sectors (chapter 2.4). However, we have also seen that specific aspects of copyright metadata management remain challenging (chapter 2.5 and 2.6), at least for some stakeholders and in some sectors. Against this context, it is important to acknowledge the achievements reached by various rights metadata-related initiatives in the different creative industries, to “keep their momentum” and to efficiently build on them. This way we can strengthen the strengths of the current rights management framework while also improving on weaknesses.

Instead of delving into recommendations for specific actions for specific actors in specific industries, in the following we sketch some general thoughts of what steps could be reasonable to take in order to improve the situation regarding rights metadata and to strengthen the existing rights data system.

The proposed steps aim to make a contribution for metadata systems to be able fulfil legal requirements (i.e., purposes laid out in relevant European Directives such as the European Directive 2001/29/EC¹⁹⁷ on copyright in the information society), but also specific stakeholder needs. The latter were discussed as current “pain points” in chapter 2.5 of this report and relate to aspects such as awareness, costs and efficiency issues in managing copyright data (e.g., when managing royalty payment processes to rightsholders).

In order to alleviate these pain points, we have described the following requirements for metadata. As stated at various points of this report, rights metadata must be exhaustive, current, interoperable, trustworthy, and accessible:

- **Exhaustive metadata.** Metadata are necessary to remunerate creators in the digital era. In the analogue-only world, one could say, ‘no money, no content (for the consumer)’, in the digital era, one can say, ‘no data, no money (for the rightholder)’. Additionally, metadata are protected as ‘rights management information’ as defined in Article 7 of Directive 2001/29/EC. However, still some creators or smaller actors (e.g. composers, session musicians) do not consistently use metadata or are not aware of their importance and benefits. This can contribute to a lack of metadata. Further awareness raising activities and educational efforts are therefore still needed to avoid this situation.
- **Current metadata.** As mentioned before, the content value network is highly complex, dynamic, fluid, and transitional. Business models come and go. More and more professionals and prosumers create content distributed through more and more channels. RMI must not only be complete and error-free, but - permanently - updated.
- **Interoperable metadata.** When metadata is shared among multiple players in a sector or across industries, interoperability must be ensured. However, existing databases of rights still at times lack common technical standards (identifiers, certification, etc.), or the formats used are not interoperable. An example where this has been recognised and acted accordingly is the recent harmonisation of the registration process between EIDR and ISAN in the audio-visual sector. Other

¹⁹⁷ Article 7(2) of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, OJ L 167, 22.6.2001, p. 10.

important initiatives include the DDEX initiative in the music industry. However, cross-sectoral interoperability is still a challenge.

- **Accessible metadata.** Metadata must be accessible; this is a prerequisite for effective and efficient data management. The accessibility of usage metadata – for example – is a condition for transparency which is first necessary to restore a level playing field between rightholders and distributors and, second, essential for a fair remuneration of rightholders.
- **Authoritative metadata.** Databases containing information on rights and licensing terms and conditions are still fragmented (several sectors, territories, CMOs, online platforms, etc.). Market participants may not always have access to reliable authoritative data to obtain licences for the use of content. This lack of reliable authoritative data has a negative impact on rightholders, particularly SMEs, authors and performers, which miss revenue streams linked to their unidentified content.¹⁹⁸

Against the background of the manifold existing initiatives, it seems reasonable to assume that many **paths already taken** will be further followed in the future. In this sense, the current trajectories of stakeholders will most likely be continued. For example, this implies that the **European institutions** will pursue their ongoing regulatory and non-regulatory initiatives. In relation to copyright data, the EU has proposed new regulations on AI, digital services and markets, and data governance that would impact rights data management. On the non-regulatory side, boosted by the [Resilience and Recovery Fund](#), its digital priority and growing attention to culture, long-term initiatives will be launched, and finite projects funded with the aim to solve problems at the junction of media, copyright, and technology. Finally, EUIPO, specifically the EUIPO Observatory, will continue to play an important role, be it for awareness campaigns, specific databases such as orphan works, or portals such as out-of-commerce works.

In addition, the Intellectual Property Action Plan is an important EU initiative with may foster further improvement activities on copyright metadata.

There are also other important European initiatives to take into account. This includes, for example, the **European Data Strategy** (addressing, for example, EU-wide common, interoperable data spaces in strategic sectors to overcome legal and technical barriers to data sharing across organisations), new **regulations on Artificial Intelligence** with relevance for rights management (e.g. for machine learning to create, curate and clean rights management information), licensing and enforcement, as well as the **Digital Services Act and the Digital Market Act**. These initiatives may possibly impact or influence the sharing of copyright metadata between parties in the future.

Equally important are, of course, the manifold industry initiatives as well as research and innovation projects identified in this report (see Annex 5.3) – from DDEX in the music industry, to Ardito for text and images to EIDR in the audio-visual industry. The momentum built up by these initiatives in different areas needs to be kept and reinforced.

Overall, the following broad areas for future action would, in the opinion of the study authors, be useful to address further.

¹⁹⁸ The annex 5.5 details further characteristics of rights data management.

2.7.1. Raising general copyright awareness

The study (see for example the study section on the industry “pain points” related of misappropriation and other rights infringements, section 2.5.4) showed that creators and rightholders on one side, and users and consumers on the other side, are too often not aware of their rights or the rights of others. Many organisations try to remedy this issue. From our overview of ongoing initiatives (see chapter 2.4 and the corresponding annex), we can highlight, for example, the UK project *Music Copyright Explained* and add to that example:

- **CopyrightUser.eu**¹⁹⁹ aiming to scale up the successful **CopyrightUser.org**²⁰⁰ by building upon and extending the methodologies, features and models of collaboration between academics and creatives designed as part of the Copyright User initiative in the UK. Copyright knowledge needs of different creative and cultural sectors in the EU are identified from the bottom-up in collaboration with researchers and key stakeholders across the **reCreating Europe consortium**²⁰¹.
- **WIPO for Creators**²⁰², an open public–private partnership (PPP) launched by WIPO and the [Music Rights Awareness Foundation](#) aiming at raising awareness and increasing knowledge of creators’ rights and related management practices, ensuring recognition and fair reward for all creators regardless of their geographical, cultural or economic conditions. the PPP has already been joined²⁰³ by the International Confederation of Societies of Authors and Composers (CISAC), the Digital Data Exchange (DDEX), International Authors Forum (IAF), International Confederation of Music Publishers (ICMP), International Federation of the Phonographic Industry (IFPI), Independent Music Publishers International Forum (IMPF), International Publishers Association (IPA) and the Societies’ Council for the Collective Management of Performers’ Rights (SCAPR), most of them already mentioned in this report. WIPO also provides a public database of worldwide IP outreach initiatives²⁰⁴, including in Europe and for copyright.

Apart from awareness among creators/ rightholders, part of the challenge is also that users not always respect copyright to its full extent. Under the heading of “piracy” and other labels, this has been a topic for discussion for a long time, of course, and has become even more present in the digital/Internet era. Legal (e.g., the first HADOPI²⁰⁵ laws in France) and technical (e.g., Digital Rights Management Tools) measures were rapidly taken to stem the misappropriation of copyright-protected content – with limited success. Therefore, further education among users seems necessary and awareness must still be addressed²⁰⁶.

¹⁹⁹ See <https://zenodo.org/record/5070439#.YQmHYo5KiUI>

²⁰⁰ See <https://www.copyrightuser.org/>

²⁰¹ “reCreating Europe aims at bringing a ground-breaking contribution to the understanding and management of copyright in the DSM, and at advancing the discussion on how IPRs can be best regulated to facilitate access to, consumption of and generation of cultural and creative products”, see <https://www.recreating.eu/>

²⁰² See <https://www.wipo.int/wipoforcreators/en/>

²⁰³ See https://www.wipo.int/pressroom/en/articles/2021/article_0005.html

²⁰⁴ See <https://www.wipo.int/ip-outreach/en/tools/practice/>

²⁰⁵ The French “Haute Autorité pour la Diffusion des Œuvres et la Protection des droits d’auteur sur Internet” (HADOPI) was introduced in 2009 with the general task to encourage compliance with copyright laws.

²⁰⁶ Recent EUIPO studies show for example that awareness of legal offers appears to reduce consumption of pirated film. See Online copyright infringement in the European Union available at:

https://euiipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/reports/2020_Online_Copyright_Infringement/2020_Online_Copyright_Infringement_in_the_EU_Title_Level_Study_FullR_en.pdf

Raising copyright awareness is not an easy task. Placing a “do not copy” warning as first screen of a DVD, for example, is not really a deterrent, or that asking Internet users for consents when they do not read terms and conditions is not really protecting their privacy or consumer rights. Therefore, further action in this area is needed.

Some avenues for a way forward

In essence, ongoing copyright awareness campaigns (see annex 5.3) may benefit from several approaches:

- Behavioural **research** on rightholders and users’ sides to identify effective incentives to respect copyright. This could, for example, include research on the development of intelligent, adaptive, **contextual interfaces** that would distil awareness messages (vs. threats) at specific, individual, appropriate moments of use of a creative work.
- **Impact analysis** of the measures taken during the last 20 years in order to adapt and finetune current approaches.
- **Coordination** of efforts to avoid duplicate developments and convey stronger messages.
- Incentives for online platforms to better **distinguish copyright-protected content** from other content and display rights information based on existing standards (e.g., ARDITO, Google image search).
- **Formal education** at the levels of primary and secondary schooling as well as in higher education.

Successful work in this area normally happens when it is targeted to specific sectors or applications where copyright awareness has a practical relevance to the creator, intermediary or user. This could be taken into account when assessing a potential **copyright awareness competence programme**²⁰⁷, to address problems, issue guidance, and coordinate research and development.

2.7.2. Mastering metadata skills

A subsequent step could extend **awareness raising activities** regarding copyright in general to metadata expertise. The interviews in this study and many other reports indicated that many individuals (especially creators and rightholders) have a relatively weak understanding of what metadata are and how to handle them. This lack of expertise ends up in non-exhaustive metadata at various distribution stages of works and related subject matters.

Several initiatives are active in the area of education and expertise regarding metadata. We can highlight, for example, the Working party on IP at the Council of the European Union, the Finnish project around the copyright infrastructure (investing in training programmes on copyright data), the UK project *Music 2025*, the Content Authenticity

²⁰⁷ A programme of this kind might be managed within a **competence centre** in a way comparable to the IP Helpdesk managed by DG GROW, see https://intellectual-property-helpdesk.ec.europa.eu/index_en, or to the EUIPO Observatory, named here as mere examples. A **public private partnership** could also be envisaged. Competence centres in this study are considered as **hubs**, i.e., **nodes** of a network with specific responsibilities, not as “**headquarters**”.

Initiative, or the #MetadataPays awareness campaign by Session (formerly Auddly) and PRS for Music.

Campaigns to raise **awareness** are often linked to offerings of **education**. Everybody using the Internet, needs some copyright awareness since copyrighted content is ubiquitous. Metadata awareness, on the other hand, is a more specific aspect and not something that is relevant for a wider public. While only professionals dealing with content and rights management, or licensing, or enforcement, or remuneration, need to master the creation and use of metadata at the highest level, it would be certainly useful to also increase basic metadata awareness among all parties involved in content creation in creative industries. It should be noted, however, that because of the breadth, complexity and variety of metadata in use or required, expertise is normally related to the specific systems or schemas in use in the area in which a professional is working. Therefore, general programmes on metadata skills and awareness are less likely to be successful than initiatives related to specific sector standards, systems or workflows. Metadata know-how is an interdisciplinary matter at the crossing of law, business practice and technology. During our interviews, we have encountered people highly skilled in one of these domains, rarely in two, never in three. Most of them accumulated their knowledge “on the job”.

Some avenues for a way forward

Coordinated actions could be programmed at two levels:

- The formal education and guidance of **professionals** handling content and rights metadata when developing or using rights and royalties systems
- The development of **intelligent user interfaces** in rights data management systems. Indeed, most existing rights and royalties systems presuppose an important IT and IP knowledge on the side of the users. Should this IT and IP knowledge be encapsulated in the interfaces, many more users – creators, rightholders, prosumers, producers of user-generated content – would access and use sophisticated tools to manage rights management information.

Essentially, a **metadata education programme**²⁰⁸ could be set up to address metadata problems, issue guidance, and coordinate research and development; this is distinct from the copyright awareness competence programme, but the development of both should be co-ordinated.

2.7.3. Making use of new technologies relevant for metadata management

New technologies like Artificial Intelligence as well as blockchain/distributed ledger technologies have the potential to contribute significantly to improved rights metadata management²⁰⁹. Artificial intelligence, for example, can help in **rights management** (e.g. through pattern recognition, rule-based algorithms, and machine learning to create, curate and clean rights management information), **rights licensing** (AI could be used to inform

²⁰⁸ The footnote above on a copyright awareness programme also applies here.

²⁰⁹ See for example the Intellectual Property Infringement and Enforcement Tech Watch Discussion Paper 2020, prepared by The European Observatory on Infringements of Intellectual Property Rights, EUIPO, with support from the Impact of Technology Expert Group (https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/reports/2018_Study_on_legislative_measures_related_to_online_IPR_infringements/2018_Study_on_legislative_measures_related_to_online_IPR_infringements_EN.pdf)

rightholders about the qualitative and quantitative use and monetisation of their works) or **rights enforcement**.²¹⁰

On the other hand, the European Union is investing heavily in the European Blockchain Services Infrastructure, to support the creation of cross-border services, based on blockchain technologies. The EBSI is currently focusing on four use cases, implementing a Self-Sovereign Identity model in Europe (allowing users to create and control their own identity across borders), notarisation in the area of education (digital control of educational credentials, verification of diplomas), traceability of documents (automate compliance checks, prove data integrity) and secure data sharing among public authorities in Europe.

There are clear links to the management of rights management information, such as in the notarisation of digital works or assets) or the creating a self-sovereign identity of authors, rightholders or other stakeholders). The potential in this area was mentioned, for example, in the EU Blockchain procurement documents, where a possible use case of an "EU wide management of IP rights (like patents, trademarks, copyrights), including also the management of copyrights that can be directly associated to digital content in near real-time" was featured²¹¹.

Some avenues for a way forward

While the technologies like Artificial Intelligence and Distributed Ledger/Blockchain and their application are advancing more and more, much (applied) research and development work is still needed to make full use of these new technologies for metadata management. We therefore suggest to

- Support and promote ongoing research on the potential for new technologies in metadata management.
- Work towards concretely making use of the European Blockchain Infrastructure and further scope potential applications in the area of metadata management based on it.
- Take into account aspects related to metadata management in ongoing regulatory processes, for example regarding the current Proposal for an AI Regulation laying down harmonised rules on artificial intelligence²¹², the Digital Services Act (DSA) and the Digital Markets Act (DMA).

²¹⁰ An important aspect in discussing about AI in general, but also for metadata management purposes is algorithm transparency. As mentioned in the introduction of the proposed Digital Services Act, the use of metadata by online platforms to feed AI algorithms and promote content still lacks transparency, e.g., for the establishment of playlists or display of search results. Authors and performers are requesting more transparency on the exploitation of their works and more information about monetisation to balance the bargaining powers between them and the platforms, increase fairness and secure cultural diversity (see G. Mazziotti, A Data-Driven Approach to Copyright in the Age of Online Platforms, EUI Department of Law Research Paper No. 2020/07, 2020 available at SSRN: <https://ssrn.com/abstract=3655027> or <http://dx.doi.org/10.2139/ssrn.3655027> and D. Antal, A. Fletcher, P. Ormosi, Music Streaming: Is it a Level Playing Field? 2021 available at <https://www.competitionpolicyinternational.com/music-streaming-is-it-a-level-playing-field/>)

²¹¹ see <https://etendering.ted.europa.eu/cft/cft-document.html?docId=81917>

²¹² see <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206>

2.7.4. Integrating the existing rights data framework(s)

While the previous paragraphs have addressed the important aspects of awareness and education on copyright and metadata issues especially among creators and smaller rightsholders, in the following we broadly sketch a concrete avenue towards an enhanced, integrated copyright framework and its basic semantic, technical and governance implications.

International **rights data framework(s)**, including many standards related to metadata issues, already exist. Based on the results for this as well as previous studies, the study team argues, however, that it still has gaps and weaknesses. Most notably, it lacks **interoperability** across different media or content sectors. This becomes apparent, for example, when looking at the many ongoing initiatives which are almost all sector-specific. Most of the interoperability initiatives identified in section 2.3 focus on issues *within* market sectors. At the all-media level, generic technical identifiers like URI and markup/exchange/coding languages like HTML, XML and JSON successfully provide essential resources for data interoperability in the digital network. However, so far there are few data standards for content and rights which are *designed* to be sector-neutral. In a world, where multi-media content combines creative works from various sectors (for example text, video, music on a website)²¹³, this could help bring rights data management for these different creative works more closely together and streamline remuneration processes.

Among the few examples of sector-neutral standards, the following stand out:

- ISNI (for parties) and DOI (for content, although it can be used for anything) are two standard identifiers which have begun to have success across more than one distinct market sector. ISNI (and with the closely related ORCID) in particular is emerging as a potentially invaluable interoperability tool as it can cover any content types and is taking root in both the bibliographic and commercial sectors,
- Content metadata schemas and datasets often include rights elements, but those in common use are specific to particular media/content/market/cultural sectors or functions (such as orphan works) and are inherently not generally extensible. This is unavoidable: the way to schema interoperability is through trusted mappings and transformations which allow content and rights data to pass accurately from one schema to another, and while methodologies and tools exist for this there is no co-ordinated or trusted framework in place,
- ODRL (Open Digital Rights Language) is the broadest in scope of licensing-oriented rights schemas, although its adoption to date has been in very specific sectors. Creative Commons and others offer generic licences which may apply to any type of content, but have limited scope,
- ISCC (International Standard Content Code) is an example of content-derived identification mechanism which offers, in combination with industry identifiers,

²¹³ The need for a cross-sectoral rights management is not new. Already four different copyrights could be applied to a vinyl recording, for example: one for the composition, one for the performance, one for the cover art, and one for the sleeve notes. The stepwise digitisation of music (from vinyl to CD to sound download to video streaming) did not alter the need. Nowadays, modular digital educational content including static, animated and augmented images (photography and graphics), texts, music, sounds and voice-over that can be configured and remixed in a multitude of ways by a multitude of contributing rightsholders is another good example of the need. Video games and user-generated content provide even more examples.

various possible distributed solutions for some essential mapping, linking, deduplication, and matching issues.

A **cross-sector rights data network could bridge gaps** between standard content identifiers such as ISRC, ISWC, ISBN or ISAN and digital manifestations of the content they denote. It could therefore enable the matching of digital objects. The ultimate objective of this endeavour would be to break the silos of different creative industries and improve the efficient rights data management and licensing across sectors.

An open integrated rights data framework would have several benefits. For example, it would:

- Support releasing much more of the **digital potential** of Europe's creative sectors and contribute to the development of a **single market for data**.
- Address **pain points** discussed in section 2.5 (and addressed in annex 8), including interoperability issues. Specifically it would make rights management simpler, more accurate, faster, and more affordable for **all stakeholders on the content value network**
- Provide **trustworthy rights information** which can then be relied upon for rights licensing and rights enforcement, as well as for a fair, appropriate, proportionate, and transparent rights remuneration
- Restore a **level-playing field** between major actors and the European small and medium sized creative enterprises through an inclusive approach catering for interests of any rightholder, stakeholder, incumbent or new intermediary

The emergence of an all-media, cross-sector digital content and data network is relatively recent and is still poorly supported in terms of standards and services for interoperability. Its structure would include a set of foundational standards and technologies which underlie the exchange of rights management information. This *infra*-structure supports a fragmented network of rights declarations, attributions, verifications, and queries in the digital era. In more concrete terms this could mean authoritative mappings available as services supporting automated "translation" of metadata.

Of course, an integrated rights data framework like the one sketched above also needs an effective, integrated governance structure. In this sense, a European multistakeholder governance²¹⁴ would need to be set up comprising, e.g., rightholders, publishers, collective management organisations, distributors, and users. Indeed, while data governance is an object of attention to publishers, collective management organisations and distributors, it still lacks awareness and understanding in other parts of the value network, such as in the creation, retail, and consumption of content.

Some avenues for a way forward

Opening and integrating the framework – as described in more detail in Annexes 6 and 7 below – could boost the value of the creative content sectors significantly. The challenge

²¹⁴ "Multistakeholder governance is a practice of governance that employs bringing multiple stakeholders together to participate in dialogue, decision making, and implementation of responses to jointly perceived problems. The principle behind such a structure is that if enough input is provided by multiple types of actors involved in a question, the eventual consensual decision gains more legitimacy, and can be more effectively implemented than a traditional state-based response. While the evolution of multistakeholder governance is occurring principally at the international level, public-private partnerships (PPPs) are domestic analogues" from Wikipedia.

is to make it **trustworthy**, **interoperable**, and as **accessible** and **comprehensive** as possible. To do so it must be made **extensible**, capable to support any business model, present or future, for any sector in any jurisdiction.

Opening and integrating the rights data framework means enhancing the existing framework, it does not mean replacing it. Moreover, it should only be done on an as-needed basis for the sake of accessibility and interoperability.

Different aspects or layers would need to be addressed. The foundational layer of the framework is **legal**, as content rights do not exist without agreements and statutes being in place. Above that, we see at the practical level two parts: a **semantic** layer, in which the meaning of identifiers and metadata terms in agreements and rights needs to be clear and consistent, and a **technology** layer which provides mechanisms for trust and interoperation. Finally, as mentioned above, there is a **governance** layer to oversee necessary integration.

The following steps could be useful to consider:

- **Further develop the idea and align the scope of the suggested open rights data framework** with the (a) *acquis communautaire*, (b) ongoing regulatory initiatives, (c) definition of copyright infrastructure as formulated by WIPO, the Council of the European Union, and in the IP Action Plan, and future-proof this scope by considering (d) the EU Strategy for Data and the concept of Common European Data Space, and potentially (e) the European Blockchain Services Infrastructure (EBSI).
- **Discuss** the concept of an open rights data framework within the **inclusive industry dialogue**²¹⁵ foreseen in the Commission's IP Action Plan, to further align the proposal with industry needs and priorities.
- **Encourage research, development, and standardisation** of identifications and protocols necessary to open the rights data framework. Bring together representatives from sector metadata and identifier standard groups/bodies²¹⁶ with other major interested parties to develop a board/committee structure to oversee the strategic, technical, and administrative work needed.
- **Promote** non-proprietary content-dependent identification mechanisms, digital identity wallets, identifier binding protocols and verifiable credentials.
- **Let the industry set up business models and software applications** based among others on the suggested open rights data framework, whereby this framework would remain neutral to content genre and business models.

²¹⁵ Including professional associations, trade bodies and collective management organisations.

²¹⁶ These groups, such as DDEX, EDItEUR/ONIX, IPTC, DOI, Creative Commons, ISXX identifier agencies and cultural heritage standards groups typically already have engagement from rightholders, intermediaries and platforms/users in their sectors and so bring with them representation of a broad range of interests on data issues.

3. Second Part: Copyright and artificial intelligence in the field of creative industries

3.1. Introduction

3.1.1. Context

In recent decades, there has been a sharp increase in the number of creative works available in digital format, either because they came into existence in the form of digital files, or because they were digitalised after having been created and stored in analogue formats (e.g. hand-made paintings, tape recordings, etc.). Thus, **digitalisation** reached the creative and cultural sectors.

Also in the last few decades, we have witnessed how data analysis (the process of collecting, cleaning and extracting information of data, i.e., through pattern identification /inference deduction), and the strong increase in computing power opened the way to ever more advanced algorithms that could process larger and continuously updated datasets. These systems that are increasingly more powerful are today commonly known as **artificial intelligence** (AI) systems due to their “ability, for a given set of human-defined objectives, to generate outputs such as content, predictions, recommendations, or decisions which influence the environment with which the system interacts, be it in a physical or digital dimension”.²¹⁷

In recent years, AI has moved out of the pre-square of high-tech companies' research departments and is nowadays used across different industries in a wide range of applications. The **creative industry**, despite being one of the most inherently human sectors, which rely on capabilities traditionally regarded as intrinsically humans such as imagination, inspiration, and creativity, is no exception to this trend and the deployment of AI tools across its value chains can be observed.

3.1.2. Policy context and motivation of the study

The reliance on these technologies might be responsible for the emergence of **new tensions with intellectual property**, including the copyright legal framework. Large amounts of datasets of creative works are indeed given as input to AI. While many AI applications in industrial sectors use datasets not protected by copyright,²¹⁸ AI applications in the creative and cultural sectors might concern or reuse protected subject matter. AI solutions might then permit the generation of cultural outputs, the protectability of which under copyright law is questionable. Therefore, the relationship between the copyright legal framework and the use of AI in the creative sector deserves special attention.

Intellectual property issues raised by the use of AI have **recently attracted the attention of international organisations and of the EU institutions**.

At the **international level**, the WIPO engages in a “Conversation on Intellectual Property Policy and Artificial Intelligence” with the aims of “formulating the questions that the policymakers need to ask”.²¹⁹ In this regard, the WIPO published a first “draft issues Paper”

²¹⁷ Comp. Commission proposal COM/2021/206 for a Regulation laying down harmonised rules on Artificial Intelligence (artificial intelligence act), (Annex 1, identifying 3 types of AI techniques and approaches, Rec. 6).

²¹⁸ Comp. Commission proposal COM/2021/206 final for a Regulation laying down harmonised rules on Artificial Intelligence (artificial intelligence act), (Annex 1, identifying 3 types of AI techniques and approaches).

²¹⁹ https://www.wipo.int/about-ip/en/artificial_intelligence/conversation.html

in December 2019²²⁰, followed by a “revised issues Paper”²²¹ in May 2020, taking into account the feedback received on the first paper.

Concerning copyright, the second WIPO paper first noted especially that “AI applications are increasingly capable of generating literary and artistic works. This capacity raises major policy questions for the copyright system, which has always been intimately associated with the human creative spirit and with respect and reward for, and the encouragement of, the expression of human creativity. The policy positions adopted in relation to the attribution of copyright to AI-generated works will go to the heart of the social purpose for which the copyright system exists. If AI-generated works were excluded from eligibility for copyright protection, the copyright system would be seen as an instrument for encouraging and favoring the dignity of human creativity over machine creativity. If copyright protection were accorded to AI-generated works, the copyright system would tend to be seen as an instrument favoring the availability for the consumer of the largest number of creative works and of placing an equal value on human and machine creativity.”²²² Under this general tension, the WIPO further identify several more specific issues that the policymakers will be confronted with as for instance: “Do AI generated-works require copyright or a similar incentive system at all?”, “Should copyright be attributed to original AI-generated literary and artistic works or should a human creator be required?”²²³. Lastly, the WIPO’s paper also notes that several issues arise with regard to the training of AI applications since the “the data used for training the AI application may represent creative works”²²⁴

Initially, the focus of the **European debate** on the law and policy of AI has been on the ethical aspects²²⁵, the liability of autonomous systems²²⁶, the legal personality of robots²²⁷, the concerns of data access²²⁸, the protection of personal data. More recently, the Commission adopted a proposal for a transversal regulation (**the AI act**) which follows a risk-based approach for the development and deployment of AI systems within the internal market.²²⁹

In the field of copyright policy²³⁰, the changing creative environment shaped by AI was not the main reason for the most recent copyright reform leading to the **2019/790 directive**

²²⁰ WIPO 13 December 2019, Issues paper on intellectual property policy and artificial intelligence, WIPO/IP/AI/2/GE/20/1, https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_ai_2_ge_20/wipo_ip_ai_2_ge_20_1.pdf

²²¹ WIPO 21 May 2020, Revised issues paper on intellectual property policy and artificial intelligence, WIPO/IP/AI/2/GE/20/1 REV., https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_ai_2_ge_20/wipo_ip_ai_2_ge_20_1_rev.pdf

²²² Idem, p. 7.

²²³ Idem

²²⁴ Idem, p. 8.

²²⁵ High level expert group on AI, *Ethics Guidelines for Trustworthy AI*, <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>, (accessed in July 2021).

²²⁶ Register of Commission Expert Groups and Other Similar Entities, <https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&docid=36608>, (accessed in July 2021).

²²⁷ European Parliament, Resolution 2015/2103(INL) of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics.

²²⁸ EU Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions (A European strategy for data), COM(2020) 66 final, 19 February 2020.

²²⁹ EU Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts, COM(2021) 206 final, 21 April 2021.

²³⁰ European Parliament, Report on intellectual property rights for the development of artificial intelligence technologies, 2020/2015(INI), 2 October 2020.

(DSM dir.), although this directive contains highly relevant provisions, in particular, the two text and data mining (TDM) exceptions (Art. 3 and 4) but also the new press publisher's right (Art. 15).

However, in its **IP Action Plan**,²³¹ the EU Commission underlined the need for a "reflection on how and what is to be protected. AI technologies are creating new works and inventions. In some cases, for instance in the cultural sector, the use of inventive machines may become the norm. These developments raise the question of what protection should be given to products created with the help of AI technologies. Discussions on the impact of AI on IPRs are ongoing both in Europe and internationally"²³².

The 2020 IVIR/JIPP study²³³ concluded that if the "current EU IP framework and the European Patent Convention appear broadly suitable to address the challenges raised" by AI, "[h]owever, harmonisation gaps and room for improvement remain"²³⁴. As a first step, the Commission therefore announced that it "will map and analyse all issues and engage in stakeholder discussions"²³⁵.

This Communication from the Commission was welcomed by the **EU Parliament** in its **Resolution on artificial intelligence** in education, culture and the audiovisual sector.²³⁶ The EP furthermore "calls for the intellectual property action plan announced by the Commission to address the question of AI and its impact on the creative sectors, taking account of the need to strike a balance between protecting IPR and encouraging creativity in the areas of education, culture and research"²³⁷. In this regard, the EP:

- "calls on the Commission and the Member States to address the issue of AI-generated content and its challenges to authorship and copyright infringement;²³⁸
- asks the Commission, in that regard, to assess the impact of AI and related technologies on the audiovisual sector and the CCSI [Cultural and creative sectors and industries], with a view to promoting cultural and linguistic diversity, while respecting authors' and performers' rights"²³⁹,
- "[e]mphasises the role of an author's personality for the expression of free and creative choices that constitute the originality of works;²⁴⁰

²³¹ EU Commission, Communication from the commission to the European Parliament, the Council, the European economic and social Committee and the Committee of the regions (Making the most of the EU's innovative potential - An intellectual property action plan to support the EU's recovery and resilience), COM(2020) 760 final, 25 November 2020.

²³² *Idem*, p. 6

²³³ C. HARTMANN, J. ALLAN, P.B. HUGENHOLTZ, J.P. QUINTAIS et D. GERVAIS, "JIIP & IVIR report on AI and IP", *op. cit.*, p. 92.

²³⁴ EU Commission, EU Commission, Communication from the Commission to the European Parliament, the Council, the European economic and social Committee and the Committee of the regions (Making the most of the EU's innovative potential - An intellectual property action plan to support the EU's recovery and resilience), COM(2020) 760 final, 25 November 2020, p. 7

²³⁵ *Idem*.

²³⁶ European Parliament, Resolution on artificial intelligence in education, culture and the audiovisual sector, 2020/2017(INI), 19 May 2021, §58-76. See also European Parliament, Report on intellectual property rights for the development of artificial intelligence technologies, 2020/2015(INI), 2 October 2020, §6

²³⁷ European Parliament, Resolution on artificial intelligence in education, culture and the audiovisual sector, 2020/2017(INI), 19 May 2021, §73.

²³⁸ See also also European Parliament, Report on intellectual property rights for the development of artificial intelligence technologies, 2020/2015(INI), 2 October 2020, §14-15.

²³⁹ European Parliament, Resolution on artificial intelligence in education, culture and the audiovisual sector, 2020/2017(INI), 19 May 2021, §65.

²⁴⁰ See also European Parliament, Report on intellectual property rights for the development of artificial intelligence technologies, 2020/2015(INI), 2 October 2020, §15

- underlines the importance of limitations and exceptions to copyright when using content as data input, notably in education, academia and research, and in the production of cultural and creative output, such as audiovisual output and user-generated content”²⁴¹, and lastly
- “asks the Commission to assess the impact of IPR on the research and development of AI and related technologies, as well as on the CCSI, including the audiovisual sector, with particular regard to authorship, fair remuneration of authors and related questions”²⁴²

Taking into account that context, **the present study** aims at offering a comprehensive analysis, first, of the current use of AI in selected creative sectors, and second, of the suitability of the copyright framework to face the implication of this new technological development.

This study is not the first one that the Commission has commissioned in the field of AI. A study entitled “Trends and Developments in Artificial Intelligence - Challenges to the Intellectual Property Rights Framework”²⁴³ was released before the communication of the Commission on its IP action plan. The scope of this previous study was, however, different: regarding the economic sectors, it was broader as it was not limited to the cultural and creative industries which are the sole focus on the present study; also it reviewed not only the copyright, but also the patent framework; however the previous study was somewhat narrower, since it focused on the output side of the AI process (i. e. the protection of the AI assisted or generated artefacts) and did not rely on an extensive survey of the stakeholders to propose and discuss policy options for copyright. Additionally, a further study on “Opportunities and challenges of Artificial Intelligence Technologies for the Cultural and Creative Sectors” is currently (2021) carried out for European Commission. It, however, concentrates on different questions, namely i) to give an overview of the cultural and creative ecosystems’ readiness to adopt and deploy AI and related technologies, ii) to identify AI and related technologies that are currently being used and/or will/can be deployed in future for the cultural and creative sectors, and iii) to examine the challenges and opportunities that AI and related technologies raise for the promotion of, and access to, a culturally diverse offer of European works. Ultimately, the study aims to identify the main investment needs of the ecosystem and indicate which support measures and instruments are available to fund them both. It thus serves very different purposes than the present study.

3.1.3. Scope of the study

This **second part of the present study** aims to provide insight into the copyright and related rights issues connected to the use of AI in the selected cultural sectors, in particular for visual arts, music, audiovisual and film, gaming and cultural heritage. The development and use of AI solutions in relation to creative productions challenge the application of fundamental copyright principles, which have taken shape over the years while mass-digitisation or AI technologies were not yet as advanced and sophisticated as today. The recent challenges stemming from those new technologies and practices are identified in this report. They might have implications for copyright at the international, EU and national level. The evidence on which the study relies has been systematically gathered through

²⁴¹ Idem, §70.

²⁴² Idem, §73.

²⁴³ C. HARTMANN et al, “Trends and Developments in Artificial Intelligence. Challenges to the Intellectual Property Rights Framework”, <https://op.europa.eu/en/publication-detail/-/publication/394345a1-2ecf-11eb-b27b-01aa75ed71a1/language-en>, (accessed in July 2021).

desk research, semi-structured interviews, and a Delphi survey with industry stakeholders (see below on the methodological aspects). The responses of the participants during the interviews and surveys reflect the copyright challenges they identify. The views and claims of the stakeholders do not necessarily require a response from the EU authorities and legislator as copyright policy making also takes place elsewhere (at least at international and national levels). Nevertheless, the collected evidence was analysed with the view to identify possible obstacles and uncertainties in the EU copyright framework, as well as to formulate certain options to address these challenges under various policy scenarios and to reflect on their likely impacts.

This second part of the report is **structured** as follows. In the next section, the methodology followed in this study is described. Then, we offer an overview of the copyright relevant applications of AI in the different creative and cultural sectors covered in this study and provide insights into selected uses cases found in the industry. The second section (p. 148) presents the challenges raised by the implementation of these AI applications and the potential shortcoming of the existing copyright legal framework in this regard. The last section (p. 203) of this study proposes potential policy scenarios to tackle the issues identified. Each policy scenario is analysed and discussed with special consideration of its impact on stakeholders. The policy scenarios are grouped into those related to input data for AI training and those related to the output generated by AI.²⁴⁴ **Considering the early stages of development and adoption of AI in the selected cultural sector, these policy scenarios are merely offering some options for legislators to address the identified challenges should they decide to intervene in the future.**

Before diving into the topic and the findings, the following box explains the **specific terminology used in this study**.

Terminology used in this study is as follow:

- **AI solution / AI tool / AI application / AI system**²⁴⁵: a software that is developed with one or more techniques (such as machine-learning) and can process and/or generate creative or cultural outputs.
- **AI input**: datasets, comprising protected works and other subject matter, used to train or to trigger the AI solution and generate the AI output. AI input may refer to the data used by the developer to train an AI solution or to the data used by the user of the AI solution in order to generate an AI output.
- **AI output**: products/outcomes resulting from the use of an AI solution.
 - (i) **AI-autonomously generated output**: An output entirely generated by AI without human creative choices being made at any relevant stage of the generation process of the art piece.
 - (ii) **AI-assisted output**: An output created with the help of an AI solution, but that is the result of creative choices (either at the input or output level of the AI process) made by a human.

²⁴⁴ The copyright protection of the AI solutions or training model is out of the scope of this study.

²⁴⁵ The four terms are used as synonyms.

- **AI training:** the process by which the learnable parameters of an AI model are optimised through its feeding with data to perform a specific task.

3.1.4. Methodological remarks

This part of the study focuses on copyright issues concerning the use of AI in five creative sectors, namely visual arts, music, audiovisual & film and gaming, and cultural heritage. Legal and empirical techniques were employed to collect and analyse information and data provided by stakeholders.

As a first step, an in-depth analysis of the EU relevant legal framework regarding copyright and copyright-related rights was conducted. The latter aims to identify various legal issues of copyright in relation to AI from a theoretical perspective. This desk research included examining the applicable international conventions, the EU secondary legislation, and specific examples stemming from national legislations, which might be relevant as elements of comparison. Case law, and especially the decisions of the Court of Justice, were also duly taken into account. The analysis of the legal framework was also lastly flanked by an intensive review of the existing literature. In parallel, desk research was conducted to map applications of AI in the selected creative sectors. For each sector, a global overview of the use of AI is presented, examples of AI applications are discussed and a case study of one application is showcased.

A total of 28 in-depth semi-structured interviews were conducted with legal experts with diverse professional experience on the topic of copyright and AI, such as academics, lawyers, legal advisors of company associations, developers of AI solutions. The experts provided valuable insight into the views of the industry stakeholders in the creative sectors covered in this study.

The literature review, desk research and interviews contributed to the identification of issues and policy scenarios. The opinions of legal experts and industry stakeholders on the policy scenarios were gathered through a Delphi survey with two rounds. In the first survey round, the copyright issues identified through the literature review and interviews were presented to the survey participants, together with the different policy scenarios identified so far. Participants were asked to provide their opinion on the different policy scenarios through a Likert scale and were prompted to elaborate on them and suggest additional scenarios through text fields. The analysis of the first round of the survey guided the design of the second round.

The 64 participants who completed the first round of the survey were invited to participate in the second round, of whom 19 completed the second survey. These participants were briefly informed of the preliminary results collected in the first round: for each of the copyright issues covered in the first round of the survey, a short description of the results of the first round was provided. This second round allowed the inquiry of opinions on the additional policy scenarios suggested by participants in the first round of the survey, as well as a more in-depth analysis of each of the policy scenarios, in terms of impact. For each of the policy options that were considered to be more adequate by the participants of the first round of the survey, questions on their impact on costs, revenues, investments, litigation, etc. were formulated.

The sample of participants was formed by representatives of cultural associations in the different creative fields covered by this study, academics, companies using/developing AI in the creative sector, and individual artists with experience in the use of AI. Regarding the professional activity of the participants, most were legal experts (31), followed by CMOs

(19) and individual artists/authors/s (10) (see Figure 11). This is important background and contextual information when interpreting the survey results.

It should be noted as a caveat for the interpretation of all results that some participants answered as individuals, while others answered on behalf of several individuals/organisations they represent, as it is the case, for instance, of collective management societies. Therefore, while the Delphi surveys give an interesting perspective on the stakeholders' opinion on this matter, their result should not be considered as reflecting a fully-fledged stakeholder consultation.

Furthermore, the composition of the sample by sector needs to be taken into account when interpreting the results. As it can be seen in Figure 12, most participants who completed the first round of the survey (48) indicated intellectual property as their field of expertise, followed by audio-visual creations (12), visual arts and design (10), IT-related law (9), music (6), cultural heritage (2) and consumer law (2), video games (1) and consulting (1). It should also be noted that 7 of the participants indicated that they operate in the publishing sector, which is not within the scope of this study. Representatives of this sector manifested their interest in participating and the invitation to participate was extended to them. Regarding country distribution, 84% of the participants are located in 15 EU Member States, most of them in Belgium (13 participants; 20%), Germany (7; 11%) and Austria (5; 8%).

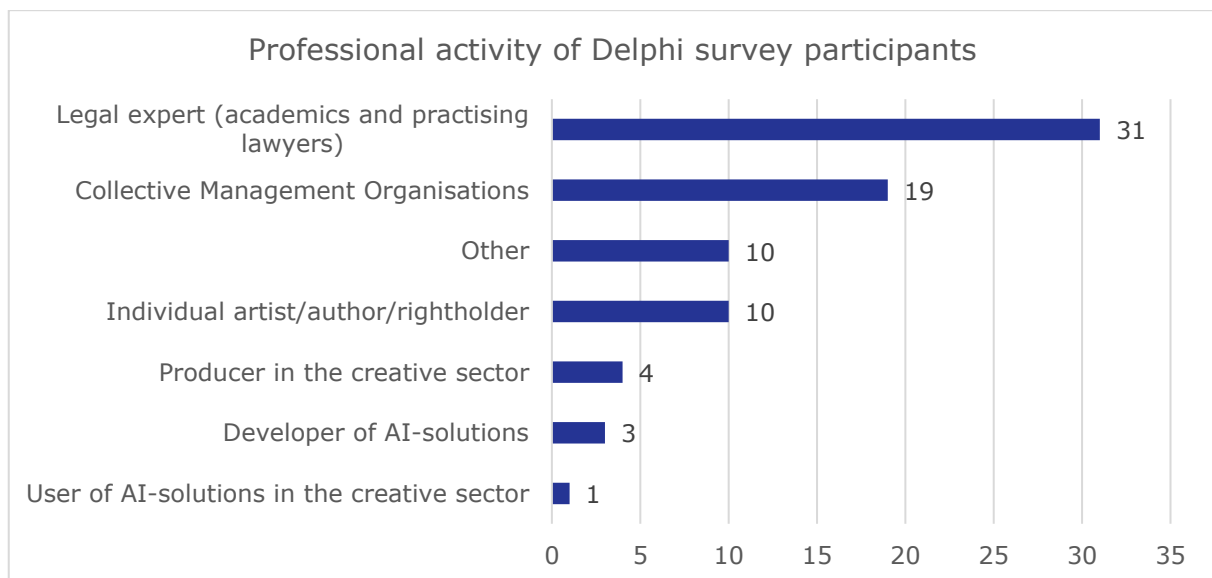


Figure 11: Professional activity of Delphi survey participants (N=64)

Source: Technopolis Group Delphi Survey

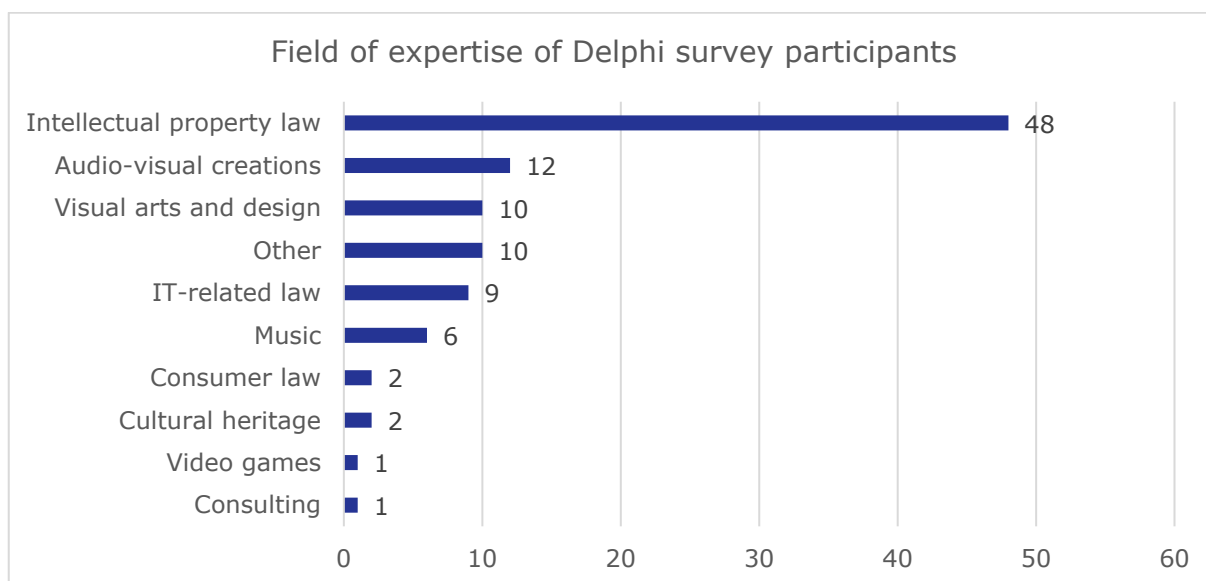


Figure 12: Field of expertise of Delphi survey participants (N=64)

Source: Technopolis Group Delphi Survey

In general, the study team is of the opinion, that the Delphi survey gives a unique empirical basis for providing bottom-up insight of practitioners from various creative industries on the study questions. Limitations of the approach include a certain bias in the respondent base towards IP law experts (see Figure 12). Individual artists, content producers or AI developers” have taken part to a smaller extent than law experts. This should be taken into account as a relevant context factor for the results presented in the following chapters. However, it should also be noted that respondents having identified themselves as law experts also qualify to some extent as “content producers” or “developers”, for example in the case of respondents from law departments of companies working with AI applications.

To ensure a robust interpretation for the study, all the evidence collected through literature review, desk research, interviews, Delphi survey was analysed and triangulated. For each of the issues identified, the study team provides an independent and non-partisan expert opinion on the issue and the possible policy scenarios, without any prescriptive intention but rather leaving room for further debate.

Apart from the Delphi, another survey consultation was conducted with 11 academics specialised in copyright law attending a session of a European Copyright Society (ECS) conference. A brief introduction on the copyright issue was given, followed by a multiple-choice question to select possible actions to be taken as adequate policy scenarios.

3.2. Current use of artificial intelligence in selected creative sectors

Broadly speaking, there are **two main types of uses** of AI in the creative and cultural sector: one that aims to assist or replace humans in the creative process of generating creative outputs (e.g. a melody, a visual effect, etc); and another that aims to automate tasks other than the creative process, such as in the investment, distribution and business processes (e.g. AI-powered support for the decision whether or not to produce or distribute some content, the cataloguing of creative works, the detection of copyright infringements,

the content recommendations, voice to text transcriptions, etc.; see sections 2.2.3.6 as well as 2.3.3.2 in the first part of the study). This second type of use of AI tools remains outside the scope of this study.

In the course of this study, we examine the reliance on AI solutions in **four creative sectors**: (i) visual arts, (ii) music, (iii) audiovisual and film, and (iv) video games. Our research demonstrates that the degree of adoption of AI solutions varies, depending on the sector. Concerning **cultural heritage**, the use of AI is not widely adopted. Applications are mostly in the digitalization, description and classification of art collections (e.g. automatic classification of cultural heritage objects, automation of the annotation of iconographic elements in works of art etc.). The efforts in developing technical tools are mainly aimed at improving the management of the collections and making them more searchable. Finally, the existing applications in the cultural heritage sector do not appear to involve AI tools directly linked to the creation process and are less relevant than those used in the other sectors analysed in this study (especially visual art). For this reason, no distinct section is devoted to the cultural heritage – but references to applications in the cultural heritage sector will be included in the other sections (where relevant).

In the following four sub-sections, we therefore present successively AI use cases having some copyright relevance in the four other sectors – visual arts (3.2.1), music (3.2.2), audiovisual and film (3.2.3), gaming (3.2.4) – with references to the cultural heritage sector when relevant. The presentation of each sector is subdivided into three parts: (i) an overview of the AI development observable within that sector, (ii) a list of existing AI solutions and (iii) an in-depth analysis of the copyright-relevant aspects of a case study. A fifth and concluding sub-section (3.2.5) offers a transversal analysis of the AI uses across the cultural sectors and attempts to common characteristics.

3.2.1. Visual arts

AI applications in the sector of the visual arts are quickly emerging. The copyright questions arise in relation to the paintings, photographs and representations of three-dimensional works, such as sculptures or design, that are used to train the AI applications, they also concern the (protection of the) AI-output, which may be very similar to, if not indistinguishable from, human-created works of visual art.

3.2.1.1. Global overview

Image processing is a heavily researched sub-domain of AI, due to this field's high potential for practical applications (including, for example, applications in medicine, automated driving etc.). Therefore, there already exists a wide range of advanced AI technologies for image processing that have found application in the creative sector as well.

Regarding **AI-based image generation**, a useful distinction between different types of generation processes may be drawn, according to the types of input data that are used to generate the output of the AI.²⁴⁶

- **Non-image-to-image:** In these applications, the AI is trained on categorical or numerical features such that the algorithm learns (during the training phase) a mapping from these inputs to images. From the perspective of the active user of the AI, the inputs provided to the algorithm may be completely generic and non-

²⁴⁶ N. ANANTRASIRICHAJ and D. BULL, "Artificial Intelligence in the Creative Industries: A Review", *Artif Intell Rev* 2021, <https://doi.org/10.1007/s10462-021-10039-7>.

proprietary. An example of this would be 'Generated Photos'²⁴⁷ (described in detail in the case study below), whereby the AI yields an image of a person's face based on generic inputs that include e.g. a variety of descriptive features such as sex, make-up, emotion displays, skin tone. Another example of this type of AI-based image production would be 'The Next Rembrandt'²⁴⁸ project in which a team of researchers assembled a training set for the AI including not only a database of Rembrandt paintings but also a set of heavily processed features (to capture typical stylistic properties of Rembrandt's paintings such as motifs, lightening, typical facial expressions of the subjects portrayed, geometric patterns etc.), which required a substantive amount of additional labour in the creation of the training database.

- **Image-to-image:** In these applications, the AI is trained to learn a mapping from images to images. To generate an image, users of the AI therefore need to provide an image as the input. This circumstance may potentially raise additional questions regarding the relationship between the ownership of the creative content of the input and the ownership of the output of the AI. An example of an implementation of this idea in the creative sector is the work done by Artbreeder²⁴⁹ and GANVAS Studio²⁵⁰. In this project, the AI produces images – many of them rather surrealistic – based on images provided by the user. Other prominent examples of this type of AI-based image generation in the creative sector are several important image post-production techniques (such as enhancement, colourisation, data compression, denoising etc.), where an input image is transformed by the AI in a way that meets the purposes of the active user.

AI applications in the visual sector **other than image generation** seem to be less common and are still in the early stages of their development. However, there are still potential ways in which AI could be relevant to this area of the creative sector as is demonstrated by examples such as 'Art Verified by A.I.', 'ArtPI' by Atrendex and 'ArtRank', which are AI-based evaluations of art authenticity, art recommendation, and valuation forecasting of art, respectively.

3.2.1.2. Examples of artificial intelligence solutions

Amongst examples of current AI applications is **Prisma**²⁵¹ a photo editor mobile app that uses AI algorithms to transform active user's photos into works of art, change the background or foreground, overlay objects with different objects and clone/copy the style or effects from another image. The AI uses styles inspired by various famous artists (e.g., Picasso, Munch, or Salvador Dali).

MyHeritage, an Israel-based company, launched **Deep Nostalgia**²⁵² which allows active users to animate their photos. The technology used was licensed by MyHeritage from D-ID, an Israel-based company specialising in video re-enactment using deep learning. The Deep Nostalgia algorithm was trained on video footage of real human movements and gestures produced by MyHeritage. After a picture is uploaded to the platform, Deep Nostalgia enhances it and then automatically decides which sequence to apply to the face based on its orientation. However, users can depart from the default sequence and select a different animation made available to them through the platform. The result is a

²⁴⁷ Generated photo website, <https://generated.photos/>, (accessed in July 2021).

²⁴⁸ The Next Rembrandt, <https://www.nextrembrandt.com/>, (accessed in July 2021).

²⁴⁹ Artbreeder, <https://www.artbreeder.com/>, (accessed in July 2021).

²⁵⁰ GANVAS studio, <https://ganvas.studio/>, (accessed in July 2021).

²⁵¹ Prisma, <https://prisma-ai.com/>, (accessed in July 2021).

²⁵² Deep Nostalgia, <https://www.myheritage.com/deep-nostalgia>, (accessed in July 2021).

technological simulation of how the person in the photo would have looked and moved if they had been captured on video. MyHeritage places motion icons on all animated photos so that users can distinguish them from the original. The solution can be used for free, but the paid version gives access to more animation features. MyHeritage states that it does not own the uploaded photo nor the resulting video, instead they belong to the user.

DeepDream²⁵³ is a computer vision program developed by Google that uses an artificial neural network to find and enhance patterns in images through algorithmic pareidolia. The network was trained by feeding it millions of training examples and gradually adjusting the network parameters until it produced the desired classifications. The network typically consists of 10-30 stacked layers of artificial neurons. Each image is fed into the input layer, which then talks to the next layer until eventually the output layer is reached. Each layer progressively extracts higher and higher-level features of the image, until the final layer essentially decides on what the image shows. Soon after Google made their code open-source²⁵⁴, several tools appeared on the market, such as the **Deep Dream Generator**²⁵⁵, which allows active users to transform their own photos into dream-like images.

Deepsketch²⁵⁶ is a sketch recognition technology developed by the Numediart Institute of the University of Mons. This technology uses a deep convolution neural network (ConvNets) to enable classification and medium/high features extraction. The network features are then used as a basis for similarity search using K-nearest Neighbors in large scale datasets such as the TU-Berlin benchmark (a gathering of 2000 unique sketches) and the sketchy database (a large-scale collection for sketch-based image retrieval). Recently, a Deepsketch prototype was used in **Pierre Alechinsky's exhibition Carta Canta**²⁵⁷ at the Royal Museums of Fine Arts of Belgium. The visitors of the exhibition were invited to draw sketches on a screen placed in the exhibition room to explore and interact with Alechinsky's collection. The AI tool showed then the drawings of the artist that are most similar to the sketch.

GANVAS Studio²⁵⁸, launched by Danielle Baskin, turns images created through Artbreeder into posters or canvases. In particular, the images are curated and tuned through repetitive mixing using Joel Simon's **Artbreeder**²⁵⁹. Artbreeder uses BigGAN (image generator trained on 512x512 from ImageNet) and StyleGAN. The output, which is only 256px, is then enlarged using machine learning. GANVAS Studio hand-paints details, edges and textures onto the canvas to enhance detail and remove pixelization.

DALL·E²⁶⁰ is a transformer language model developed by OpenAI that interprets natural language inputs to generate images. It uses a dataset of text-image pairs gleaned from the Internet, i.e. it receives both the text and the image as a single stream of data

²⁵³ DeepDream, <https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html>, (accessed in July 2021).

²⁵⁴ Google-Drop Dream, <https://github.com/google/deepdream>, (accessed in July 2021).

²⁵⁵ Deep Dream Generator, <https://deepdreamgenerator.com/>, (accessed in July 2021).

²⁵⁶ O. SEDDATI, S. DUPONT and S. MAHMOUDI, "Deepsketch 3: Analyzing deep neural networks features for better sketch recognition and sketch-based image retrieval", *Multimedia Tools and Applications* 2017, 76 (4), pp. 1-27; O. SEDDATI, S. DUPONT ET S. MAHMOUDI, "Deepsketch: Deep convolutional neural networks for sketch recognition and similarity search", *CBMI* 2015, pp. 1-6.

²⁵⁷ Fine Arts Museum, <https://www.fine-arts-museum.be/en/agenda/2021/04/01/draw-and-explore-with-pierre-alechinsky#>, (accessed in July 2021).

²⁵⁸ Ganvas, <https://ganvas.studio/>, (accessed in July 2021).

²⁵⁹ Art Breeder, <https://www.artbreeder.com/>, (accessed in July 2021).

²⁶⁰ Openai, <https://openai.com/blog/dall-e>, (accessed in July 2021).

containing up to 1 280 tokens²⁶¹. Hence, DALL·E is capable of creating plausible images for a great variety of sentences that explore the compositional structure of language.

Another example is **Nutella Unica**²⁶² which is a project that was carried out in collaboration with the advertising agency Ogilvy & Mather Italia. By pulling from a database of dozens of patterns and colours, an algorithm generated 7 million different versions of Nutella's graphic identity used as packaging designs for Nutella's jars in Italy.

The Dalí Lives²⁶³ exhibition at Dalí Museum (Florida) – in partnership with the ad agency Goodby, Silverstein & Partners (GS&P) – provides its visitors with an opportunity to learn more about Salvador Dalí's life from Dalí himself. Using archival footage from interviews, GS&P pulled over 6,000 frames and used 1,000 hours of ML to train the AI algorithm on Dalí's face.

The Next Rembrandt²⁶⁴ is a 3D-printed painting made solely from data of Rembrandt's body of work. Indeed, the AI was imprinted with 346 of Rembrandt's known works. The team used high-resolution scans provided by TU Delft and the Mauritshuis Museum, and images from other sources. In order to streamline the resolution of all these images, they teamed up with a Deep Neural Network algorithm to upscale the images, thereby increasing the resolution by 300% and reducing visual noise. They designed a software system that could understand Rembrandt based on his use of geometry, composition and painting materials. A facial recognition algorithm was then used to identify and classify the most typical geometric patterns used to paint human features. The final 3D printed painting consists of more than 148 million pixels and is based on 168.263 Rembrandt painting fragments.²⁶⁵

Memories of Passersby I²⁶⁶ by Mario Klingemann at Espacio SOLO is an autonomous machine that uses a system of neural networks to generate a never-ending, never-repeating stream of portraits of non-existing people. To develop Memories of Passersby I, Klingemann trained his AI model using thousands of portraits from the 17th century to the 19th century. He created a Tinder-like application to accelerate the learning process and taught the machine his own aesthetic preferences, influenced by surrealist figures such as Max Ernst. The outputs displayed on the screen are not random or programmed combinations of existing images but unique, AI-generated artworks. The flow of images presented does not follow a predefined choreography but is the result of the AI interpreting its own output.

Computed Curation²⁶⁷ by Philipp Schmitt is a photo book that has been created by using machine learning and computer vision tools to curate a series of photos from an archive of pictures. The AI has been fed with pictures taken by Mr Schmitt. The various computer algorithms used caption each picture, categorise it using tags, analyse the composition, try to figure out the content, and then arrange the pictures in a way that would constitute a continuous flow.

²⁶¹ Openai, "A token is any symbol from a discrete vocabulary, <https://openai.com/blog/dall-e/>; for humans, each English letter is a token from a 26-letter alphabet. DALL·E's vocabulary has tokens for both text and image concepts", (accessed in July 2021).

²⁶² Youtube, <https://www.youtube.com/watch?v=RY-nK4ChLEQ>, (accessed in July 2021).

²⁶³ The Dali, <https://thedali.org/exhibit/dali-lives/>, (accessed in July 2021).

²⁶⁴ The Next Rembrandt, <https://www.nextrembrandt.com/>, (accessed in July 2021).

²⁶⁵ Another example is **Faceless Portrait of a Merchant**, one of the AI portraits produced by Ahmed Elgammal and AICAN).

²⁶⁶ Vimeo, <https://vimeo.com/298000366>, (accessed in July 2021).

²⁶⁷ Philippschmitt, <https://philippschmitt.com/archive/2018/work/computed-curation>, (accessed in July 2021).

Microsoft partnered with Iconem, a French company specialized in digitizing cultural heritage sites in 3D, to recreate the **Mont-Saint-Michel**²⁶⁸ site using a process in which AI processes hundreds of thousands of photos—taken by drones and cameras on the ground—to render models that are so precise, they look fully photo-realistic. HoloForge Interactive then used 3D renderings to bring Mont-Saint-Michel to life in holographic form.

The Metropolitan Museum of Art is using AI to simplify and scale the process of classifying and tagging each work of art. In particular, it has worked on **Art Explorer**²⁶⁹ which uses Microsoft's Cognitive Search to examine each artwork and automatically generates all the information needed to tag and classify the piece. Moreover, when a digital image is loaded into Art Explorer, Cognitive Search also brings to the surface which objects are depicted in the piece, which other artworks in the collection are visually similar, and what relevant information (geography, artist's history, etc.) it should pull in from the web. This data is then organized into a searchable index that unlocks insights, uncovers relationships between pieces in the collection, and grows the knowledge base around each piece online. No AI output is generated as such.

Thread Genius, a New York-based start-up founded in 2015 and recently acquired by Sotheby's auction house, uses complex algorithms to predict what art or luxury items clients may want to purchase based on previous purchases and searches.

Different as these AI solutions may be, they have in common that the algorithms used to create the output were **trained using datasets of protected works, such as photographs, images, art works**. The **output** of the AI-driven process can be transformations of user-submitted images or self-standing creations, which are very similar to (even undistinguishable from) human-made works of visual art and may be protected under copyright (depending on the role played by the human creator).

3.2.1.3. Case study: Generated Photos

Generated Photos

1. Technology

Generated Photos is an AI-based solution of a US-based company that generates "synthetic" (*i.e.* artificial) pictures of models.

Training data. The training of the algorithm requires Generated Photos to first build its '**real-life dataset**' by taking tens of thousands of photos of people in its professional studio. This is a crucial step to ensure that the input data are fully modelled and released as training data. Moreover, photographing models in a controlled environment allows Generated Photos to make sure that each face has a consistent look and quality. Furthermore, it created its own training data set to ensure diversity.

Generated Photos' website states that its '**real-life dataset**' is made up of the following photos.

²⁶⁸Inculture Microsoft, <https://www.microsoft.com/inculture/arts/le-mont-saint-michel-mixed-reality/>, (accessed in July 2021).

²⁶⁹ Art Explorer, <https://art-explorer.azurewebsites.net/search>, (accessed in July 2021).

92.000 portraits: Its models are people with different skin colours, various ages, and unique facial features/biometric characteristics, gender diversity, emotions. Moreover, Generated Photos uses different backgrounds, angles, poses etc.



8.000 manually masked photos, *i.e.* masked with Photoshop



77.000 AI-masked photos



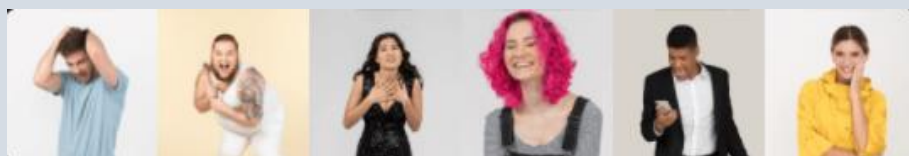
3.600 objects: e.g. clothes, food, plants etc.



1.700 surfaces: e.g. wood, metal, stone, grass etc.



6.800 real emotions: 17 emotions x 4 shooting angles x 100 models



After the shooting, a process of **manual labelling** of the photos is necessary.

Algorithms. The data set created is then used to train a generative adversarial network to produce faces that have never existed. The generative adversarial network used is called StyleGAN; this is open-source AI architecture introduced by Nvidia researchers.

The training of the AI system itself is time consuming and costly, and requires human intervention. One employee manually curates (on a full-time basis) the data, and two engineers build and train the AI tool.

Outputs. The current '**synthetic dataset**' that has already been generated is composed of 2 686 595 photos. The website allows the photos to be classified based on several parameters, such as (i) face (all, natural, beautified), (ii) head pose (front-facing, left-facing, right-facing), (iii) sex (female, male), (iv) age (young adult, adult, child, middle-aged person, infant), (v) ethnicity (white, black, Latino, Asian), (vi) eye colour (brown, grey, blue, green), (vii) hair colour (brown, black, blonde, grey), (viii) hair length (short, medium, long), or (ix) emotion (joy, natural). The user also has the opportunity to change the background colour of the photos.



Figure 13: Screenshot presenting a fraction of the 8,159 results of the “front-facing brown hair joy female with blue eyes”

Source: <https://generated.photos/>

2. Human’s creative influence on the output

Human creative influence can be exercised during the input phase, where the “**real-life dataset**” is generated by its photographers and models, as they play an important role concerning the input data.

After this stage, the remaining process, even if it is resource-intensive, does not offer much room for creative choice. The purpose of the entire process is not to generate a creative output reflecting the creative choices made by the employee working at Generated Photos, but rather to develop a massive database of pictures that consumers can use as appropriate in their projects.

3. Business model

The business model of Generated Photos is based on the monetisation of the photos they are generating. In this regard, Generated Photos offers different packages:

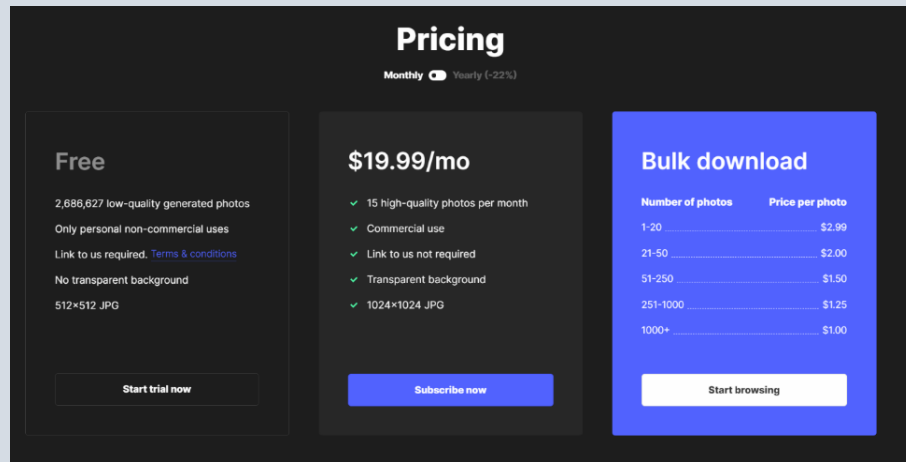


Figure 14: Screenshot from Generated Photos

Source: <https://generated.photos/>

Personal use: Generated Photos makes the photos in half-size resolution (512x512) available for free as individual downloads. However, users have then to provide a reference including a link to Generated Photos. If non-commercial photos by Generated Photos are used in a final work, then the final work needs to carry the proper reference. This includes compositions and remixes.

Commercial use. Photos can be used commercially (and without citation) only if a 'licence' is purchased. There are currently two options offered.

- **Bulk download.** This refers to a one-time download of photos. The images can then be kept without a time limit (perpetual licence). When creating final products with Generated Photos, the final product is bound by this licence.
- **API access.** This concerns authorisation for programmatic access to the photos. The photos can be used for up to 30 days after a licence has been cancelled. When creating final products with Generated Photos, the final product is bound by this licence.

Copyright reliance: Generated Photos states on its front page that "All images can be used for any purpose without worrying about copyrights, distribution rights, infringement claims, or royalties." They appear not to claim any IP rights on the photos generated by its AI system. However, it does not clarify the copyright status of the pictures generated, and they rely on the term "Licence", which is typically associated with the use of a protected subject matter (copyright, trademark, patent, etc.). It should lastly be noted that the terms and conditions themselves flag the uncertainty concerning the IP status of the photos: "[I]legal usage rights for content produced by artificial intelligence is a new, largely unknown domain. We are actively working in this area and as such, we reserve the right to amend our stated Use Licence as needed. We appreciate your understanding. »

Examples of uses of photos produced by Generated Photos can already be found in the gaming industry (e.g., to produce a crowd of zombies), in marketing material, as a dataset for ML, for medical or academic research, as avatars, or to achieve anonymity.

4. Link with copyright in the AI context

- > notion of work (see: 3.3.2.1)
- > protection of performance (see: 3.3.2.2)
- > protection of non-original photographs (see: 3.3.2.3)
- > performers moral and personality rights (see: 3.3.5.1.2)

3.2.2. Music

AI applications can also be used to generate music in an automated way. Different copyright issues can be examined. The AI input can of course consist of musical works (compositions or partitions) protected under copyright but, because musical works are of course meant to be performed, the training set may also contain other protected subject-matter, i.e. different performances and recordings of such musical works. Similarly, the AI output may be a musical creation (music score) that is meant to be performed by humans, or it may be a non-human performance of an AI-generated score that is readily available in a digital file. There may be implications for authors, performers and producers alike.

3.2.2.1. Global overview

In the music sector, the use of AI has been in practice for several decades. Already in the 1990s, David Bowie contributed to the development of an app called Verbasizer which took the literary source material and randomly recorded words to generate new combinations for lyrics.²⁷⁰ Another early AI tool is Autotune, which was released in 1997 to automatically alter pitch in vocal and instrumental music recordings and performances.²⁷¹

Currently, applications of AI in this domain include the following:

- **Creation process:** This refers to applications that assist the artist in the creative activity by providing assisted sound generation. The AI algorithm analyses data to find musical patterns (e.g. chords, tempo, length from various instruments, etc.) and automatically generates instrumental music based on musical rules and the choices of the user regarding instruments, tempo, musical notes, etc. The output can be new pieces (e.g. Artificial Intelligence Virtual Artist (AIVA) or Amper music) or the adaptation of existing compositions to fit audio-visual content (as on Muzeek), as well as autonomously generated new melodies. The amount of human input may range considerably for different AI solutions of this kind, from some that require human creative decisions to be operated, to those that automatically generate music with minimal user interaction. These AI solutions present a strong potential, and they could result in an important shift in the music industry. Indeed, if background music might not appear as significant from a cultural perspective, it

²⁷⁰ The Verbasizer was David Bowie's 1995 Lyric-Writing Mac App., <https://www.vice.com/en/article/xygxpn/the-verbasizer-was-david-bowies-1995-lyric-writing-mac-app>, (accessed in July 2021).

²⁷¹ H. A. HILDEBRAND, "Pitch detection and intonation correction apparatus and method," *U.S. Patent 5973252A* 1999.

is generating important value and revenue. AI solutions are also deployed as a tool to support creations as illustrated by the project “Hello World”.²⁷²

- **AI-based music production.** AI has also been used to provide musicians with a more affordable way to create a high-quality sound to make their music ready for distribution. Examples of this are AI-automated mastering, AI audio-mixing or AI-assisted plug-ins, step sequencers and drumming.
- **AI-based music analytics, curation and recommendations.** This refers to automatic classification of music for different purposes, such as music curation, music recommendations for the creation of automatically personalised playlists (e.g. Spotify), automated music classification (e.g. according to genre, instruments), support in the commercialisation of compositions (e.g. Muzeek) or forecasting of commercial success (e.g. Microsoft and KKBOX). These tools may be useful not only for end consumers but also for musicians, managers and production companies. For instance, they may be used for AI-supported matchmaking between artists and managers, to manage a music catalogue or to protect music rights (e.g. detection of unauthorised use of music protected by copyright, for instance as background music in online videos).

As regard performances, two main types of AI applications can be distinguished:

- **The automatic conversion of musical notes to musical sound:** Some AI solutions convert musical notes to sound to support the creation process. An example is “Magenta” a research project that has implemented tools that combine AI and music. They created NSynth (Neural Synthesizer)²⁷³, which is an AI algorithm that learns the characteristics of instrumental sounds and then creates new ones by blending acoustic qualities of the original sounds (e.g. a flute and a guitar as original sounds combined into a blend of both at once). These original sounds can be played digitally as if they were a new instrument.
- **The use of AI in performances:** The production and consumption of live performances is traditionally analogue. Although there are versions of events that are available online, these are just digitalised recordings of analogue events, without the use of AI. Although the application of AI to performances is very limited, there are a few state-of-the-art applications that have recently been developed. For instance, “Onsets and Frames” by Magenta²⁷⁴, performs an automatic transcription of recorded piano music into a digital piano roll, which can later be played on a synthesiser or just stored as a music sheet.

3.2.2.2. Examples of artificial intelligence solutions

Large companies such as Amazon, Microsoft or Sony are active in the field of AI in the music sector. Amazon created **Amazon AWS DeepComposer**²⁷⁵ which comprises hardware (MIDI keyboard) and software and works with Amazon Web Services (AWS) cloud platform. The keyboard allows the user to play a melody that is transformed into a music score through the use of a Generative AI.

²⁷² Hello World Album, <https://www.helloworldalbum.net/>, (accessed in July 2021).

²⁷³ Nsynth Super, (<https://nsynthsuper.withgoogle.com/>), (accessed in July 2021).

²⁷⁴ Magenta, Onsets and Frames, “Dual-Objective Piano Transcription”, <https://magenta.tensorflow.org/onsets-frames>, (accessed in July 2021).

²⁷⁵ Amazon, <https://aws.amazon.com/deepcomposer/>, (accessed in July 2021).

Microsoft Taiwan announced a strategic partnership with **KKBox** (a Taiwanese music streaming service) in which Microsoft's AI technology will be used to enhance KKBox's music streaming service²⁷⁶. This partnership will focus on (i) creating a predictive model to forecast the commercial success of a song (using data and AI); (ii) using Microsoft's AI technology to build an AI-assisted music arrangement system and an AI-assisted lyric generator and (iii) using AI to enhance personalized services, that is, reduce the operating cost²⁷⁷, improve the user's experience and develop services that create new monetization models.

Sony created **Sony's Flow Machines**²⁷⁸, an AI-assisted plugin that combines human creativity with AI-based music to help composers create various styles of melodies based on their own concepts. Using this tool users are also able to further edit the created output.

Spotify also uses AI and machine learning to enhance its service offer and to orient and recommend pieces of music to customers based on customer data insights (e.g. Discover Weekly).

Amper Music²⁷⁹ is a cloud-based platform that allows users to generate unique pieces of music with artificial intelligence. Users can create pieces of music by making simple choices (such as genre, length or mood) and may also edit and tweak the created piece. Amper is the AI that empowers users to create and customize original music for their content.

Melodrive is more focused on music for video games. It uses AI to compose a stream of original emotionally variable music in real-time. Melodrive composes and produce an original music piece that continuously adapts to user interaction and the scenario. It can be integrated into a game engine or as an application that provides music on top of games. Melodrive's target market is XR/game developers, players and streamers (where music is key for player engagement).

Slightly outside the music creation process, **Muzeek**²⁸⁰ is an application that uses AI to improve the adaptation and the commercialisation of compositions. Muzeek adapts existing pieces of music to fit audio-visual content. The AI can produce infinite variations of a music piece so that it follows the rhythm of images that it can recognise.

Finally, there are already compositions generated or partially generated by AI that are available to the public such as *Hello World!*, the Lost tapes of the 27 Club or Shimon.

Hello World!²⁸¹ is an album created through a collaboration between invited artists (such as Stromae, Michael Lovett or Camille Bertault) and the AI software **Flow Machines**. Flow Machines does not generate anything without human input. The user imports music files into a database to determine the style, the machine then makes a certain number of melodic and rhythmic proposals. Using an interactive interface, several round-trips between the proposals and the user's wishes are then made, until the user is satisfied with the output. The album *Hello World!* was released in January 2018 by the label Flow Records.

²⁷⁶ News Microsoft, <https://news.microsoft.com/2019/12/20/microsoft-and-kkbox-group-launch-global-strategic-partnership/>, (accessed in July 2021).

²⁷⁷ News Microsoft, "AI to identify different bitrates for video compression and transcoding, greatly reducing transmission bandwidth and saving storage space", <https://news.microsoft.com/2019/12/20/microsoft-and-kkbox-group-launch-global-strategic-partnership/>, (accessed in July 2021).

²⁷⁸ Flow Machines, <https://www.flow-machines.com/>, (accessed in July 2021).

²⁷⁹ Amper Music, <https://www.ampermusic.com/>, (accessed in July 2021).

²⁸⁰ Muzeek, <https://www.muzeek.com/>, (accessed in July 2021).

²⁸¹ Hello World Album, <https://www.helloworldalbum.net/>, (accessed in July 2021).

The **Lost Tapes of the 27 Club**²⁸² is an album partially created by an AI to raise awareness on mental health issues in the music industry. To create the album the company “Over The Bridge” fed an AI system with isolated hooks, rhythms, melodies and lyrics of musicians from the infamous 27 Club²⁸³. Based on the data received the AI system generated new hooks, rhythms, melodies and lyrics, and then an audio engineer took the music created and composed the album. In this particular case, a part of the music input was probably still protected under copyright. This raises questions about, for instance, whether this type of use of copyrighted matter (as input data) should be considered an exclusive right of the rightholder, whether it is covered by the exclusive reproduction right, or whether it should fall under an exception (e.g. TDM exceptions); and whether the output creation should be protected under copyright, related rights or fall in the public domain.

Shimon²⁸⁴ is a four-armed marimba playing robot. The robot can study large datasets from well-known musicians and then produce and perform its own original compositions by using deep learning.

Although little information was readily available on the technical process, these AI solutions mostly generate musical creations that are much like recorded human performances of human-created works. This suggests that their algorithms are set up and adjusted by relying on training data that not only contain musical works but also recorded performances. Both on the input and on the output sides, interesting questions arise on the use of copyright protected works (compositions) and other creations protected under related rights of performers and producers.

3.2.2.3. Case study: Artificial Intelligence Virtual Artist – an artificial intelligence composer

AIVA – an AI composer

1. Technology

AIVA (Artificial Intelligence Virtual Artist) is an “**AI composer**” developed by a Luxembourgish start-up which aim is to produce pieces of music.

Training data. AIVA was initially developed to compose classical music. Its algorithm was therefore trained on more than 30,000 scores from classical composers like Mozart, Bach, Beethoven, Wagner or Stravinsky. From a commercial perspective, this choice can be regarded as strategic, since this particular style of music is most often used in movies, games, commercials, and trailer soundtracks. From a legal perspective, most of these composers’ works also present the advantage of being part of the public domain.

More recently, AIVA has been trained with pieces of work that are of more recent styles. For the AI composer to be capable of generating rock music, a “re-training” of the algorithm was for instance necessary using “several hundred rock tracks”.

²⁸² Lost Tapes of the 27 club, <https://losttapesofthe27club.com/>, (accessed in July 2021).

²⁸³ Insider, <https://www.insider.com/27-club-celebrities-musicians-died-27-years-old-2017-9>, (accessed in July 2021).

²⁸⁴ Shimon Robot, <https://www.shimonrobot.com/>, (accessed in July 2021).

Lastly, it can be noted that when users of AIVA upload their own MIDI ("Musical Instrument Digital Interface") compositions (to influence the output produced)⁵, they automatically consent to provide AIVA, for perpetuity, with a worldwide, non-exclusive, commercial and transferrable licence to train AIVA's AI Systems with the uploaded MIDI.⁶

Algorithms. The development of AIVA algorithm has been made through deep learning and reinforced learning, as well as stochastic algorithms.⁷

2. Human's creative influence on the output

The users of AIVA are offered an experience of 'music composing', ranging from a minimal number of interactions (that can be reduced to 'editorial choices' of pieces automatically generated by the AI composer) to experiences in which the user plays a considerably more important role in the creation. In the latter case, the role of AIVA might be regarded as mere assistance.

Generation of pieces of music with no creative interactions from the user's side

In the case in which the user engages as little as possible, the interface only requires that the user indicate the 'style' of music that they need.



Figure 15: Screenshot from Generated Photos

Source: <https://generated.photos/>

The user can then select some options concerning the pace, the duration and the number of compositions they want the AI composer to generate:

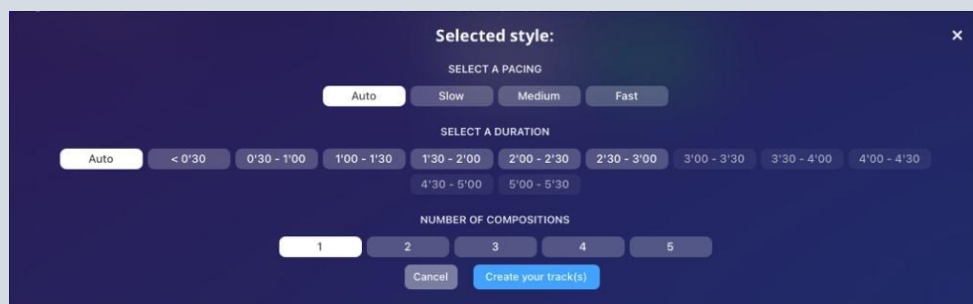


Figure 16: Screenshot from AIVA

Source: <https://www.aiva.ai>

The compositions are generated and added to the user's personal composition library:







TITLE	PARAMETERS	DURATION	CREATION DATE	
 New Composition #7	Fantasy, Bb Major, Symphonic Orches...	1:02	Feb 25, 2021	 ...
 New Composition #8	Modern Cinematic, A Minor, Piano & S...	0:57	Feb 25, 2021	 ...
 New Composition #9	Modern Cinematic, B Minor, String En...	0:59	Feb 25, 2021	 ...

Figure 17: Screenshot from AIVA

Source: <https://www.aiva.ai>

Finally, the user can proceed to the download of the composition proposed in several formats:



Figure 18: Screenshot from AIVA

Source: <https://www.aiva.ai>

The user's role might hence be reduced to the minimal interactions possible in order to generate a piece of music.

Generation of pieces of music with creative user's influence

Furthermore, AIVA offers the user an opportunity to engage in a significantly more creative experience.

First of all, the user can customize more precisely the preset of parameters within which AIVA composer will then operate. The parameters offered to customization will then depend on the style of music preselected.



Figure 19: Screenshot from AIVA

Source: <https://www.aiva.ai>

The user's creative engagement might be raised to an even higher level. Firstly, the user can choose not to pick a style of music, but rather to upload their composition (in a MIDI format). The AI composer will then analyse the uploads and suggest compositions approximating the user's style.⁸ It is also possible for the user to open a given song into an editor and adjust each element of composition (for instance, by adding instruments, modifying the tempo, the tonality, etc.).

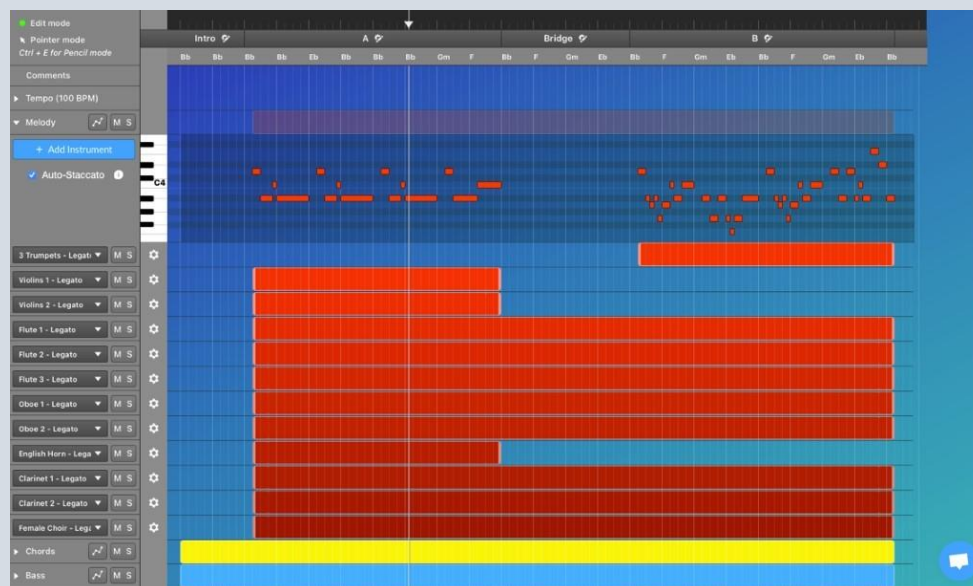


Figure 20: Screenshot from AIVA

Source: <https://www.aiva.ai>

2. Business model

Direct exploitation by AIVA of its outputs

AIVA Technologies is also in a position to generate revenue from the exploitation of AIVA-generated pieces of music. This is most likely for advertisement purposes. Indeed, AIVA has already released four albums on streaming platforms: 'Genesis' (2016, classical music) and 'Among the stars' (2018, classical/film music); '艾妮' (2018, Chinese music); and 'Romeo and Juliet' (2020, the soundtrack for a film). AIVA also generated a piece for the national holiday of Luxembourg (2020).

Moreover, 239 pieces of music are registered under the composer name 'AIVA' by SACEM.⁹ This registration should not mistakenly lead to the conclusion that SACEM recognises the composer's status as AI. In reality, a natural person working for AIVA Technologies falls under the name of 'AIVA', which in this case is used as a pseudonym. However, it means that the works produced by / with the help of AIVA may be considered a composition. Consequently, the firm developing AIVA can claim remuneration for the use of its works.

AIVA's lead investors are NetEase (China) and Kima Ventures (France). NetEase has been an investor since 2020 (which has enabled the recent international growth of AIVA).

AIVA as service for external users

The core of the AIVA business model is to sell access to its application to external users as a service. In this regard, AIVA proposes different pricing plans for three different categories of users: individuals; students and schools; and enterprises. The only pricing plan directly available without a subscription on the website is the pricing plan for individuals.

Three subscription models are offered by AIVA. Depending on the chosen subscription, the user will enjoy a different amount of possible monthly downloads, longer track durations, and a larger selection of possible download formats.

Free, Forever	Standard Monthly	Pro Monthly
€0	€15 / month + VAT, if applicable	€49 / month + VAT, if applicable
No credit card required	Billed Monthly	Billed Monthly
For beginners who want to use compositions for non-commercial use cases only, and don't mind giving credit to AIVA.	Recommended for content creators who want to monetize compositions only on Youtube, Twitch, Tik Tok and Instagram.	Recommended for creators who want to own the copyright of their compositions, and monetize without restrictions.
<ul style="list-style-type: none"> Copyright owned by AIVA No monetization Credit must be given to AIVA 3 downloads per month Track durations up to 3 minutes Download MP3 & MIDI formats 	<ul style="list-style-type: none"> Copyright owned by AIVA Limited monetization No need to credit AIVA 15 downloads per month Track durations up to 5 minutes Download MP3 & MIDI formats 	<ul style="list-style-type: none"> Copyright owned by YOU Full monetization No need to credit AIVA 300 downloads per month Track durations up to 5'30 mins Download ALL file formats Export high quality WAV files

Figure 21: Screenshot from AIVA

Source: <https://www.aiva.ai>

The chosen plan also affects the type of licence granted by AIVA.

- Free plan.** AIVA grants the user a 'non-exclusive, non-transferable and non-commercial license to use, modify and distribute the MIDI and Audio Composition, in any Content that the Licensee holds rights over'. The copyright of the composition is claimed to be 'owned by AIVA'. In the event of public use of the composition, the user must give credit to AIVA.
- Standard monthly plan.** This is the free plan plus 'non-transferable and commercial license to monetise the MIDI and Audio Composition on a limited set of third-party websites: YouTube, Twitch, Tik Tok and Instagram'. The

standard plan grants a limited commercial licence to the user. The copyright of the composition is still claimed to be 'owned by AIVA'. The user does not need to give credit to AIVA when using the composition on the given websites.

3. **Pro monthly plan.** 'Licensor assigns, grants and conveys all copyrights of the MIDI and/or Audio Composition to Licensee.' With the subscription to this plan, AIVA agrees to transfer the copyright of the composition to the user.

3. Link with copyright in the AI context

- > notion of work (see: 3.3.2.1)
- > performance of AI autonomously generated outputs (see: 3.3.2.2)
- > protection of AI autonomously generated outputs as phonogram (see: 3.3.2.3)
- > presumption of authorship (see: 3.3.3)
- > moral rights (see: 3.3.5)

3.2.3. Audiovisual and Film

The AI applications in the audiovisual sector are, at this stage, more focused on specific parts of the audiovisual production and commercialisation process. Mainstream audiovisual productions are arguably too complex to be entirely autonomously generated by AI applications, meaning that – for the time being – the human contribution is indispensable and their protection under copyright and related rights does not raise any particular difficulties. By contrast, the AI applications used for automating parts of the audiovisual creative process are based on the processing of human creations, including writings, still and moving images, potentially protected under copyright, performers' rights and/or producers' rights. Where AI applications are used to optimize the commercialisation or distribution process, for predicting the success of a new audiovisual production, for recommending similar content or for classifying audiovisual content, the input data do not necessarily contain protected works, performances or recordings.

3.2.3.1. Global overview

There is significant overlap between AI applications in the audiovisual and film sectors and these applications in both the visual arts sector (concerning, for instance, the different types of image-processing techniques mentioned above) and the music sector (e.g. creation of background music in a way that human choices are not always required). However, additional fields of application for AI-based technologies do arise, as there are typically more agents involved in, for example, film production (screenwriting, directing, camerawork, production, acting, etc.) than in the visual arts or music (where in both cases a work of art may be attributable to only one agent). Therefore, as the creation of audiovisual content or film involves a broad spectrum of different processes and stages, it is impossible to foresee which types of applications and purposes AI-based technology may find in the sector.

The following bullet points thus should be seen as providing a non-exhaustive list of possible **AI-based applications for the generation of audiovisual content and film**:

- **Script-writing:** Some applications of AI-based scriptwriting exist, such as the example of *DeepStory* by the company *ScriptBook*, with which movie scripts can be

generated and downloaded for further editing and use.²⁸⁵ In 2016, the screenplay for the movie *Sunspring* was written by AI *Benjamin* (and the AI is listed on the internet movie database as the writer, whereas human actors are listed in the categories of director and actors).²⁸⁶

- **Animation:** Animation is the process through which a sequence of images is arranged such that they appear as moving images. As this process is now overwhelmingly carried out with the aid of computers, the possibility emerges for AI to at least assist humans in various stages of the process. For example, movie studio Laika has, in cooperation with Intel, used AI to accelerate the animation process in stop-motion productions. Furthermore, AI applications have also been used for several more specific tasks in the animation process (e.g. *DeepMotion* for real-time animation of moving persons, or different AI-based animation software companies, e.g. *Raw Shorts*, *Animaker* and others, targeting companies to produce customized advertisements).
- **Post-production:** A myriad of AI applications that were specifically designed to aid editors during the post-production process has also been developed. This ranges from the automated generation of meta-data (e.g. *EditShare's EFS* and *Avid Media's Composer* facial/object recognition software), quality upscaling of images (where, additionally to the image processing techniques mentioned above, cross-temporal relationships between images are learned by the AI and may subsequently be used to smoothen the moving images; e.g. *Topaz Video Enhance AI*), or applications that fulfil specific tasks such as *CrumplePop's WindRemover AI* that removes background wind or the example of *Flawless AI*, which generates automated lip-synced visualisations (see case study below)
- **Other applications:** Other AI applications in this area are e.g. **deepfakes** (the transformation of existing audiovisual format into another synthetic type of content of similar appearance but different semantic content). While there are legitimate usages of deepfakes in e.g. the entertainment sector, they may also have been created with malicious intent, which raises the urgency of the legal treatment of deepfakes, where important intersections between copyright law and e.g. privacy and data protection law are likely to arise. Future applications include applications concerning **virtual reality** and **augmented reality**.
- **Outside the creation process,** AI application may be used to benefit the wider industry, e.g. by aiding the decision-making process for investments, by predicting commercial success, likely audience satisfaction across target groups etc. based on the script and/or movie analysis (e.g. *ScriptBook*), or by making of recommendations and promotions directed towards the users of streaming platforms (e.g. on platforms such as *Netflix* or *YouTube*).

3.2.3.2. Examples of artificial intelligence solutions

In the audio-visual sector, it seems that AI solutions are generally used to enhance decision making and recommendations. **ScriptBook**²⁸⁷ is a platform for script analysis, financial forecasting and decision support. ScriptBook uses an AI tool to guide the users in their decision making through automated script analysis by giving information about the script

²⁸⁵ Deepstory, <https://www.deepstory.ai/#/>, (accessed in July 2021).

²⁸⁶ Thereforefilms, <https://www.thereforefilms.com/films-by-benjamin-the-ai.html>, (accessed in July 2021).

²⁸⁷ Scriptbook, <https://www.scriptbook.io/#/>, (accessed in July 2021).

and financial forecast to maximize the potential for critical and commercial success (by giving information on script's DNA, financial forecasting, content validation and audience insights, ...).

A similar platform is **Cinelytic**²⁸⁸, which uses AI to support film studios and independent content companies in making faster and better-informed decisions. The start-up has compiled a wide range of data that helps the users understand in real-time how their choices (actors, scripts, release date, marketing, greenlight, finance options, ...) can impact the risk profile of projects. Cinelytic signed a partnership with the renowned studio Warner Bros²⁸⁹.

Other tools are also used in the same field such as Largo.AI, Merlin or Vault. **Largo.AI**²⁹⁰ uses AI to support the traditional content creation workflow and to help actors of the sector in their decision-making. **Merlin**²⁹¹ is a system, used by 20th Century Fox²⁹², that uses AI and machine learning to match particular films to particular genres and audiences. **Vault's** RealDemand AI platform²⁹³ analyses thousands of key elements of a story, an outline, a script, castings, a trailer and marketing efforts to determine/predict the product's audience.

On the creation side of audio-visual content, **Ginger.Studio**²⁹⁴ offers an AI-powered tool to assist users in the creation of their audio-visual content. Its face and word recognition algorithms allow the user to create and edit audio-visual content more efficiently (simple and quick selection of video sequences by the user).

Netflix²⁹⁵ also uses AI and machine learning to, among other things, to power its recommendations algorithms, shape its catalogue of movies and TV shows and optimise the production of Netflix original movies and TV shows.

Through their partnership, **YouTube** and **Google Creative Labs** created an infinite interactive music video using TensorFlow machine learning and over tens of thousands of fan covers of the song 'Bad guy' by Billie Eilish. The interactive AI experiment collects thousands of covers and blends them with the help of machine learning (each video is aligned on the original). In the infinite music video, the original video starts playing in the centre and other videos materialise around it; if the viewer clicks on another video (or hashtags presented below the player), the video will seamlessly transition to the selected video and pick up the song where it left off in the previous video. Viewers can also adapt the kind of videos presented by selecting hashtags below the player.

The first short film written completely by an AI is **Sunspring**²⁹⁶. The movie was created during the event "48-hour Film Challenge" at the Sci-Fi London film festival. The participants were given a set of prompts (props and lines) that had to appear in the short film. The participants supplied the information (the prompts to be used) to the AI writer

²⁸⁸ Cinelytic, <https://www.cinelytic.com/platform/>, (accessed in July 2021).

²⁸⁹ Theverge, <https://www.theverge.com/2020/1/9/21058094/ai-film-decision-making-warner-bros-signs-cinelytic>, (accessed in July 2021).

²⁹⁰ Largo, <https://largo.ai/>, (accessed in July 2021).

²⁹¹ Cloud Google, <https://cloud.google.com/blog/products/ai-machine-learning/how-20th-century-fox-uses-ml-to-predict-a-movie-audience>, (accessed in July 2021).

²⁹² The verge, <https://www.theverge.com/2018/11/2/18055514/fox-google-ai-analyze-movie-trailer-predict-success-logan>, (accessed in July 2021).

²⁹³ Vault AI, <https://www.vault-ai.com/>, (accessed in July 2021).

²⁹⁴ Gingalab, <https://gingalab.com/>, (accessed in July 2021).

²⁹⁵ Research Netflix, <https://research.netflix.com/research-area/machine-learning>, (accessed in July 2021).

²⁹⁶ Arstechnica, <https://arstechnica.com/gaming/2016/06/an-ai-wrote-this-movie-and-its-strangely-moving/>, (accessed in July 2021).

which created a screenplay out of it. The work was authored by an LSTM (long short-term memory) recurrent neural network self-named Benjamin.

The **Netherlands Institute for Sound and Vision** ("NISV") is one of the largest audio-visual archives in Europe with around one million hours of content (and still growing). NISV has applied AI in services of public broadcasters and focuses more particularly on speaker labelling (a technology that matches voices with persons' names – politicians or well-known celebrities – and marks where their voices appear in a programme) and entity extraction (an information extraction technique that identifies key elements from text and then classifies them into predefined categories). It relies on its own open collection of movies to feed its AI.

The **Time Machine**²⁹⁷ is another cultural project that uses AI. It is an international collaborative project that builds a large-scale historical simulator mapping thousands of years of European historical, social, cultural and geographical evolution. It is based on an information infrastructure that processes millions of historical documents, museum collections, geo-historical datasets, as well as the growing amount of 'digitally native' heritage produced today. In doing so, it helps simulate possible futures and past scenarios with predictive techniques and interactive 3D visualisations.

The diversity of these examples illustrates that the AI solutions can be used for different specific tasks in the creation, production and distribution of audiovisual content. Depending on the function of the AI tool, the input data may consist of copyright protected works (scripts, photos, moving images) and protected performances and recordings or, on the contrary, data or metadata that are not protected under copyright and related rights.

3.2.3.3. Case Study: Flawless AI – TrueSync

Flawless AI

1. Technology

Flawless AI²⁹⁸ is a UK-based start-up, co-founded by film director Scott Mann and software developer Nick Lynes. The company uses deep neural networks to create lip-synced versions of movies in multiple languages.

In their paper entitled 'Neural Style-preserving Visual Dubbing', the group of researchers explains how their style-preserving visual dubbing approach maintains the signature style of the target actor, including their person-specific idiosyncrasies.²⁹⁹

²⁹⁷ Time Machine, <https://www.timemachine.eu/>, (accessed in July 2021).

²⁹⁸ Flawlessai, <https://www.flawlessai.com/>, (accessed in July 2021).

²⁹⁹ C. THEOBALT e.a., "Neural Style-Preserving Visual Dubbing", *SIGGRAPH Asia*, <https://arxiv.org/pdf/1909.02518.pdf>, (accessed in July 2021).

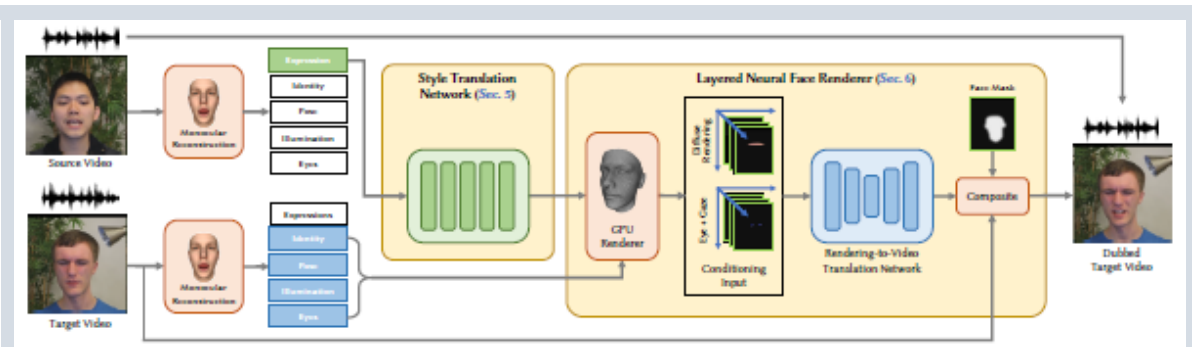


Figure 22: Overview of the style-preserving visual dubbing approach³⁰⁰

Source: Flawless AI

The process for generating AI-driven dubbing performances goes as follows. It starts with a source actor (i.e. the dubber) narrating the dialogue in the required language, similarly to dubbing.

Next, the AI system analyses both the source actor's performance and the target actor's performance and records a 3D face model by reconstructing the facial expression parameters. Specifically, it learns the subtle nuances of how the actor's lip movements relate to other facial features, including head position and eye movements. Hence, it captures the actors' performances digitally from the two-dimensional films and transforms them into a 3D model.

Then, the AI system transfers, through a recurrent generative adversarial network, the source actor's expression (lip movements) to the target actor's expression while the latter maintains the idiosyncracies and style. This network was trained in "an unsupervised fashion from unpaired training data"³⁰¹. The dataset is composed of "a collection of sequential expression parameters from individual videos"³⁰².

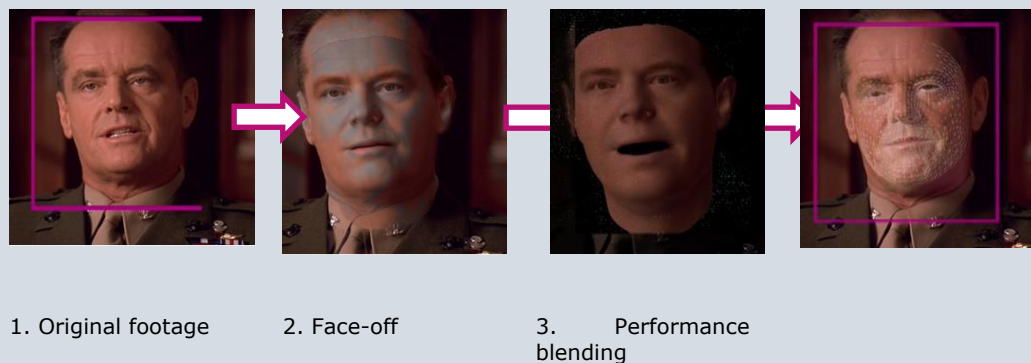


Figure 23: TrueSync technology applied to Jack Nicholson's performance in "A Few Good Men"

Source: screenshots from <https://www.flawlessai.com/product>

Finally, the AI tool synthesizes and generates a photorealistic video portrait that is rendered on the target actor's face. If necessary, the output is adjusted by a human visual effects artist.

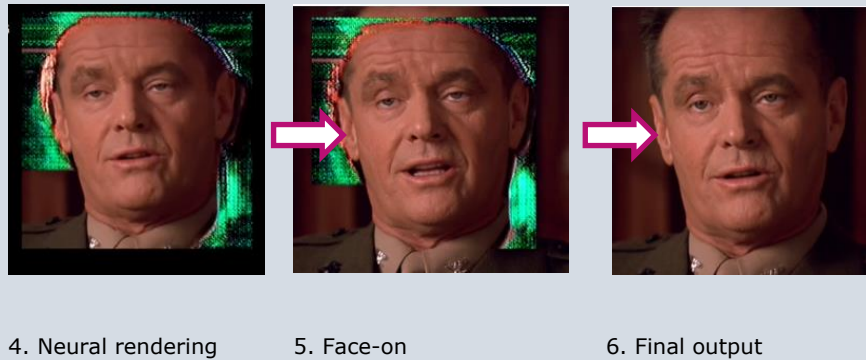


Figure 24: TrueSync technology applied to Jack Nicholson's performance in "A Few Good Men"

Source: screenshots from <https://www.flawlessai.com/product>

Flawless AI is thus able to generate a modified version of the target actor's original performance by changing their voice and facial expression to suit an entirely different language while retaining their performance and style.

As an illustration, reference can be made to the short clips on the Flawless AI website in which the technology is featured.³⁰³ These videos include Tom Cruise and Jack Nicholson conversing in fluent French in the 1992 film "A Few Good Men," Robert De Niro speaking German in the 2015 film "Heist," and Tom Hanks in the 1994 film "Forest Gump" speaking German, Spanish, and Japanese.

2. Human's creative influence on the output

Flawless AI requires human intervention at the input and output sides of the process.

First, the training data supporting the training of the AI model are composed of videos of individuals. In this regard, the scientific paper describing the technology explains that: *"facial expression styles consistently captured by approximately five-minute-long videos are typically sufficient to train our style translation network."*³⁰⁴

Second, and this time necessarily relevant under the performers' right, the source actor's performance (especially facial expression and voice) is used to generate the output.

Lastly, the output still requires some manual editing by human visual effects artists.

³⁰⁰ Ibidem.

³⁰¹ Ibidem.

³⁰² Ibidem.

³⁰³ Flawlessai, <https://www.flawlessai.com/product>, (accessed in July 2021).

³⁰⁴ *Op.cit.*

3. Business model

Flawless AI is in discussions with major streamers and studios about possible deals, without further details. There is no projected schedule yet for the release of a commercial television show or movie using Flawless AI's technology.³⁰⁵

Little has been disclosed about the pricing model, though it has been indicated that the price will be based on the number of translations and the length of the film.³⁰⁶

Eventually, it is said that Flawless AI technology, trained on the body movement patterns of actors and the captured environment of an actor in a given scene, will eventually allow the editing of certain film scenes to correct mistakes or better conform to a director's vision rather than bringing the cast and crew together for expensive reshoots.³⁰⁷

4. Link with copyright in the AI context

-> protection of performance as input (see: 3.3.2.2.1)

-> performers' moral rights and personality rights (see: 3.3.5.1.2)

3.2.4. Gaming

As for the audiovisual sector, AI applications are not used for generating entire video games but as a support for automating certain operations within the creative or commercial processes. The input used for optimising the AI applications may contain copyright protected works (software code, images) but also subject matter protected under the related rights (film fixations).

3.2.4.1. Global overview

Historically, gaming has had strong ties to AI. However, the term 'AI' has been defined rather broadly in the context of computer games, namely as the possibility of generating responsive, adaptive or intelligent behaviour in non-player characters. This exceeds the narrower definition of AI (which is intended to encompass state-of-the-art machine learning algorithms or different types of neural nets) and instead includes, for example, traditional chess computers (of course, this idea of intelligent computer behaviour is the essence of the famous Turing Test and dates back to the earliest days of modern computer science).

Regarding the current day video game industry, AI and machine learning are likely to be most beneficial to game developers **during the creation phase**. Game developers already experiment with AI to generate game content (e.g. *AI Dungeon*, *No Man's Sky*, *Yavalath*), to automate the bug-fixing process (e.g. modl:test) or to create game animations. For instance, the production of game animations is typically heavily labour intensive; the

³⁰⁵ The Verge, <https://www.theverge.com/2021/5/18/22430340/deepfake-dubs-dubbing-film-tv-flawless-startup>, (accessed in July 2021).

³⁰⁶ S. BLAKE, "You Talking to Me? This New Startup Uses Deepfake Technology for Movie Dubbing", [Flawless Launches to Use Deepfake Tech for Movie Dubbing - dot.LA](https://dot.la/flawless-ai-dubbing-2652865135.html?web=1&wdLOR=cA3AAE24F-1D36-F846-B4AB-5077F8A09330), dot.LA., <https://dot.la/flawless-ai-dubbing-2652865135.html?web=1&wdLOR=cA3AAE24F-1D36-F846-B4AB-5077F8A09330>, (accessed in July 2021).

³⁰⁷ Spectrum, <https://spectrum.ieee.org/tech-talk/artificial-intelligence/machine-learning/ai-modifies-actor-performances-for-flawless-dubbing>, (accessed in July 2021).

advantage of AI applications is therefore that less code will have to be written manually by game developers. Advances in AI applications in virtual reality and augmented reality may also be exploited by game developers.

Furthermore, AI may be used to **enhance the in-game experience**, as AI opens up the possibility of non-player characters exhibiting responsive learning that produces kinds of behaviour that are similar to the behaviour of human players. Therefore, the aim of AI is to provide a gaming experience similar to the experience of playing against other human players (e.g. modl:play, which is a software product for AI-based training bots that can act as player stand-ins).

Finally, AI might also be used for a range of **additional purposes** in the video game industry. It has, for example, been applied to analyse players' behaviour (e.g. the start-up **Spirit AI** provides an AI solution that detects bullying in video games).

3.2.4.2. Examples of artificial intelligence solutions

Among the current applications of AI to video games is, for instance, **No Man's Sky**³⁰⁸, developed by Hello Games, which presents a universe in which every rock, flower, tree, creature, and planet has been 'procedurally generated' to create a vast and diverse play area. The tens of millions of planets that comprise the universe are all unique. Each is generated when a player discovers it and is subject to the laws of its respective solar systems and vulnerable to natural erosion. The multitude of creatures that inhabit the universe dynamically breed and genetically mutate as time progresses.

Spirit AI has created '**Ally**'³⁰⁹ which uses AI and natural language processing (NLP) – trained on millions of messages – to combat toxic behaviour within game communities. It helps moderators to identify bad actors and helps keep online communities safe and healthy. Ally can also intervene automatically via webhooks that trigger muting, rule reminders, or suspensions.

Yavalath³¹⁰ is an abstract board game for two or three players, invented by a computer program called LUDI. LUDI created the game by taking the rules of existing games and scrambling them into new combinations using genetic programming (GP) techniques of crossover and mutation. It was tested through self-play trials and assigned a quality score based on its estimated potential to interest human players; hence the complete process of design, testing and evaluation are entirely automated. Ludi created a unique name 'Yavalath' for this game using a Markovian process seeded with Tolkien-style words.

Another application of AI in video games is **Remako HD Graphics Mod** which is a mod that completely revamps the pre-rendered backgrounds of the classic JRPG Final Fantasy VII. For this upscaling, it uses AI neural networks which tries to emulate the detail the original renders would have had.

Another use of AI in this sector is that of **Modl.ai**³¹¹, which provides AI-aided tools for video games studios to accelerate game development and testing, as well as to enhance player enhancement. Examples of their AI-based services are cheating detection in online games (e.g., against cheating bots, hacks and cheat tools), the creation of game playing bots (e.g., to ensure that online gamers have always someone to play against), the support in content generation (e.g., to kick-start and accelerate creative processes in the design

³⁰⁸ No Man's Sky, <https://www.nomanssky.com/>, (accessed in July 2021).

³⁰⁹ Spiritai, <https://www.spiritai.com/ally/>, (accessed in July 2021).

³¹⁰ Cambolbro, <http://cambolbro.com/games/yavalath/>, (accessed in July 2021).

³¹¹ Modl.Ai, <https://modl.ai/>, (accessed in July 2021).

of the game), and even the support in understanding player motivations in real-time so that the game can react accordingly to create meaningful player engagement.

'Steam Labs' is another case of the use of AI in the video game sector. Launched in 2019 on Valve's gaming platform Steam, is dedicated to experiments around discoverability, video, machine learning and more. One of these experiments is called "**Interactive Recommender**"³¹², which is a machine learning model that is trained to recommend games based on a user's playtime history, along with other salient data based on many millions of Steam users and many billions of play sessions. Hence, the model infers properties of games by learning what users play, not by looking at user-curated metadata (e.g., user-provided tags, user-curated lists, aggregate review scores, and sales data). The purpose is to recommend games based on the actual playing habits and interests of players with broadly similar profile.

3.2.4.3. Case Study: AI Dungeon

AI Dungeon

AI Dungeon is a text adventure video game developed by *Latitude*³¹³, which presents the specificity of being infinite since its content is continuously generated by the AI. A more in-depth description of the game is provided below, but the underlying idea of the video game is that the player – confronted with a written scenario – is free to decide how to react. All they have to do is to react – via a written interface – to the content generated by the AI. The game will then respond to the player's input and continue the scenario accordingly. Latitude summarized it as follow: "You enter your actions and the AI will continue the story..."

1. Technology

Algorithms. The versions of AI Dungeon rely successively on the natural language processing neural networks GPT-2 (*Dungeon 2*) and GPT-3 (*Dragon*).³¹⁴ Both models were developed by *Open AI* (originally a non-profit organization)³¹⁵ and were – at the time of their respective releases – technological breakthroughs. Released in 2020, GPT-3 was the largest existing AI model, with 175 billion parameters trained with 570 gigabytes of text content.³¹⁶ Although the previous model developed by Open AI was available in open-access, GPT-3 is only accessible via an API allowing *Open AI* to generate revenues.³¹⁷ At present, we do not know what it costs *Latitude* to rely on this model. *Latitude* simply explains that "*The Dragon model uses one of the most advanced AIs in the world [GPT-3] and is extremely expensive to run.*"³¹⁸

³¹² Steam, <https://store.steampowered.com/recommender>, (accessed in July 2021).

³¹³ Latitude, <https://latitude.io/about/>, (accessed in July 2021).

³¹⁴ The abbreviation GPT stands for "Generative Pre-trained Transformer".

³¹⁵ Technology Review, <https://www.technologyreview.com/2020/02/17/844721/ai-openai-moonshot-elon-musk-sam-altman-greg-brockman-messy-secretive-reality/>, (accessed in July 2021).

³¹⁶ Arxiv, <https://arxiv.org/abs/2005.14165>, (accessed in July 2021); Hai Standford, <https://hai.stanford.edu/blog/how-large-language-models-will-transform-science-society-and-ai>, (accessed in July 2021).

³¹⁷ Openai, <https://openai.com/blog/openai-api/>, (accessed in July 2021); as a matter of precision, it should be noted that GPT-3 was recently exclusively licensed to Microsoft: see <https://www.technologyreview.com/2020/09/23/1008729/openai-is-giving-microsoft-exclusive-access-to-its-gpt-3-language-model/>, (accessed in July 2021).

³¹⁸ Play Aidungeon, <https://play.aidungeon.io/main/frequentlyAskedQuestions>, (accessed in July 2021).

Training data. Although GPT models are pre-trained, their training can be fine-tuned to the specific task they will be asked to accomplish. In the case of AI Dungeon, regarding the first version of the game, the training of GPT-2 was performed on a dataset of "30 MB of text adventure stories (...) scraped from "chooseyourstory.com"³¹⁹. Chooseyourstory.com is a website where an online community of "Choose-Your-Own-Adventure style storygames" fans create and share their own story games.³²⁰ For the moment, we do not know if a "fine-tuning" of GPT-3 was necessary.

2. Human's creative influence on the output

When starting an adventure, the users of AI Dungeon have to make several choices concerning the initial setting of the adventure. The interface itself offers some options, with pre-defined worlds and atmospheres, including "Fantasy", "Mystery", "Apocalyptic", "Zombies" etc. Nonetheless, it is possible to start completely from scratch by clicking on "custom". In this case, the player is asked to describe these contextual elements themselves.

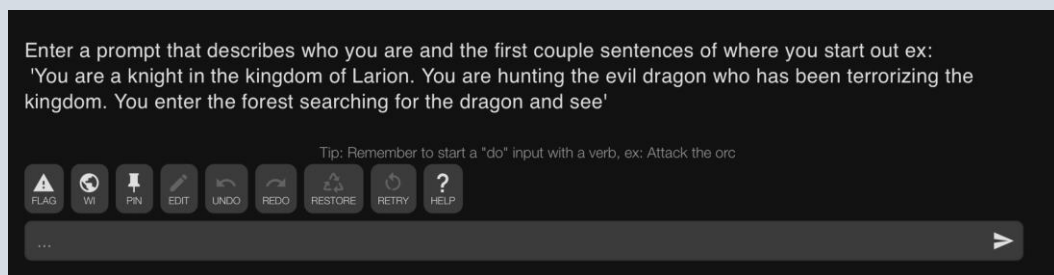


Figure 25: Screenshot from AI Dungeon

Source: <https://play.aidungeon.io>

By way of example, we have submitted the following prompt:

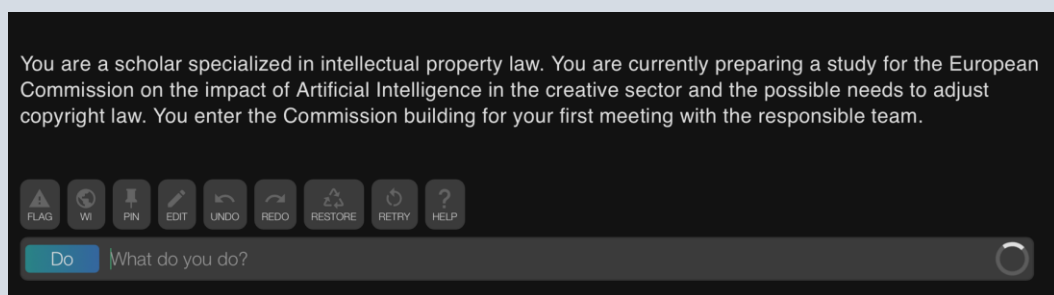


Figure 26: Screenshot from AI Dungeon

Source: <https://play.aidungeon.io>

³¹⁹ Interview of AI Dungeon creator, <https://towardsdatascience.com/the-creator-of-ai-dungeon-2-shares-gpt-2-finetuning-advice-e5800df407c9>, (accessed in July 2021).

³²⁰ Choose your Story, <http://chooseyourstory.com>, (accessed in July 2021).

The AI then comes up with a sequel to the story (own highlighting):

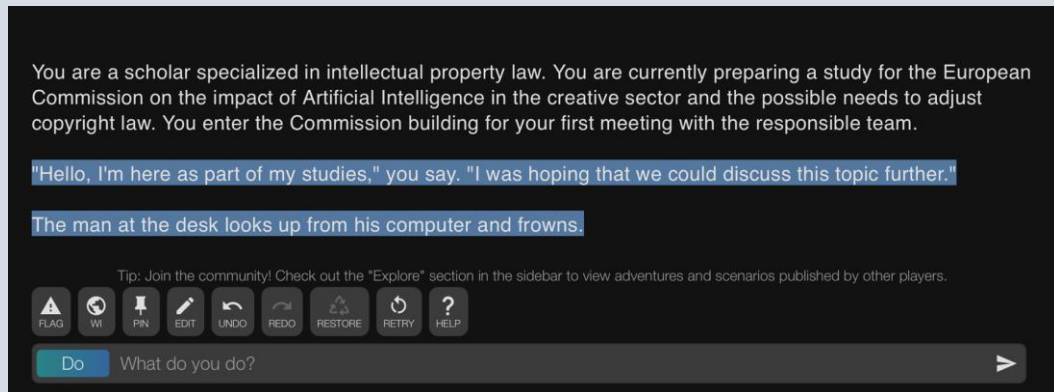


Figure 27: Screenshot from AI Dungeon

Source: <https://play.aidungeon.io>

In replying to the AI-generated scenario, the user can choose between three different interaction methods (three types of input):

1. "Do": the player describes an action that they want their characters to perform.
2. "Say": the player writes the dialogue, which they want their characters to deliver.
3. "Story": the user writes a description of what he wants to happen during the adventure.

In the following two screenshots, you can see how the AI reacts to our input (still our own highlighting).

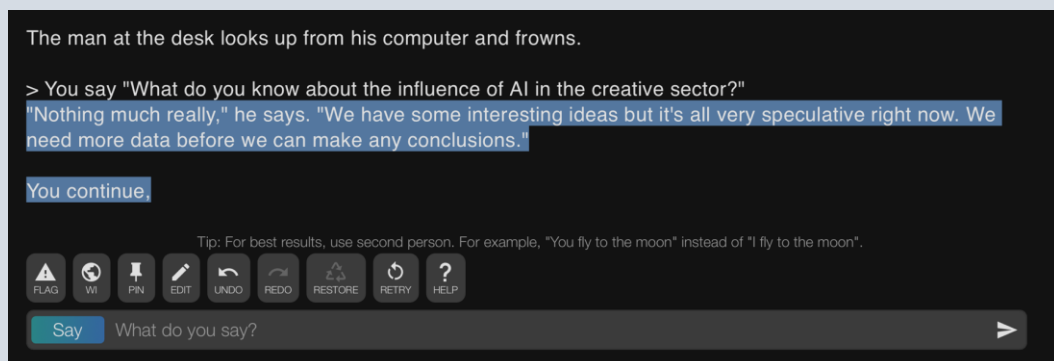


Figure 28: Screenshot from AI Dungeon

Source: <https://play.aidungeon.io>

> You say "What do you know about the influence of AI in the creative sector?"
 "Nothing much really," he says. "We have some interesting ideas but it's all very speculative right now. We need more data before we can make any conclusions."

You continue,

> You say "But do you think it might have an impact on copyright law? "
 "How did you know?" the man asks surprised. "Yes, there are likely to be some changes," he says.

Figure 29: Screenshot from AI Dungeon

Source: <https://play.aidungeon.io>

The creativity of the user, therefore, has a strong influence on the course of the adventure. However, it is the AI that still creates the content of whatever happens next in the adventure. Yet, the user can control what the AI has generated. Indeed, there is an "undo" button that allows the user to go back to what the AI has created.

Indeed, there is an "undo" button that allows the user to go back to what the AI has created. Besides, there is a "pin" button that allows the user to change what has been created by the AI. The user can also press the enter key while leaving the text blank, which will cause the AI to generate more stories.

3. Business model

The business model of *Latitude* relies on several formulae. First of all, Latitude makes a free version of the game available to the users;³²¹ players then have the possibility to buy "scales".³²² These scales enable users to unlock certain features of the game, such as new worlds.³²³ Finally, there are also several monthly subscription plans. It is only by purchasing a gold or platinum subscription that the player can access the latest version of the game trained on GPT-3.³²⁴

Pick your Premium

1 week free not available for users who have already tried Premium.

<p style="text-align: center; font-size: small;">1 week free</p> <p style="text-align: center;">Silver</p> <p style="text-align: center; font-size: small;">\$4.99/month after free trial</p> <ul style="list-style-type: none"> ✓ 100 Monthly Scales* ✓ Unlimited Griffin ✓ Scripting ✓ Randomness ✓ Length ✓ Author's Note ✓ Audio Narration 	<p style="text-align: center; font-size: small;">1 week free</p> <p style="text-align: center;">Gold</p> <p style="text-align: center; font-size: small;">\$9.99/month after free trial</p> <ul style="list-style-type: none"> ✓ 200 Monthly Scales* ✓ Unlimited Griffin ✓ Dragon ✓ Scripting ✓ Randomness ✓ Length ✓ Author's Note ✓ Audio Narration 	<p style="text-align: center;">Platinum</p> <p style="text-align: center; font-size: small;">\$29.99/month no trial</p> <ul style="list-style-type: none"> ✓ 500 Monthly Scales* ✓ Unlimited Griffin ✓ Unlimited Dragon ✓ Scripting ✓ Randomness ✓ Length ✓ Author's Note ✓ Audio Narration
--	--	--

Silver

\$4.99/month
after free trial

Gold

\$9.99/month
after free trial

Platinum

\$29.99/month
no trial

Figure 30: Screenshot from AI Dungeon

Source: <https://play.aidungeon.io>

Latitude has also run a Patreon campaign to support the development and operations of the game and cover associated costs. In February 2021, it was announced that Latitude had raised \$3.3 million in seed funding (led by NFX, with participation from Album VC and Griffin Gaming Partners) to take AI content creation beyond the purely text-based nature of the AI dungeon to new levels of immersion and interactivity.³²⁵

4. Link with copyright in the AI context

-> notion of work (see: 3.3.2.1)

-> TDM exceptions (see: 3.3.6)

3.2.5. Transversal analysis of the use of AI

This section discusses aspects and issues that are common to the above-mentioned sectors.

3.2.5.1. Different usages of artificial intelligence in the creative industries

AI solutions are particularly successful and useful in performing tasks that are either **repetitive or time-consuming for humans** (such as searching through a large data set and extracting the features of the data to identify patterns or classifications), but also when humans lack the necessary skills, instruments or investments to perform a task (for instance, the lack of music knowledge to compose a melody in a certain style, or translating from one language to another). The post-production of creative content often involves a set of repetitive actions to be performed according to different processes and parameters, therefore, post-production workflows are clear candidates for continuing gains in efficiency and efficacy linked with the increased use of AI.

Additionally, AI solutions also have a lot to offer in the **assistance during the creative phase**. Here, the human intervention varies significantly from AI assisting the creative process for instance by suggesting creative options (i.e., as a source of inspiration), over performing technical tasks based on abstract commands from the user (i.e. give a more joyful rhythm to a melody), to entirely generating the output in an automated manner.

3.2.5.2. Artificial intelligence as a service

Most of the AI solutions discussed above are offered online as a service (i.e., the so call Machine Learning As A Service (MLaaS)), which is a common way of delivering applications over the internet. Instead of installing and maintaining software, the user only accesses it via the internet. While this model frees the user from the complexity of software and hardware management, it also brings limitations. For instance, for the sake of simplicity

³²¹ Play Aidungeon, <https://play.aidungeon.io/main/home> (accessed on 31 March 2021).

³²² Play Aidungeon, <https://play.aidungeon.io/main/currencyStore> (accessed on 31 March 2021).

³²³ Play Aidungeon, <https://play.aidungeon.io/main/worlds> (accessed on 31 March 2021).

³²⁴ Play Aidungeon, <https://play.aidungeon.io/main/subscribe> (accessed on 31 March 2021).

³²⁵ Techcrunch, <https://techcrunch.com/2021/02/04/latitude-seed-funding/> (accessed on 31 March 2021).

and user accessibility, the user is typically bound to use the trained model³²⁶ of the service provider as a black box, without being offered the possibility to check or update the training data, to prevent changes to the trained model. Additionally, the creation and download of personalised trained models do not seem to be common practices.

This service-model leading to what might be called CaaS (Creation as a Service) affects the risks incurred by the AI solution providers: they control access to the solution, which can be modified or terminated in case of non-compliance with the (contractual) conditions. The risk of using the AI solution for processing unauthorised content is reduced. The AI developers have factual control over access and hence a possibility to gain revenues, if demand exists, from their AI solution. The circumstance that the AI developer does not have any rights to control the AI output, may not be an obstacle for their business model, which is based on control over the access to and use of the AI solution and is determined by the contractual conditions.

3.2.5.3. Importance of input data

The **input data** is of particular importance because currently, the most effective AI solutions rely on algorithms that implement supervised learning. This means that prior to using the algorithm, it needs to “learn” from labelled datasets that are used to train the model that extracts patterns out of the training data.

However, this poses several challenges. First, the labelling of data in the creative sector may not always be straightforward or even possible, as the artists’ creativity often involves creating as well as abstractly combining ideas, and it may be affected by experience and emotions. Second, the task of data collection and data labelling can be resource-intensive as this needs to be done or supervised by humans. It may even be argued that the labelling itself implies human creative choices. Third, there is uncertainty on whether creative works protected by copyright may be used or not for the training of the AI algorithms, irrespectively of the nature and purpose of the AI output, without the need for prior authorisation or licences from the rightsholders. Regarding the two first above-mentioned challenges, unsupervised or self-supervised algorithms (that do not need labelled datasets) are prime candidates for underpinning the next generation of AI solutions, particularly so as the amount of unlabelled data dramatically and continuously grows on the internet. Concerning the third challenge, this is one of the issues that will be further discussed in the subsequent sections of this study.

In order to assess the copyright relevance of the data processing, it should consequently be verified (i) which training data are used and whether they contain copyright protected works, protected performances and/or protected fixations or recordings, (ii) which technical acts of copying and communicating the content are performed during the training process and (iii) in which context and for which purpose the processing is made.

3.2.5.4. The variety of artificial intelligence outputs

The type of output ranges according to the purpose of the AI solution, from categorisations of creative works into a taxonomy, personalised recommendations of creative works, post-processed creative works, generation of AI-made creative output. The output is rarely

³²⁶ Cloud services for building and deploy AI algorithms (e.g. Google’s Vertex AI) allow the upload of a trained model, but these tools are not for the general user as they require software programming skills to operate them.

deterministic, meaning that the same input data may lead to different outputs, particularly as the model is updated.

Additionally, the output benefits significantly from large amounts of diverse and unbiased data (e.g. to avoid racial, sexual bias), but this generally cannot be checked by the user, as the software is typically presented to the user as a simple interface without access to the training model.

Lastly, the amount of human creative choices impacting the output might vary from one AI solution to another. This feature will determine the eligibility for protection under copyright and under the performer's rights. Depending on the type of output, it may also be protected under the producers' rights, for which no qualitative conditions apply.

3.3. Challenges of the use of artificial intelligence within the current copyright framework

In this section, the copyright framework (in its broad sense, covering also related and *sui generis* rights) will be analysed with a focus on the identification of the possible challenges of the use of artificial intelligence within the current copyright framework.

After a short presentation of copyright and related rights (section 3.3.1), the subject matter protected by these rights (section 3.3.2), the attribution of authorship and ownership of rights (section 3.3.3), the economic and moral rights granted under copyright and related rights (sections 3.3.4 and 3.3.5), the relevant exceptions (section 3.3.6) will be discussed in order to describe how these rights apply to the use and development of AI solutions in the cultural sector. For each matter, the potential challenges will be illustrated with applications drawn from section 3.2 in which the current use of AI in the selected cultural sectors is described.

3.3.1. A first overview of copyright, related and sui-generis rights

The use of AI applications in the cultural sector raises several copyright related questions: the AI input (training data or other submitted data) may contain protected subject matter (works, performances, recordings, databases), the AI solution may be a copyright protected software and the AI output may result in creations that are very similar to traditional cultural creations and that may consequently be protected under copyright or the related rights (if the conditions are met). In order to assess, at a later stage, whether the processing of AI input requires the right holder's consent under copyright or under the related rights and whether the generation of AI output should be protected, the starting point of the reflexion should be the main principles and justifications for the protection under copyright and the related rights.

3.3.1.1. Copyright

Copyright confers **legal exclusivity** over literary and artistic creations for a limited period of time. Therefore, copyright protects a work and allows the rightholder to prevent others from copying the work without their permission.

The **justifications** for copyright protection traditionally present two dimensions: a personhood rationale and an economic incentive rationale. The *personhood rationale* justifies the existence of these rights based on the personal imprint that "a work of authorship bearing the personal imprint of its creator is in effect an extension of the

author's personality"³²⁷. This author-centred theory is often justified upon natural law and justice or fairness arguments. The *economic incentive rationale* is less concerned with the interest of the creators but aimed at maximising the welfare of society as a whole. Under this theory, copyright aims at allowing creators to exploit their works within markets, despite the initial public good nature of their creation. The awarding of time-limited exclusive rights over the fruit of their work indeed allows markets to generate economic incentives for the creators, supposed to stimulate both the creation and dissemination of cultural outputs.

Copyright confers two types of **rights** to the author: moral rights and economic rights. *Moral rights* include the integrity of the work and a right of attribution to the author. Moral rights are not transferable (but can usually be waived), in opposition to the fully transferable *economic rights*, which permit authors to obtain financial benefits from their creation.

Copyright is largely **harmonized** in the European Union by the thirteen directives and two regulations adopted so far. The main legislation is the 2001/29 InfoSoc Directive, which harmonised to a certain extent the reproduction right (Art. 2), the right of communication (Art. 3), the distribution right (Art. 4), and the exceptions and limitations (Art. 5). In addition, the Court of Justice of the EU has clarified the EU laws by offering a reasonably coherent interpretation of several blocks of copyright: the substantive conditions of protection, the scope of the rights and of the exceptions, the balancing with other fundamental rights (freedom of expression, privacy, freedom to conduct a business), the responsibilities of online intermediaries (including for hyperlinking and aggregating content), etc. However, certain matters are not harmonised under EU law, for example the moral rights. On many other issues (such as copyright levies or contractual rules), the EU laws only offer some partial guidance. Therefore, the national laws need also be considered to complete the EU copyright framework.

3.3.1.2. Related and sui generis rights

Next to copyright, a swarm of rights, qualified as **neighbouring / related and sui generis rights**, exist. Creative content and assets may be protected under both copyright and one or more related/sui generis rights, which apply cumulatively. Each related/sui generis right confers to the owners **some degree of legal exclusivity** – often coming close to the exclusivity granted by copyright³²⁸ – upon the outcomes of given efforts or investments.

From a terminological point of view, some of these rights (rights of **performers, phonogram producers, broadcasters, film producers, press publishers**), because they are traditionally connected to the creation or exploitation of a work, are qualified as **related rights** (or **neighbouring rights**). Additionally, **sui generis rights** were provided

³²⁷ P. GOLDSTEIN and P.B. HUGENHOLTS, *International Copyright: principles, law and practice*, Oxford, Oxford University Press 2019, <https://books.google.be/books?id=D7emDwAAQBAJ>, (accessed in 14 July 2021); See also: M. SENFTLEBEN and L. BUIJTELAAR, "Robot Creativity: An Incentive-Based Neighboring Rights Approach", *SSRN* 2020, 11 et seq.

³²⁸ See for instance: P. GOLDSTEIN and P.B. HUGENHOLTZ, *International Copyright: Principles, Law, and Practice*, Oxford, Oxford University Press 2019, according to whom "the civil law countries, particularly as spurred by E.C. Directives, have increased the protection that they grant to neighboring rights to a point at which these rights often approach author's rights in rigor and effect."

for creations with a more remote link to copyright protected works, i.e. the “database right”.³²⁹

In comparison to copyright, these rights are less harmonised, and the number of those rights even varies between Member States. As an example, some countries such as Germany or Spain also provide protection for **non-original photographs** (*Lichtbilder*³³⁰, *meras fotografías*³³¹) that are not protected in the other Member States.

With the exception of performers’ rights, these rights “have in common that **they reward economic or entrepreneurial expenditure rather than human creativity**”³³². This specificity explains why, in the context of AI, where the link with human creativity of some outputs might be weak, these rights could be called upon to play an important role.

3.3.2. Protected subject matter – when are artificial intelligence inputs and outputs protectable?

The use of any given dataset as AI input only has copyright relevance if the content of the dataset meets the conditions for protection under copyright or the related rights. The same conditions will determine whether the AI output can be protected.

3.3.2.1. Copyright protected works

Copyright protects “**literary and artistic works**” or “works”. The notion of a “work”, as an intellectual creation protected by copyright, is, according to the Court of Justice of the EU (CJEU), “an autonomous concept of EU law which must be interpreted and applied uniformly, requiring two cumulative conditions to be satisfied”: **a form of expression** and **an original subject matter**.³³³

3.3.2.1.1. *The expression-requirement and the absence of protection of ‘styles’*

The form of expression under EU law

Regarding the **form of expression**, in the EU, copyright law distinguishes ideas, principles, theories, concepts, systems, methods, etc., from **the forms in which these are expressed**. Copyright protects only that form, as expressed.³³⁴

The Court of Justice of the EU stated, on this point, that “the subject matter protected by copyright must be expressed in a manner which makes it identifiable with sufficient

³²⁹ European parliament and Council, Directive 96/09/EC of 11 March 1996 on the legal protection of databases, (O.J. L 77 28 March 1996, p. 20).

³³⁰ German Copyright Act, §72.

³³¹ Spain Copyright Act, art. 128.

³³² C. HARTMANN, et al., “Trends and Developments in Artificial Intelligence. Challenges to the Intellectual Property Rights Framework”, 2020, <https://op.europa.eu/en/publication-detail/-/publication/394345a1-2ecf-11eb-b27b-01aa75ed71a1/language-en>, (accessed in July 2021).

³³³ CJEU 12 september 2019, Cofemel / G-Star Raw CV, EU:C:2019:721, paragraph 29.

³³⁴ For examples in Belgium: Brussels 18 September 1998, (Plan K/F Flamand v. Béjart), *IRDI* 1998, p. 346 (applying the distinction to a choreographic sequence, protecting the sequence, and finding infringement); Brussels 28 February 2008, (Verbraeken v. VRT and BVBA De Filistijnen), *A&M* 2008, 2, p. 114 (applying it to a format for a television show, protecting the format); See also Brussels 3 October 2013, (Permis d’rire), *JLMB* 2014, 1, p. 446 (no protection for the idea of TV shows whose format is not described in a written document).

precision and objectivity.”³³⁵ It is commonly accepted that artistic styles fall within the realm of “ideas” that therefore cannot be protected under copyright³³⁶.

AI-related challenges (1): Use of protected works as input for an AI solution

Where cultural content is used as input for an AI solution, it is commonly subject to technical operations to make it suitable to train the algorithms. For example, the digital files can be arranged in a different technical format and enriched by labels or metadata to facilitate supervised machine learning.

For this preparatory phase, a digital transcription of the cultural content is used: the image, text or recording is no longer recognisable to the human senses but is, conceptually, still the same work or protected fixation – albeit in another form.

This means that any given dataset, used to train an AI solution, may contain copyright-protected works and that, as a matter of principle, the developer or user of the AI solution who processes such works should verify whether the authors’ consent is required. Considering the volume and the nature of the data needed to optimise or use an AI solution, the resulting obligations would entail additional challenges for the development of AI applications (see also below 3.3.4 and the policy scenarios sub 3.4.1).

AI-related challenges (2): AI output and the protection of “style”

As described in part 3.2, several advanced AI applications³³⁷ allow generating outputs which are similar to previous works and sometimes famous predecessors, without necessarily including any discernible protected element of the works used as inputs.³³⁸ In some sectors, this might lead to the substitution of human-created works by machine-generated productions, which may have an impact on the economic situation of authors and derived right holders. Despite those potential economic consequences, the author cannot rely on copyright to prohibit the making and commercialising of AI-generated productions that imitate their style but that do not actually take over any distinguishable parts that qualify for copyright protection.

If **the disruptive potential** of these AI applications seems obvious and was also flagged by participants during the interviews, the participants to the Delphi survey were less alarmed. Participants were asked to provide their opinion on whether copies of style made by AI systems would impact artists more than copies of style made by humans. In the opinion of only 47% of them, artists will be significantly or slightly more affected by AI-generated copies of a creative style than by human-made takings. The two main reasons given by the stakeholders were:

- AI applications may produce vast amounts of works overwhelming the possible market with similar or identical products to those created by human artists. Therefore, this threatens the revenue stream of artists and could drive several artists out of business. Among the respondents of the second round of the Delphi survey, 77% strongly agreed (13) or somewhat agreed (7).

³³⁵ CJEU 13 November 2018, *Levola Hengelo v. S. milde Foods*, EU:C:2018:899, paragraph 40.

³³⁶ The Belgian supreme court has confirmed that copyright protection does not extend to ideas, concepts, styles, fashions or genres, which are merely the expression of a general sense of aesthetics: Cass. 17 January 2017, RG N° C.15.0144.N

³³⁷ For instance the examples of ‘Prisma’ and ‘the Next Rembrandt’ in the visual art sector in section 3.2.1.2 and ‘OpenAI Jukebox’ and ‘music sector Lost Tapes of the 27 Club’ in the music sector in section 3.2.2.2.

³³⁸ Other relevant examples can be found in: J.-M. DELTORN, “Deep creations: Intellectual property and the automata”, 4 *Frontiers in Digital Humanities* 3 2017, p. 8–9.

- AI mimics can be very disruptive as the level of precision and the production rate can be immensely higher than that of humans. This was the opinion of 81% of the participants, of which 14 strongly agreed and 7 somewhat agreed.

In contrast, 8% of respondents argued that they would be less affected by AI-generated copies compared to human-made copies (see Figure 31).

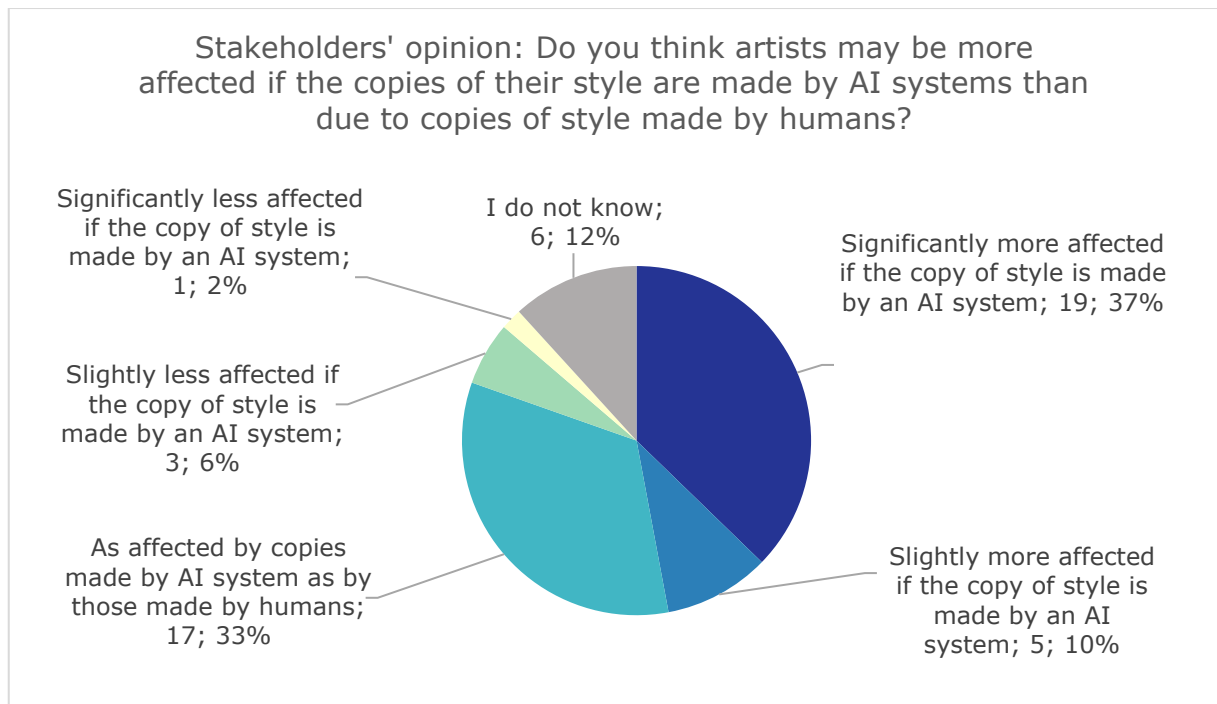


Figure 31: Stakeholders' opinion on the impact of copies of style made by AI vs. by humans

Source: Technopolis Group Survey

It should be noted that the question was formulated in very general terms, in which the definition of "copy" was diffuse and broad, and the use of the copy was also not stated (e.g. personal use, homage, plagiarism). It is possible that the participants understood it in a purely objective way, (i.e. does a copy of a style generated with the help of an AI system affect more artists as if a human makes the copy) and not in a more systemic sense (i.e. if artists will due the deployment of this technologies be more affected by this type of practices).

Due to the systemic effect that such practices may have in the future on cultural endeavours, the question may be asked whether copyright and the related rights should protect authors and performers against such practices.

Such protection would however come down to the **protection of a "style"**, which in principle falls outside the scope of copyright law. A style can indeed be defined as "any distinctive, and the therefore recognisable way in which an act is performed, or an artefact

made or ought to be performed and made”³³⁹. The delineation of a style may be based on several works of one artist or on the œuvre of several artists, active during the same period or belonging to the same school. The "style" of an author's work is hence described at a higher level of abstraction than the elements that are protected under the existing copyright rules (see also below subsection 3.3.4 on Economic rights and infringement). In principle, an AI output indeed does not constitute a servile 'copy' of any of the protected works or performances that have been used as input to train the AI model. If an AI model has been trained with several works of a given author, it is expected to produce outputs that will not replicate any of the pieces of work it has been trained with but will constitute new item. This new item draws on the analysis of the works which have initially been fed to the model.

Nevertheless, where an AI solution is trained to generate cultural output "in the style" of a particular author³⁴⁰, this possibility to "imitate" is likely to generate tensions and authors and performers may request some kind of legal protection against this type of practice. These challenges are addressed below (section 3.4.2.2).

3.3.2.1.2. *The condition of originality*

“Originality” under EU law

Although there is no explicit general requirement of “originality” for copyright protection, the CJEU has derived such condition from the different directives and has harmonised this notion for all works.

- Free and creative choices
- In the selection, arrangement and combination of elements

According to EU law, works “are protected by copyright only if they are original in the sense that they are **their author's own intellectual creation**”³⁴¹. This requirement supposes a *human* effort or contribution. This condition in the current copyright legislation results from a policy choice concerning the function of the copyright system, which serves as an instrument to encourage human creativity and not the production of cultural outputs in general. Unlike patent and design law, copyright does not require any novelty. The practical or utilitarian function of the object embodying the original creation does not prevent the protection by copyright but copyright extends only to the elements of the work which are not dictated by technical or functional considerations.

Firstly, the originality of a creation can be deduced from the author's free and creative choices in the expression of her creative ideas. For the Court of Justice of the EU, “if a subject matter is to be capable of being regarded as original, it is both necessary and sufficient that the subject matter **reflects the personality of its author, as an expression of his free and creative choices**”.³⁴²

That « **free and creative choices** » are needed and sufficient was confirmed by the CJEU for a photograph in the **Painer** case³⁴³: “As regards a portrait photograph, the photographer can make free and creative choices in several ways and at various points in

³³⁹ Gombrich, Ernst, *Style*, THE ART OF ART HISTORY: A CRITICAL ANTHOLOGY, 120–140, 2009, p. 129.

³⁴⁰ In such a case, the moral right of the author could be used as a basis for interesting claims, see section 3.3.5.

³⁴¹ CJEU 16 July 2009, Infopaq International A/S/ v. Danske Dagblades Forening, EU:C:2009:465, paragraph 35.

³⁴² CJEU 12 september 2019, Cofemel / G-Star Raw CV, EU:C:2019:721, paragraph 30.

³⁴³ CJEU 1 december 2011, Eva-Maria Painer / Standard VerlagsGmbH, EU:C:2011:798.

its production. In the preparation phase, the photographer can choose the background, the subject's pose and the lighting. When taking a portrait photograph, he can choose the framing, the angle of view and the atmosphere created. Finally, when selecting the snapshot, the photographer may choose from a variety of developing techniques the one he wishes to adopt or, where appropriate, use computer software³⁴⁴.

By making those various choices, the author of a portrait photograph can stamp the work created with their '**personal touch**'³⁴⁵. « Free and creative choices » can be made even when the objective, as for a portrait photograph, is to best render the features of a human physiognomy: even in this case, "the freedom available to the author to exercise his creative abilities will not necessarily be minor or even non-existent"³⁴⁶.

Secondly, **originality** might result from the **choice, arrangement and combination of elements** that, taken independently, are not new or original. In *Infopaq*³⁴⁷, the Court of Justice ruled on whether a short sequence of elements can be original and protected. This case involved a **sequence of eleven words** "which, considered in isolation, are not as such an intellectual creation of the author who employs them"³⁴⁸. For the Court, « It is only through the choice, sequence and combination of those words that the author may express his creativity in an original manner and achieve a result which is an intellectual creation »³⁴⁹.

This also applies to the choice and arrangement of words in a technical document, such as a user manual accompanying software on which the court ruled in *SAS Institute*.³⁵⁰ Three operations governed by the author can reveal the creative touch: to choose, to arrange and to combine the elements, which, considered independently (or "as such"), can be non-original or original.

Thirdly, the **practical or utilitarian function** should not prevent the copyright protection of the creation it embodies. The category of "**works of applied art**" is expressly mentioned in the list of protectable works in Article 2(1) of the Berne Convention: it covers the works embodied in the items or articles having a practical application or a utilitarian function. Besides, in the EU design directive and the design regulation, the European legislator has imposed **the rule of full cumulation with copyright**.³⁵¹

Fourthly, **copyright will not subsist** where the expression of a work is **dictated only by a technical function** as, in such cases, the idea and expression of the work are indissociable.

The CJEU confirmed this in a case involving **the graphic user interface (GUI) of computer games**³⁵² : the national judge "must take account, inter alia, of the specific arrangement or configuration of all the components which form part of the graphic user interface" and "where the expression of those components is dictated by their technical

³⁴⁴ CJEU 1 december 2011, Eva-Maria Painer / Standard VerlagsGmbH, EU:C:2011:798, paragraph 91.

³⁴⁵ CJEU 1 december 2011, Eva-Maria Painer / Standard VerlagsGmbH, EU:C:2011:798, paragraphs 90-92.

³⁴⁶ CJEU 1 december 2011, Eva-Maria Painer / Standard VerlagsGmbH, EU:C:2011:798, paragraph 93.

³⁴⁷ CJEU 16 July 2009, Infopaq International A/S/ v. Danske Dagblades Forening, EU:C:2009:465.

³⁴⁸ CJEU 16 July 2009, Infopaq International A/S/ v. Danske Dagblades Forening, EU:C:2009:465, paragraph 45.

³⁴⁹ CJEU 16 July 2009, Infopaq International A/S/ v. Danske Dagblades Forening, EU:C:2009:465, paragraph 45.

³⁵⁰ CJEU 29 November 2011, SAS Institute Inc. v. World Programming Ltd., EU:C:2011:787, paragraphs 66-70.

³⁵¹ European Parliament and the Council, Directive 98/71 of 13 October 1998 on the legal protection of designs (O.J. L 3, 05.01.2002, p. 1) and Regulation 6/2002 of 12 December 2001 on community designs (O.J. L 3, 05.01.2002, p. 1).

³⁵² CJEU 22 December 2010, Bezpečnostní softwarová asociace Svaz softwarové ochrany / Ministerstvo kultury, EU:C:2010:816, paragraphs 48-50.

function, the criterion of originality is not met, since the different methods of implementing an idea are so limited that the idea and the expression become indissociable. In such a situation, the components of a graphic user interface do not permit the author to express his creativity in an original manner and achieve a result which is an intellectual creation of that author”³⁵³.

Regarding the **originality of a database**, the CJEU confirmed in *Football Dataco* that the application of labour and skill, without the exercise of creative freedom, in the making of a utilitarian article is insufficient to demonstrate the originality required for copyright protection: the originality “criterion is not satisfied when the setting up of the database is dictated by technical considerations, rules or constraints which leave *no room for creative freedom*”.³⁵⁴ The **exercise of the freedom of creation** is thus the leading condition for copyright protection. Often this freedom in the creative process can only be demonstrated negatively, i.e. through the absence of any compelling factor or constraint restricting the choice of the author to zero.

This was again confirmed by the 2019 *Cofemel* decision, where the Court underlined that there is no originality when the expression of the creation “has been dictated by technical considerations, rules or other constraints, which have left *no room* for creative freedom”.³⁵⁵ The analysis below shows that no « technical considerations, rules or constraints » can explain the various choices made by the designer of the appearance of the electric truck.

In *Brompton*, the CJEU held that “a subject matter satisfying the condition of originality may be eligible for copyright protection, even if its realisation has been dictated by technical considerations, provided that it is being so dictated has not prevented the author from reflecting his personality in that subject matter, as an expression of free and creative choices”.³⁵⁶

AI-related challenges: copyright protection of AI-assisted and AI-generated output

The condition of originality plays a decisive role in the protectability of cultural outputs generated with the support of an AI solution. A distinction needs to be made between (i) creations that are made by a human author who uses AI as a tool (AI-assisted creation) and (ii) creations that are autonomously generated by an AI-solution.

A human who uses **an AI-solution as a tool** in the creative process, i.e. **AI-assisted creations**, may make a sufficient intellectual effort to qualify for copyright protection under the current rules. The creation of the album “Hello world!” previously described is a good example: the artists operate free and creative choices both on the input (by selecting the works to be part of the dataset used to train the model) and on the output side (by choosing and modifying the music proposed by the AI, adding (human) vocals, instruments, arrangements). The AI solution appears hence to come only as a tool supporting the creative process. In such a case, the free and creative choices operated by the creators left no doubt on these cultural outputs’ copyright protectability.

³⁵³ CJEU 22 December 2010, *Bezpečnostní softwarová asociace Svaz softwarové ochrany / Ministerstvo kultury*, EU:C:2010:816, paragraphs 48-50.

³⁵⁴ CJEU 1 March 2012, *Football Dataco Ltd v. Yahoo UK Ltd.*, EU:C:2012:115, paragraph 39.

³⁵⁵ CJEU 12 September 2019, *Cofemel / G-Star Raw CV*, EU:C:2019:721, paragraph 31.

³⁵⁶ CJEU 6 February 2020, *Brompton Bicycle Ltd. / Chedech/Get2Get*, EU:C:2020:79, paragraph 26.

Different shades of human intervention by different kinds of actors require however a case-by-case assessment whether the effort results in an "original" creation³⁵⁷. The selection of the dataset to train the algorithms might be done by an engineer, in cooperation with the artist who instead describes the result they want to achieve. The artist may make a skilled and artistically motivated choice among various AI-generated outputs, but such **curatorial effort** is traditionally not protected as a work. Should a human choice between several outputs generated autonomously by an AI composer be considered sufficient to meet the threshold? The effort of the human is evident, but it is less certain whether the resulting output meets the originality threshold.

Autonomously generated output, produced **without human creative choices** at the input or output side of the AI process, necessarily falls **outside the scope of copyright** protection³⁵⁸. This would for instance be the case for pieces of music autonomously generated with the AIVA AI composer.³⁵⁹ Photos or music that are autonomously created by an AI solution are not protected under copyright, meaning that the person who triggered the creation cannot prohibit the reproduction or communication to the public by any user. The absence of copyright protection does not mean that this AI-output cannot be monetised, nor controlled by other legal means. The producer of the output may seek other types of protection, such as contracts or perhaps related rights.

As a matter of principle, **human intervention is required** to qualify for copyright protection. It will be a matter for the courts to apply this general rule and determine, on the sliding scale of human/robot collaboration, whether human efforts are required to find an "original expression"³⁶⁰.

As it can be observed in Figure 32, the majority of the experts (43; 75%) participating in the Delphi survey were aware that creative outputs generated by AI without human intervention cannot be protected by copyright. However, three of them noted that although this situation is theoretically possible, they doubt creative content can be created without human creative choices. They argued that within the current state of the art, the human is always involved in the creative process in different roles, i.e. as the creator of the data that is fed to the AI, as the developer of the AI software or as the user of the AI.

³⁵⁷ As set out in the JIIP-IVIR study: C. HARTMANN et al. "Trends and Developments in Artificial Intelligence. Challenges to the Intellectual Property Rights Framework", <https://op.europa.eu/en/publication-detail/-/publication/394345a1-2ecf-11eb-b27b-01aa75ed71a1/language-en>, (accessed in July 2021); As the authors of this study correctly point out, whether or not a creation, made with the help of an AI solution, can be protected under copyright should be analysed in the light of all the circumstances of each case. The exact role of the human and the machine in the conception, execution and redaction phases should be analysed to determine whether the creation is an AI-assisted output (eligible for copyright protection) or an AI-generated output (not eligible for copyright protection).

³⁵⁸ In this sense: J. DREXL et al., Artificial Intelligence and Intellectual Property Law-Position Statement of the Max Planck Institute for Innovation and Competition of 9 April 2021 on the Current Debate, *Max Planck Institute for Innovation Competition Research Paper 2021 No. 21-10*, p. 21; M. SENFTLEBEN L. BUIJTELAAR, *Robot Creativity: An Incentive-Based Neighboring Rights Approach*, 2020 p. 11; See also the examples of weather and news reports which may be generated in a fully automated way by AI systems and consequently not protected under copyright: C. HARTMANN et al. See also the examples of weather and news reports which may be generated in a fully automated way by AI systems and consequently not protected under copyright: C. HARTMANN et al., "Trends and Developments in Artificial Intelligence. Challenges to the Intellectual Property Rights Framework", <https://op.europa.eu/en/publication-detail/-/publication/394345a1-2ecf-11eb-b27b-01aa75ed71a1/language-en>, (accessed in July 2021).

³⁵⁹ See section 3.2.2.3

³⁶⁰ For different applications: Senftleben and Buijtelaa, *supra* note 89.

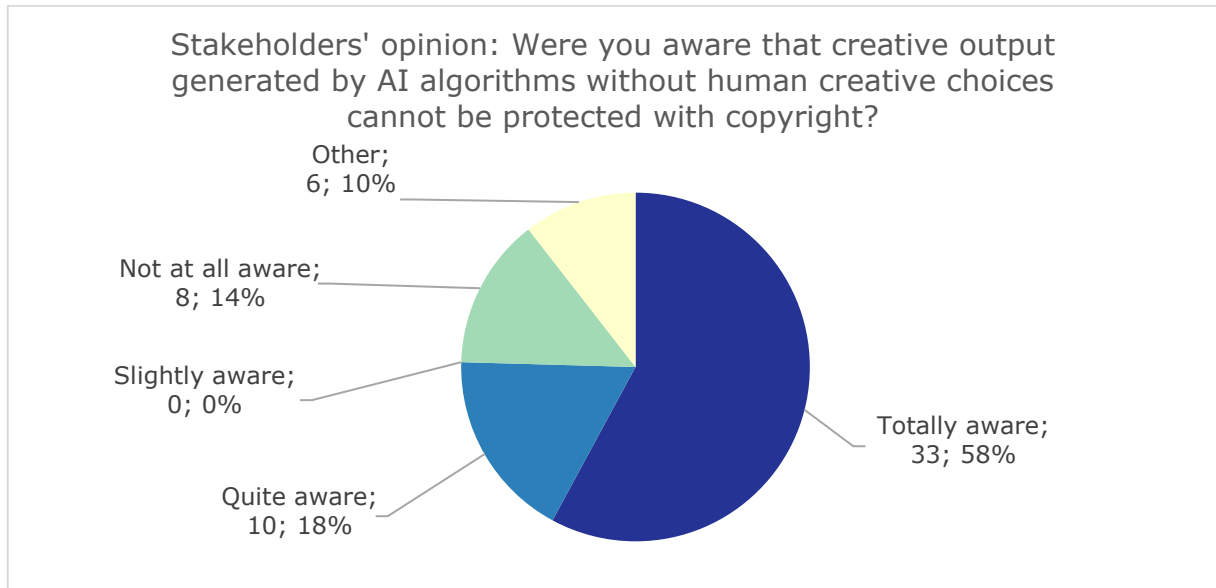


Figure 32: Stakeholders' awareness on the absence of protection for AI-generated output

Source: Technopolis Group Survey

Even if – at least in theory – the copyright conditions are clear, the application of these principles might be rendered more difficult where it is **practically impossible to distinguish between human and non-human output** (e.g. creations of GAN solutions, images of people produced by Generated Photos or music generated by AIVA). Where it is not obvious to the public that creation is actually produced by an AI solution, without protected human contribution, the public might be led to believe (and the producer of the output might be tempted to claim) that the creation is actually human-created, and therefore, protected. This would mean that a user could see the need to clear the rights to use the creation – as if it were a copyright-protected work – even though no prior authorisation of any author is legally required. The use of such creations would thus be restricted, solely due to the user's belief that the creation is copyright protected.

This raises questions regarding the **burden of proof** of copyright protection. It will remain to be seen how the creators can establish that they have made an original – human – contribution. Inversely, it may be difficult for a member of the public who has used an AI output without asking for prior authorisation, to refute the claim of copyright infringement on the account of the creation not being created by a human and consequently not being protected under copyright: the alleged infringer does not know about the creative process and can only guess whether the alleged author has actually made a sufficient creative effort – without however being able to prove that the output was actually generated by a machine. Some AI output cannot be distinguished from a human-created work in the same style (e.g., electronic music or digital art) so there are no indications in the output itself that it is machine-made. An unsubstantiated claim of copyright protection by the producer of the AI output may thus hamper the free use of a creation which is not protected.

The effect on the creative market is hard to predict at this stage (also considering that the business models of those AI solutions are not so much copyright-backed as service-based). Depending on the direction of future developments, different policy options may be considered (below 3.4.2.6).

3.3.2.2. Performances

AI solutions are increasingly used in the musical or audiovisual sectors to generate musical tunes or videos that can be played without the intervention of any human musicians, actors or singers. Such solutions rely on the processing of human performances (AI input), which they aim to mimic, and result in performance-like output.

Where such human performances are protected under related rights, their use as AI input might raise legal questions. On the output side, a peculiar issue arises when a human musician, actor or singer performs an AI-generated creation, which is not protected under copyright. Depending on the articulation between copyright and the performer's rights under the applicable national law, such human performance of AI-generated creations may not actually be protected under the performer's rights scheme.

3.3.2.2.1. The notion of 'performance'

Definition of "performance" and "performer"

In addition to literary and artistic works, the international and European legal framework organises the protection of "performances". The international and European legal instruments offer **no direct definition of the subject matter** of the protection – the performance – but provide for a description of "performers" instead.³⁶¹

At the **international level**, Art. 3(a) Convention of Rome defines "performers" as:

*actors, singers, musicians, dancers, and other persons who act, sing, deliver, declaim, play in, or otherwise **perform literary or artistic works***³⁶²

An equivalent definition can be found in Art. 2(a) WIPO Performances and Phonograms Treaty (WPPT),³⁶³ accordingly to which performers are

*actors, singers, musicians, dancers, and other persons who act, sing, deliver, declaim, play in, interpret, or otherwise **perform literary or artistic works or expressions of folklore***³⁶⁴

The WPPT hence expands the scope of the definition at two levels: it first adds "interpretations" to the list of acts that can be regarded as performance and it adds that expressions of folklore (which may not be protected under copyright) may also give rise to a protected performance. The WPPT definition is also included in Art. 2(a) the Beijing Treaty on Audiovisual Performances.³⁶⁵

At the **European level**, the Rental Directive, the Term Directive and the InfoSoc Directive harmonise certain aspects of the protection of "performances" in favour of "performers" without however defining either notion. Confronted to this lack of definition in secondary legislation, the Court of Justice interpreted this notion in its RAAP decision³⁶⁶, as necessitating "*an autonomous and uniform interpretation*" (§46) and relied then expressly

³⁶¹ World Intellectual property Organization (WIPO) Guide to the Rome Convention and to the Phonograms Convention 22 (1981)), <https://www.wipo.int/publications/en/details.jsp?id=3161&plang=EN>

³⁶² Emphasis added.

³⁶³ As a matter of completeness, the WPPT definition is also included in Art. 2(a) the Beijing Treaty on Audiovisual Performances.

³⁶⁴ Emphasis added.

³⁶⁵ It is added in agreed statement concerning Article 2(a): "It is understood that the definition of "performers" includes those who perform a literary or artistic work that is created or first fixed in the course of a performance".

³⁶⁶ CJEU 2 July 2020, Recorded Artists Actors Performers Ltd. / Phonographic Performance (Ireland) Ltd, Minister for Jobs, Enterprise and Innovation, C-265/19, ECLI:EU:C:2020:512.

upon the WPPT definition, “to which the European Union and all its Member States are contracting parties” (§52). According to the Court, under this definition, “the concept of ‘performers’ refers to all persons ‘who act, sing, deliver, declaim, play in, interpret, or otherwise perform literary or artistic works or expressions of folklore’”.

A similar definition can be found **at the national level**:

- §73 **German** Copyright Act states that “[a] performing artist within the meaning of this Act is anyone who performs, sings, plays or otherwise presents a *work or form of expression of folk art* or who participates artistically in such a performance.”
- The **French** Code of Intellectual Property provides in its Art. L.212-1 that “the performer is the person who represents, sings, recites, declaims, plays or performs in any other way a *literary or artistic work, a variety, circus or puppet act*.”
- In **Poland**, Art. 85 of the Act on Copyright and Related rights³⁶⁷ provides that “1. Any performance of a *work or product of the folk art* shall be protected irrespective of its value, purpose or way of expression. 2. Performances, within the meaning of paragraph 1, shall include, in particular, the actions of: actors, reciters, conductors, instrumentalists, singers, dancers, mimes and other persons making a creative contribution to the creation of a performance.”

The convoluted relationship between the protection of the “performance” and the copyright protection of what is being performed

The relationship between copyright and performers’ right is quite complex. It is repeatedly stated that the protection of neighbouring rights does not affect the protection under copyright³⁶⁸. Furthermore, it is not necessary for a performance to meet the copyright originality threshold in order to be protected under the related rights.

However, it is traditionally required for the “performance” to be *related* to a copyright-protected creation.³⁶⁹ This requirement is particularly visible in the Rome Convention, where the right owner needs to perform “literary or artistic works”. The fact that the piece that is being performed fell in the public domain and is no longer protected under copyright is without influence on the performance protectability³⁷⁰: the performance of a musical composition by Mozart will enjoy protection under the performer’s right.

The assumed purpose of this definition is in fact to exclude certain categories of performance from the protection. The *WIPO Guide to the Rome Convention* clarifies:

The reference to “works” means that the Convention does not protect a number of people who, *although* undoubtedly performers in the accepted sense, do not perform works as this is meant in copyright. Examples are variety and circus artists (jugglers, acrobats and clowns). In case there should be any doubt, it also excludes sports personalities.³⁷¹

[...]

The great difficulty about extending protection to those who do not perform works is to decide how far to go without making too many difficulties for the broadcasting

³⁶⁷ Polish Act of 4 February 1994 On Copyright and Related rights.

³⁶⁸ Art. 1 Rome Convention, art. 1(2) WPPT.

³⁶⁹ F. BRISON, “art. XI.204 X”. *Hommage à Jan Corbet, Le droit d’auteur Belge*, Brussels, Larcier, 2018.

³⁷⁰ F BRISON, “Rome Convention, art. 3”, *Concise European Copyright Law*, Brussels, Kluwer, 2016, 146.

³⁷¹ WIPO, *supra* note 122 at 21, §3.2.

organizations, i.e., to keep protection within defined limits. Some countries wish to include footballers; others only those who play in official competitions; most refuse altogether.³⁷²

In countries with equivalent definitions, courts have excluded protection for performers on this basis. In France, for instance, the Court of cassation denied this qualification to a school teacher who was the subject of, and main actor in, a documentary film,³⁷³ and to TV-reality show participants³⁷⁴.

However, the **international framework does not oblige countries to limit the protection to performers of literary or artistic works**. The “expressions of folklore” were first added to the definition since the WPPT and “variety”, “circus artists” or “performers of folkloristic acts” are explicitly mentioned as “performers” under many national copyright acts. More importantly, Art. 9 of the Rome Convention states that “any Contracting State may, by its domestic laws and regulations, extend the protection provided for in this Convention to artists who do not perform literary or artistic works.”³⁷⁵

The different descriptions of the protected “performances” suggest that the notion is understood differently in various jurisdictions. In some countries, the national laws require that the performance relates to a literary or artistic work, which is, or at least was, protected under copyright. In other countries, the rendering of a creation may be protected as a performance without explicitly or implicitly requiring the creation to be protected under copyright (but it might have to be in the “artistic sphere”).

3.3.2.2.2. *AI-related challenge: uncertain protection for performances of AI-autonomously generated creation.*

A narrow understanding of the notion of “performance”, in the sense that an artist only enjoys protection for their performance only if they perform a protected “work”, has consequences in an AI context. This would indeed mean that artists playing, reciting, acting or dancing a creation autonomously generated by an AI system (hence not protected under copyright) are left without protection - even though their performance takes place within the traditional creative sphere, requires the same creative efforts and the creative outcome might be similar to the performance of a protected work. An artist playing a piece of music generated autonomously by an AI composer such as AIVA³⁷⁶ or the performers of the “lost tapes of the 27 club” would then not enjoy any protection³⁷⁷.

This potential absence of protection was flagged at least in one national report,³⁷⁸ but this finding has not caused much discussion in either academia or in the industry. This silence might be explained by the different – more open – reading of the legal framework by other experts or stakeholders.

³⁷² *Ibidem*, 42, §. 9.3.

³⁷³ French Court of Cassation 13 November 2008, 06-16.278, *Etre et avoir*, explaining that “the teacher appeared exclusively in the reality of his activity without interpreting, in the service of the work, a role”.

³⁷⁴ French Court of Cassation 24 April 2013, 11-19.091, *L’île de la tentation*, explaining that: “the participants in the program in question had no role to play and no text to say, that they were only asked to be themselves and to express their reactions to the situations with which they were confronted and that the artificial nature of these situations and their sequence was not sufficient to give them the quality of actors”.

³⁷⁵ F BRISON, “Rome Convention, art. 3”, *Concise European Copyright Law*, Brussels, Kluwer, 2016.

³⁷⁶ Case study presented in Section 3.2.2.3.

³⁷⁷ The example is presented in Section 3.2.2.2.; Lost Tapes of the 27 club, <https://losttapesofthe27club.com/>, (accessed in July 2021).

³⁷⁸ A. BENSAMOUN et J. FARCHY, *Mission Intelligence artificielle et Culture: Rapport Final*, Paris, CSPLA, 2020.

Several participants of the Delphi survey stated that, in their understanding, the current legal framework would offer performers' rights protection to performances, even if the piece is entirely generated by AI and not protected under copyright. They base their argumentation on the fact that the rendition of some cultural expressions, which are not protected under copyright (e.g. musical compositions in the public domain or folklore), are protected as "performances".

This reasoning by analogy may however not be supported by the wording of the national provisions and of the WPPT to which the CJEU referred in its RAAP decision³⁷⁹, which may explicitly link the protection of the performance to the copyright protection of the performed work. Moreover, a bold reading ignoring such explicit link may have unwanted consequences in that it may extend the protection under the performers' rights to renditions outside the artistic or creative sphere that are otherwise not protected (e.g. news anchors or celebrities in reality shows). The legal scholars consulted in the frame of the interviews confirm that, depending on the wording in national law and on the interpretation by the CJEU, the existing definitions may become an issue for performers of AI-generated outputs in the future. Depending on the AI evolutions in the cultural sector, these challenges may need to be addressed (sub 3.4.2.3).

3.3.2.3. Phonograms, films, broadcasts, press publications, non-original photographs: other subject matters protected by related and sui-generis rights

In addition to the human-made works and performances, the production or use of cultural content often involves fixations or recordings of sorts, which may also be protected under related rights. Developers and users of AI solutions may have to take these rights into consideration when they use cultural content in their training sets or as input for the AI solution. Also, the producers of AI output may wonder whether their output is protected under these related rights. Unlike the authors and performers who are human persons, the producers of films, recordings, broadcasts or press publications may be legal entities.

3.3.2.3.1. Protected objects

Phonogram

The phonogram producer holds several economic rights to the "phonogram".

The European directives do not define "phonogram" but refer to the international framework.³⁸⁰ An initial definition can be found in the Rome Convention, according to which a 'phonogram' is "any exclusively aural fixation of sounds of a performance or other sounds".³⁸¹ This definition is further specified in the WPPT according to which "phonograms are defined as the "fixation of the sounds of a performance or other sounds, or of a representation of sounds, other than in the form of a fixation incorporated in a cinematographic or other audiovisual work" and 'fixation' as "the embodiment of sounds, or of the representations thereof, from which they can be perceived, reproduced or communicated through a device;"

³⁷⁹ See above: 3.3.2.2.1

³⁸⁰ CJEU 29 July 2019, Martin Haas / Ralf Hütter, Florian Schneider-Esleben, EU:C:2019:624, paragraph 3 to 4.

³⁸¹ Rome Convention No 328/E of 26 October 1961 for the protection of performers, producers of phonograms and broadcasting organisations. See also art. 1(a)-(b) of the Convention of 29 October 1971 for the Protection of Producers of Phonograms Against Unauthorized Duplication of Their Phonograms.

In essence, the “fixation” of a “sound” is the central criterion for finding a “phonogram”. **No additional conditions** are expressed: it is not required that the fixation be original or relate to “work” or a “performance”. Fixations of non-creative renditions of non-original sounds might still be “phonograms”.³⁸² Digital fixations are covered as well.³⁸³

In the same vein, no explicit condition of investment is stipulated. This might first be surprising since the ratio legis of the phonogram producer’s right is, precisely, to protect the producers’ **investment**. The latter is even assumed to be “considerable”, explaining why “adequate legal protection of intellectual property” is necessary, so the producer can pursue “satisfactory returns” (rec. 5 Rental Dir., rec. 10 InfoSoc Dir).³⁸⁴ This apparent gap between the ratio legis of the right and its protection condition can be explained by the technical developments that have taken place over time. When the need for phonogram protection emerged, the producer had no other choice than to make substantial investments, specifically in the recording of sound (studio and material, technician, etc.). The simple act of making a recording implied such investments, so this requirement did not need to be explicitly stated. With the emergence of digital technologies, the necessary link between recording and investments in the recording has disappeared. Sounds can be recorded, at acceptable quality, for a reasonable price with the help of a computer and a microphone.

At the time of the conception of the phonogram producer’s right, there was a clear distinction between the creation of a musical work, the performance of the musical work and the recording of the performance of the musical work. In a digital context, electronic musical compositions can be created, played and recorded with the use of a computer in one single process. In those cases, the acts of creation, performance and fixation are merge into one operation, to which the cumulative rights of copyright, performers’ and producers’ rights could apply.

The first fixation of a film

The producer of the “first fixation of a film” is granted several economic rights in respect of the original and copies of their “film”.

A definition of **film** is provided in the rental directive: films are “cinematographic or audiovisual work or moving images, whether or not accompanied by sound”.³⁸⁵

Similarly to the protection of phonograms, the protection of films applies to the ‘fixation’ of any type of “moving images”. No additional requirements exist with regard to the fixation of the film, which may be a recording of a copyright-protected work, a protected performance or non-protected moving images with – or without – sound.

Similarly to the phonogram producer’s rights, it is not required that the film producer demonstrate any type of “investment” in the fixation, but this ratio is underlying the protection under the related rights (rec. 5 Rental Dir., rec. 10 InfoSoc Dir).³⁸⁶

³⁸² See also JIIP-IVIR Study, pp. 89-90.

³⁸³ F. BRISON, “WPPT, art. 2”, *Concise European Copyright Law*, Brussels, Kluwer, 2016.

³⁸⁴ See on this point: B. HUGENHOLTZ, “Neighbouring rights are obsolete”, *IIC - International Review of Intellectual Property and Competition Law*, 006-1011, 2019.

³⁸⁵ European Parliament and the Council, Directive 2006/115/EC of 12 December 2006 on rental right and lending right and on certain rights related to copyright in the field of intellectual property, (O.J. L 27.12.2006, p. 28).

³⁸⁶ See on this point: HUGENHOLTZ, *op.cit.*

Broadcasts

The **subject matter** ("broadcast") and the rightholder ("broadcasting organisation") are not defined in the European instruments. No threshold for protection is provided: it is **not required** that the broadcast **content be original or creative, no substantial investment** in the broadcast needs to be demonstrated.

The **act of broadcasting** is defined in the Rome Convention as the "transmission by wireless means for public reception of sounds or images and sounds".³⁸⁷ This notion does however not coincide with the protected subject matter of the related right of the broadcaster, since the act of broadcasting is limited to wireless transmissions, whereas the "broadcast", as the protected subject matter, covers both transmissions by wire or over the air, including by cable or satellite.³⁸⁸

Press publications

Art. 15 DSM dir. provides for the protection of press publications concerning online uses. Accordingly, publishers of press publications established in a Member State should enjoy a right of reproduction and communication to the public "for the **online use of their press publications** by information society service providers".

Press publications are further defined in Art. 2(4) of the same Directive as "a collection composed mainly of literary works of a journalistic nature, but which can also include other works or other subject matter, and which: (a) constitutes an individual item within a periodical or regularly updated publication under a single title, such as a newspaper or a general or special interest magazine; (b) has the purpose of providing the general public with information related to news or other topics; and (c) is published in any media under the initiative, editorial responsibility and control of a service provider".

As noted by the IVIR/JIPP report³⁸⁹ and by other commentators,³⁹⁰ this definition does not require an element of originality or human creativity, since the collection can be composed both of "literary works of a journalistic nature", of "other work" but also of "other subject matter". Concerning this "other subject matter", the Recital 56 of the DSM dir. explains that it included "in particular photographs and videos". These photographs or video hence would fall under the scope of protection even if they do not reach the conditions to be considered as a work - for instance in the case of snapshot taken instantly without the possibility of any creative choices. In the same way, a report produced by "robot-journalism" could be protected under this new right, as part of a press publication, even if they are not protected under copyright.

³⁸⁷ Rome Convention No 328/E of 26 October 1961 for the protection of performers, producers of phonograms and broadcasting organisations, Art. 3(f).

³⁸⁸ European Parliament and the Council Art 7 (2) of Directive 2006/115/EC of 12 December 2006 on rental right and lending right and on certain rights related to copyright in the field of intellectual property, Art 7 (2), (O.J. L 27.12.2006, p. 28); and European Parliament and the Council Directive 2006/116/EC of 12 December 2006 on the term of protection of copyright and certain related rights, art 3 (4), (O.J. L 372, 27.12.2006, p. 12).

³⁸⁹ C. HARTMANN, J. ALLAN, P.B. HUGENHOLTZ, J.P. QUINTAIS et D. GERVAIS, "JIIP & IVIR report on AI and IP", *op. cit.*, p. 92.

³⁹⁰ See for instance V. MOSCON, "Position Statement of the Max Planck Institute for Innovation and Competition on the Proposed Modernisation of European Copyright Rules PART E Protection of Press Publications Concerning Digital Uses, *Max Planck Institute for Innovation & Competition Research Paper*, No. 17-02, paragraph 8, already noting that the proposal "might bring about the paradoxical and unacceptable consequence of always guaranteeing, regardless of the protectability of the work as such, the protection of the neighbouring right in the work published online by the publishers".

Non-original photographs (not harmonised)

Original photographs are protected under copyright, i.e., if they are the author's own intellectual creation, without any other requirements. The Member States are however free to protect "other photographs".³⁹¹ Some EU countries grant protection to non-original photographs.³⁹²

This is notably the case in **Germany**, where *Lichtbilder* are protected even if they do not meet the standard of originality (§72 German Copyright Act). The protection offered to these photographs amounts to the one conferred by copyright, with the difference that the protection duration is limited to 50 years after the publication.

A similar form of protection can also be found in **Italy**, where Art. 87§1 Italian Copyright Act provides that "The images of persons, or aspects, elements or events of natural or social life, obtained by photographic or analogous processes, including reproductions of works of figurative art and stills of cinematographic film, shall be considered photographs for the purpose of this Chapter." The producers are then granted a right of "reproduction, dissemination and marketing" (Art. 88) lasting for 20 years from the making of the photograph (Art. 92).

Spain also provides protection for *meras fotografías* (Art. 128 Spain Copyright Act), which is equivalent to the one offered to authors of photographic works but lasts for only 25 years.

3.3.2.3.2. AI related challenge: the potential protection of outputs autonomously generated by AI

On the output side, the cultural content that results from the use of an AI solution may enjoy protection as a **phonogram, film or press publication** (independently from the protection under copyright and performers' rights). The existence and the scope of the protection might however be uncertain, which may have an impact on the producer of such AI-generated cultural content.

One technical process triggers the production of AI-generated images, music or text, thus generating the creation, the performance (where applicable) and the fixation at the same time. Especially for the phonogram and film fixations, the question arises whether such process entails a "fixation" or a "recording" of a sound that qualifies for protection under the producers' rights.

Furthermore, since no requirement of effort or investments concerning the recording or the fixation is provided, the mere saving on a hard disk of a cultural piece generated by the AI solution (for instance a piece of music generated by AIVA), might suffice to trigger phonogram protection.

On this subject, participants in the Delphi survey were asked whether, in their opinion, related rights could be used as a way to protect creative output that is not protectable

³⁹¹ European Parliament and the Council, of Directive 2006/116/EC of 12 December 2006 on the term of protection of copyright and certain related rights, Art. 6, (O.J. L 372, 27.12.2006, p. 12); See however art. 14 CDSM, which provides that "when the term of protection of a work of visual art has expired, any material resulting from an act of reproduction of that work is not subject to copyright or related rights, unless the material resulting from that act of reproduction is original in the sense that it is the author's own intellectual creation".

³⁹² For an overview of the protection of these non-original photographs within the EU, it can be referred to: T. MARGONI, "Digitising the public domain: non original photographs in comparative EU copyright law", *Copyright, Property and the Social Contract*, Springer, 2018.

through copyright. Around half of them responded that they do not know, and the opinion of the other half was divided between those considering this a very or rather common practice and those believing it is uncommon (see Figure 33). One of the respondents added that protection through related rights will be applicable in some but not all cases where AI is used for generating creative output.

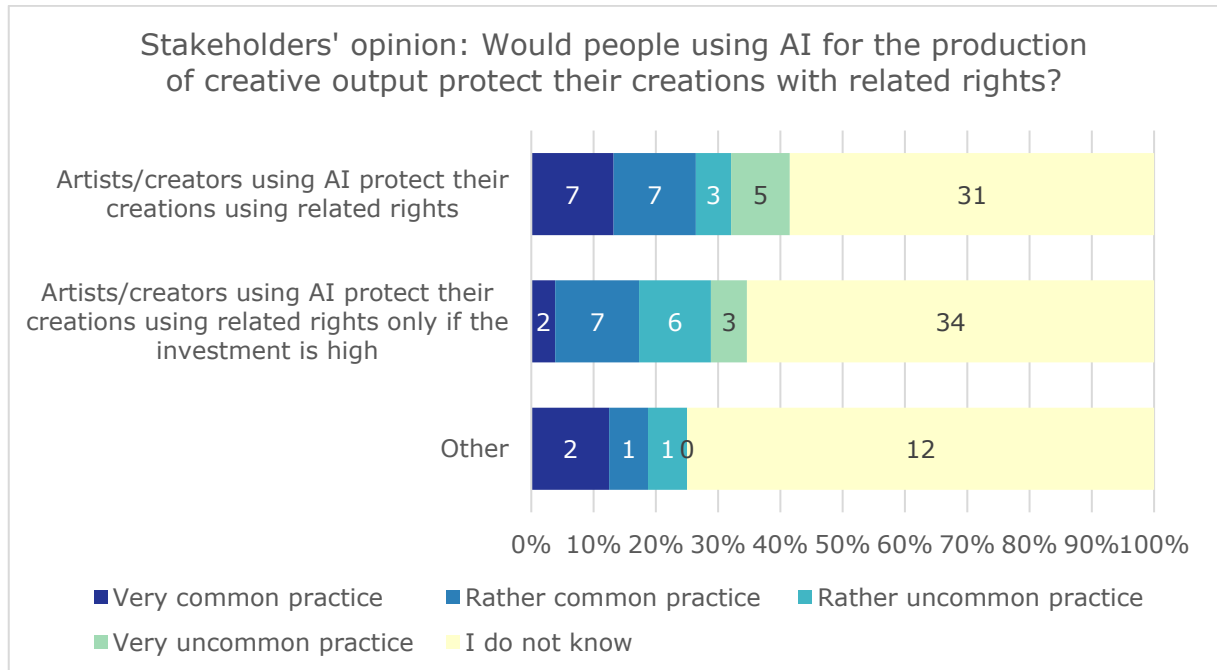


Figure 33: Stakeholders' opinion on whether people using AI would protect their creations with related rights

Source: Technopolis Group Survey

From a public interest perspective, the shift from protection under copyright to the sole protection under the producers' rights for cultural content might also present certain challenges. AI-generated cultural content would then not be protected under copyright (see Section 3.3.2.1) but would still be subject to the exclusive control of the holder of the producers' rights. This possibility, described in the IVIR/JIPP study as the "substitution" of copyright by related rights (rights of phonogram producers, right of broadcasters, rights of press publishers)³⁹³ is questionable. The protection of AI-produced cultural content under related rights (even without copyright) challenges the policymaker's decision to shape the copyright system as an instrument to encourage human creativity and not merely the production of cultural goods.³⁹⁴ Secondly, the producer of the AI-output might enjoy such a right even if they did not make any actual investment in the fixation or the recording that could economically justify the protection of the outcome (the traditional justification of the producers' rights that, however, does not translate in the conditions for enjoying this right, as underlined above).

³⁹³ C. HARTMANN et al., "Trends and Developments in Artificial Intelligence. Challenges to the Intellectual Property Rights Framework", *op.cit.*, pp. 89-93; Arguably, the related rights do not substitute copyright: producers' rights may exist both in human and machine creations, but in case of machine creations, the copyright protection simply does not come into being.

³⁹⁴ See: 3.3.2.1.2

Where **non-original photographs** are protected under national law (not harmonised at EU level), visual and photograph-like AI output might be protected as such, even autonomously generated does not raise the same issue. In this case, the national lawmakers explicitly decide to protect cultural outputs that are *not* the result of creative decisions. Although this protection might be questionable in itself³⁹⁵, and the lack of harmonisation could lead to a parcelling of the internal market, the ratio legis would not be violated if such right was used to protect AI autonomously generated outputs. The margin of the national legislators to protect non-original photographs is more limited since the adoption of the DSM dir. which provides in Art. 14 that “when the term of protection of a work of visual art has expired, any material resulting from an act of reproduction of that work is not subject to copyright or related rights, unless the material resulting from that act of reproduction is original in the sense that it is the author's own intellectual creation”. Therefore, this provision should at least avoid the protection of non-original reproductions of public domain works, which could consequently be used for AI training without restrictions

Depending on the AI evolutions in the cultural sector, the protection of AI-output under related rights may need to be addressed (sub 3.4.2.4).

3.3.2.4. Databases (sui generis right)

The database rights are especially relevant for the use of sizeable datasets for the purpose of training an AI solution. In addition to the copyright and related rights protection of each of the elements in the dataset, the AI developer may have to take into account that the collection of data, as such, may also be protected under copyright or under the sui generis database rights.

3.3.2.4.1. Protected databases

Databases are protected under copyright, and under the so-called sui generis right. Both regimes of protection may be accumulated.

A **database** is defined as “a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means”.³⁹⁶ The Database Directive clarifies that the notion “database” includes “literary, artistic, musical or other collections of works or collections of other material such as texts, sound, images, numbers, facts, and data; [it] should cover collections of independent works, data or other materials which are systematically or methodically arranged and can be individually accessed; [this] means that a recording or an audiovisual, cinematographic, literary or musical work as such does not fall within the scope of this Directive”.³⁹⁷

The CJEU has interpreted this definition in several decisions. It is required that the database be a collection of “independent” materials, i.e. they are “separable from one another

³⁹⁵ See CYRILL P. RIGAMONTI “On the New Copyright Protection for Non-Original Photographs in Switzerland”, *GRUR int.* 2020, 69(10), pp. 987–988: “it is not at all clear whether protecting non-original photographs beyond the scope of unfair competition law is sound policy, given that the standard incentive rationale for exclusive rights hardly applies to them”.

³⁹⁶ European Parliament and the Council, Directive 96/9/EC of 11 March 1996 on the legal protection of databases, Art. 1 (2), (O.J. L 77, 27.03.1996, p. 20).

³⁹⁷ European Parliament and the Council, Directive 96/9/EC of 11 March 1996 on the legal protection of databases, Rec 17, (O.J. L 77, 27.03.1996, p. 20).

without their informative, literary, artistic, musical or other value being affected".³⁹⁸ The components should be systematically or methodically arranged and individually accessible in one way or another. The collection should be contained in some "fixed base" and, where the arrangement of the components is not apparent, it should include technical means (such as electronic, electromagnetic or electro-optical processes) or other means, such as an index, a table of contents, or a particular plan or method of classification, to allow the retrieval of any independent material in the collection.³⁹⁹

Each of the constituent materials should thus be individually retrievable from the collection and should be separable from other materials without the value of their contents being affected.⁴⁰⁰ Each material included in the database should indeed have autonomous informative value, after extraction from the database, even if admittedly the components' autonomous informative value is higher by their arrangement in the database.⁴⁰¹ This autonomous informative value after extraction should be assessed from the perspective of each third party interested in the extracted material.⁴⁰²

A database may be protected under copyright if it meets the **originality** requirement on account of the selection or arrangement of its content.⁴⁰³ It should be demonstrated that the database is the author's "own intellectual creation". No other criteria may be applied to determine whether copyright protection is available. By contrast, the "intellectual effort and skill of creating that data" are not relevant, neither are the "significant labour and skill required for setting up that database" and, finally, "whether or not the selection or arrangement of that data includes the addition of important significance to that data" must not be considered for finding originality and consequently copyright protection.⁴⁰⁴

This copyright protection does not extend to the content of the database.⁴⁰⁵ Both the database (the container) and the data (the content) may be protected, independently, if the respective conditions for protection are met.

A database may be protected under the sui generis database right, if there has been a "qualitative and/or quantitative **substantial investment** in either the obtaining, verification or presentation of the contents".⁴⁰⁶ No originality or creativity is required in the conception or production of the database.⁴⁰⁷ The exact threshold for protection is uncertain

³⁹⁸ CJEU 9 November 2004, Fixtures Marketing Ltd / Organismos prognostikon agonon podofairou (OPAP), C-444/02, ECLI:EU:C:2004:697, Rec 17, point 29.

³⁹⁹ CJEU 9 November 2004, Fixtures Marketing Ltd / Organismos prognostikon agonon podofairou (OPAP), C-444/02, ECLI:EU:C:2004:339, point 30.

⁴⁰⁰ CJEU 9 November 2004, Fixtures Marketing Ltd / Organismos prognostikon agonon podofairou (OPAP), C-444/02, ECLI:EU:C:2004:339, point 32.

⁴⁰¹ CJEU 9 November 2004, Fixtures Marketing Ltd / Organismos prognostikon agonon podofairou (OPAP), C-444/02, ECLI:EU:C:2004:339, point 23-24.

⁴⁰² CJEU 29 Oktober 2015, Freistaat Bayern / Verlag Esterbauer GmbH, C-490/14, ECLI:EU:C:2015:735, points 27.

⁴⁰³ European Parliament and the Council, Directive 96/9/EC of 11 March 1996 on the legal protection of databases, Art. 3(1), (O.J. L 77, 27.03.1996, p. 20).

⁴⁰⁴ CJEU 1 March 2012, Football Dataco and Others / Yahoo! UK Limited and others, C-604/10, ECLI:EU:C:2012:115.

⁴⁰⁵ European Parliament and the Council, Directive 96/9/EC of 11 March 1996 on the legal protection of databases, Art. 3(2) (O.J. L 77, 27.03.1996, p. 20).

⁴⁰⁶ European Parliament and the Council, Directive 96/9/EC of 11 March 1996 on the legal protection of databases, Art. 7(1), (O.J. L 77, 27.03.1996, p. 20).

⁴⁰⁷ CJEU 9 November 2004, Fixtures Marketing Ltd / Organismos prognostikon agonon podofairou (OPAP), C-444/02, ECLI:EU:C:2004:339, point 26.

but financial, human (time, effort, skill) and technical efforts can be taken into account.⁴⁰⁸ Only the investment in the obtaining, verification or presentation of the content of the database can be taken into account, not the investment in the creation of such content. The purpose of database protection is indeed “to promote the establishment of storage and processing systems for existing information and not the creation of materials capable of being collected subsequently in a database”.⁴⁰⁹

The **obtaining** of data refers to the collecting of pre-existing materials to include them in a database. The investment in the **verification** of the contents refers to “the resources used, intending to ensure the reliability of the information contained in that database, to monitor the accuracy of the materials collected when the database was created and during its operation”.⁴¹⁰ The investment in the **presentation** of the contents refers to “the resources used to give the database its function of processing information, that is to say, those used for the systematic or methodical arrangement of the materials contained in that database and the organisation of their individual accessibility”.⁴¹¹ Examples could include the investment in the user interface or a search engine.

3.3.2.4.2. AI-related challenges: the protectability of AI-training data sets

Raw machine-generated data as such may not qualify for database protection.⁴¹² However, a curated **dataset used for training an AI solution** could be considered a “database” and be protected by the sui generis right if the criterium of substantial investment is met.⁴¹³ It is less likely to be protected under copyright.

If the training data are somewhat organised, indexed or arranged, so the images, sounds, texts, audiovisual strings are converted into labelled “data” that can be processed to be analysed, used in the training of an algorithm, enriched or otherwise used in a machine learning process, the collection of which they are part might be considered a “database”.⁴¹⁴ Moreover, the investment in the gathering, selecting, formatting or labelling might amount

⁴⁰⁸ CJEU 9 November 2004, *Fixtures Marketing Ltd / Organismos prognostikon agonon podosfairou (OPAP)*, C-444/02, ECLI:EU:C:2004:339, point 44; Study in support of the evaluation of Directive 96/9/EC on the legal protection of databases, Annex 1, In-depth analysis of the Database Directive, article by article, <https://op.europa.eu/s/oUKD>, (accessed in 14 July 2021).

⁴⁰⁹ CJEU 9 November 2004, *Fixtures Marketing Ltd / Organismos prognostikon agonon podosfairou (OPAP)*, C-444/02, ECLI:EU:C:2004:697, point 40.

⁴¹⁰ CJEU 9 November 2004, *Fixtures Marketing Ltd / Organismos prognostikon agonon podosfairou (OPAP)*, C-444/02, ECLI:EU:C:2004:697, point 43; See also CJEU 3 June 2021, *CV-Online Latvia / Melons SIA*, C-762/19, ECLI:EU:C:2021:434, paragraphs 24-28.

⁴¹¹ CJEU 9 November 2004, *Fixtures Marketing Ltd / Organismos prognostikon agonon podosfairou (OPAP)*, C-444/02, ECLI:EU:C:2004:697, point 43.

⁴¹² C. HARTMANN, J. ALLAN, P.B. HUGENHOLTZ, J.P. QUINTAIS et D. GERVAIS, “JIIP & IVIR report on AI and IP”, *op. cit.*, p. 92.

⁴¹³ M. LEISTNER, “Big Data and the EU Database Directive 96/9/EC: Current Law and Potential for Reform”, 2018, <https://ssrn.com/abstract=3245937> or <http://dx.doi.org/10.2139/ssrn.3245937>, (accessed in July 2021).

⁴¹⁴ In this sense: J. DREXL et al., *Artificial Intelligence and Intellectual Property Law-Position Statement of the Max Planck Institute for Innovation and Competition of 9 April 2021 on the Current Debate*, *Max Planck Institute for Innovation Competition Research* 2021 No 21-10, p. 7; Several databases might even be distinguished: the collection of training data and the collection of values resulting from training the model. In this sense: A. BENSAMOUN, J. FARCHY and P.-F. SCHIRA, “Mission Intelligence artificielle et Culture”, Paris, *CSPLA* 2020, I, p. 30; See also CJEU 3 June 2021, *CV-Online Latvia SIA / Lemons SIA*, C-762/19, ECLI:EU:C:2021:434, paragraph 28: “(...) it is for the referring court to examine, where appropriate, whether the conditions laid down in Article 7 of Directive 96/9 are satisfied for the grant of protection by the *sui generis* right, including whether the meta tags provided by CV-Online could themselves be regarded as constituting a substantial part of the protected database”.

to a “substantial investment” in the “obtaining, verifying or presenting of the contents of the database.

This might even be the case, if this dataset is collected with the sole purpose of training algorithms and even if the data set is, as such, not meant to be consulted by the public.

By contrast, the cultural output, in the form of a musical composition or a recorded performance of musical creation, an audiovisual production, an artistic or photographic image is unlikely to be protected as a database.⁴¹⁵

The consequence might be that even training datasets might be protected under exclusive rights and that the prior authorisation of the maker of the database might be required for the reutilisation or the extraction of the contents of the database for training purposes (see however 3.3.4.2 and 3.3.6). This possibility does not seem to be of particular concern for the stakeholders. Among all participants of the study, only one interviewee (a representative of a company developing an AI application for the creative sector) expressed concerns about the database protection of datasets used for training AI solutions. It will remain to be seen how the business practices will evolve in this respect.

3.3.2.5. Cumulative protection – stacking of rights and resulting incoherencies

AI-related challenges (1): the possibility of cumulative protection

The protection under copyright, related rights and sui generis rights is cumulative, meaning that “creative content” may be protected as a work, a performance, a recording (audio or film), contained in a protected database. This stacking of rights might have consequences in the AI context both at the input and output stages.

The use of one digital file (e.g. a music video) can consequently cumulatively trigger copyright, performers' rights and several producers' rights. This means that the AI developer or user must clear the producers' related rights (unless the use is covered under an exception), in addition to the authors' and performers' rights.

With the massive use of data to train an AI solution, the clearing of this accumulation of rights can be challenging for the developers and users of AI solutions. Moreover, where the rights are not held by one person, some right holders may authorise the use of their subject-matter, while other refuse to authorise the same. This means that the use of the file, the recording or other fixation containing the different protected matters will only be authorised if all right holders agree.

This issue is even more complex if the data (containing protected subject-matter) are contained in several databases. This could lead to a situation where the AI developer can have access to the same data under different conditions, determined by the different database maker.

Finally, new rights may arise in this preparatory process: where the maker of the AI solution arranges the training data or enriches them with labels or meta-data, they may be investing in the obtaining, verification or presentation of the content of a database, which results in exclusive rights in the newly created database. Where the same database is made available for training other algorithms, the maker of that database may exercise their rights and negotiate a remuneration.

⁴¹⁵ This might be different for informational creations, such as the topographical maps under scrutiny in the Verlag Esterbauer ruling of the CJEU (Judgment in *Verlag Esterbauer* (C-490/14)).

As a result, this accumulation of rights might render it difficult for AI developers in the creative sector to identify all relevant rights related to the data they need to rely on at the input level.

At this stage, the stakeholders have not mentioned any specific difficulties relating to the use of creations, protected under several exclusive rights, for training AI solutions. Several reasons may explain this: rightholders might not be aware that their creations are used as AI input, the AI developer may not have requested permission or, on the contrary, they may have cleared all rights. This may also change when the demand for AI training data changes and a market for such data develops.

AI-related challenges (2): a potential shift from copyright to related rights to protect AI autonomously generated output

The capacity to generate by AI **cultural output** autonomously, i.e. with minimum human intervention, is unevenly developed among the cultural sectors. While it does not seem likely that audiovisual productions will be generated in a fully automated and autonomous way any time soon, fast progress is made in other sectors and we might experience quasi-autonomously generated photograph-like images, artistic images and even music, which may be rendered by human or non-human performers.

AI generated visual creations, such as photograph-like images or pieces of visual art, may thus not be protected under copyright or related rights (unless related rights exist for non-original photographs under national law). Other creations need to be presented, played or performed in some form for the public to enjoy them (typically music, theatre, audiovisual creations, films or dance choreographies): the rendering by non-humans of such AI-generated creations is not protected under the performer's rights (e.g. electronic music). Even the playing of AI-generated output by humans may raise questions, if performances are only protected to the extent that they are a rendering of copyright-protected works.

While the absence of (a qualified) human intervention may cast a doubt on the protection of the AI output under copyright and performers' rights, this is less the case for the protection of producers of phonograms or first fixations of films. These fixations may or may not be recordings of protected works and/or protected performance and even without meeting any threshold of creativity, originality or investment, they may result in an exclusive right for the producer of the phonogram or first fixation of film.

The application of the current principles might lead to a **shift in the centre of gravity of the protection from the human contributions to the non-human contributions** to the AI-generated cultural output. Traditionally, the human authors and performers have a central role in the exploitation of the result, in addition to rights for the producers of sound or audiovisual fixations. With autonomously created AI output, the human factor may disappear entirely. While the cultural output may be generated without the human author and, potentially, performer, it will in many cases still be fixated (temporarily) or communicated, thus creating rights for producers. In those cases, only the producers will have exclusive rights, allowing them to control the exploitation of the AI-generated cultural output (see also 3.3.3). The utility of such exclusive rights might be questioned to the extent that the investment in fixations, that is traditionally underlying the attribution of exclusive rights (e.g. the investment in the recording of a performance of musical works or an audio-visual production), may be absent or negligible for generating and fixating AI-generated cultural output.

This in turn could lead to **technology-driven legislative dissonance** where the *ratio legis* and the application of the law are no longer in line: the protective regime traditionally intended to cover human-generated, cultural content and, incidentally, related recordings

will now apply to protect anything but the human-generated output in an AI-driven creative process. From a public interest point of view, depending on the evolution in the development and adoption of AI in the cultural sector, the question may rise whether overall the right balance is struck between the protected subject-matter (considering the *ratio legis*) and the freedom of information and enterprise (see also 3.4.2.4).

In the same vein, it was reported in one expert interview that some audiovisual output, generated using innovative data processing techniques applied to a sports competition, was actually not commercialised because the stakeholders were not certain whether the resulting output could be protected under an exclusive right and, if so, how the rights would be distributed (see also below 3.3.3).⁴¹⁶

3.3.3. Authorship and Ownership – Who is the holder of rights to AI output and how are these rightholders identified?

Closely related to the question whether AI-outputs can be protected under copyright and related rights is the issue of authorship and ownership of rights. Only the right holder can exercise the economic and moral rights protecting the work, performance, fixation or database. The various rightholders must be easily identifiable to the user of protected cultural content, who may need to negotiate a licence or who may be confronted with an infringement claim. The rightholders enjoy the comfort of a legal presumption of ownership but in an AI-context this presumption may give rise to new challenges.

3.3.3.1. Legal conditions of authorship and ownership

While the initial right holders of copyright and performers' right are in principle human, this is not the case for the producers and database makers. This distinction has obvious relevance for autonomously generated AI-output.

3.3.3.1.1. Authorship of copyright

At the European level, the notion of authorship **does not have a general definition**.⁴¹⁷ The EU legislator touches upon this notion only with respect to specific types of works. Hence, concerning computer programs, Art. 2(1) Software directive⁴¹⁸ provides that "the author of a computer program shall be the natural person or group of natural persons who have created the program or, where the legislation of the Member State permits, the legal person designated as the rightholder by that legislation". The Database directive⁴¹⁹ uses equivalent wording. According to Art 4(1): "The author of a database shall be the natural person or group of natural persons who created the base or, where the legislation of the Member States so permits, the legal person designated as the rightholder by that legislation." Lastly, and concerning cinematographic or audiovisual works, Art. 2(1) of the Term protection directive⁴²⁰ determines that "the principal director of a cinematographic or audiovisual work shall be considered as its author or one of its authors. The EU Member States shall be free to designate other co-authors".

⁴¹⁶ Confidential expert interviews conducted by the study team.

⁴¹⁷ CJEU 9 February 1012, *Luksan v. van der Let*, C-277/10, EU:C:2012:65.

⁴¹⁸ European Parliament and the Council, Directive 2009/24/EC of 23 April 2009 on the legal protection of computer programs, (O.J. L 111, 05.05.2009, p. 16).

⁴¹⁹ European Parliament and the Council, Directive 96/9/EC of 11 March 1996 on the legal protection of databases, (O.J. L 77, 27.03.1996, p. 20).

⁴²⁰ European Parliament and the Council, Directive 2006/116/EC of 12 December 2006 on the term of protection of copyright and certain related rights, (O.J. L 372, 27.12.2006, p. 12).

Usually, a work's authorship is recognised, and the copyright's ownership is vested in the natural person who created the work (**'creator doctrine'**).⁴²¹ This principle knows certain exceptions both in civil and common law countries. Such exceptions are more common in the common law tradition, due to the economic logic supporting copyright protection, whereas civil law countries animated by a personhood tradition are more reluctant to admit exceptions. However, examples can be found in both traditions, especially when it comes to establishing initial rights ownership. For instance, the initial rights' ownership might, in the UK, be vested in the employer (in case of work made for hire) or, in France, in a corporation in the case of collective work. Despite these exceptions, it is generally assumed that, when it comes to authorship, the author is "the flesh-and-blood, natural person who conceived and executed the work".⁴²² This requirement for a human person as a matter of principle goes hand-in-hand with the subject matter of copyright which demands the operation of free and creative choices.

A notable exception needs to be mentioned: the UK and Irish special regime for **computer-generated works**. These works are defined in the UK as those "generated by a computer in circumstances such that there is no human author of the work"⁴²³ (and in Ireland as "generated by computer in circumstances where the author of the work is not an individual"⁴²⁴

Consequently, according to sec. 9(3) of the UK act and 21 of the Irish act "in the case (...) work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken." UK law also provides in such cases a shorter copyright duration (sec. 12(7) UK act and 30 Irish act) and the absence of moral rights (sec 89 and 81 UK act). Since the adoption of these two acts, the concept of copyright originality however evolved following inter alia the above-mentioned case law of the CJEU.⁴²⁵ Today, to be original and therefore protectable under copyright law, a work must be the result of free and creative choices. This requirement, therefore, makes impossible the existence of work "in circumstances such that there is no human author of the work". The two legal regimes are therefore criticised as incompatible with EU law.⁴²⁶ It should lastly be mentioned that the UK government recently launched a consultation on "Artificial intelligence and intellectual property".⁴²⁷ Summarising and reacting to the contributions made concerning the legal regime of computer-generated works, the UK government recognises this issue and concludes that "there is some ambiguity about the status of AI-assisted works. In light of the above, clarification of these provisions may be needed."⁴²⁸

3.3.3.1.2. Ownership of related and sui generis rights

In the following section the ownership of different holders of related rights is discussed:

⁴²¹ GOLDSTEIN and HUGENHOLTZ, *supra* note 89 at 229.

⁴²² GOLDSTEIN AND HUGENHOLTZ, *supra* note 89 at 229.

⁴²³ UK Copyright, Design and Patent Act of 1988, sec. 178.

⁴²⁴ Ireland Copyright and Related Rights Act of 2000, Art. 2(1).

⁴²⁵ See section 3.3.2.1. As a matter of clarification, following the Brexit, European law no longer applies in the United Kingdom.

⁴²⁶ B. HUGENHOLTZ, J.P. QUINTAIS, "Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?", *IIC* 2021, §4.3.

⁴²⁷ Gov UK, <https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views>, (accessed in July 2021).

⁴²⁸ Gov UK, <https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views/artificial-intelligence-call-for-views-copyright-and-related-rights>, (accessed in July 2021).

Performers

"Performers" are "actors, singers, musicians, dancers, and other persons who act, sign, deliver, declaim, play in, or otherwise perform literary or artistic works" (art. 3(a) Rome Convention). Only natural persons can be performers and only artistic performances can give rise to performers' rights (although the Rome Convention does not explicitly require the "performer" to be human). Hence, no protection is given to purely technical performances.

The rendering of an AI-generated musical score by software imitating a human voice or musical instruments will not vest any performers' rights in the user of the software. Similarly, no performer can hold rights on the manifestations of computer animations or cartoons (although the designers of those animations may be the authors or co-authors of an audiovisual work).

Phonogram producers

Contrary to performers, the holders of the other related rights (film producers, phonogram producers, broadcasters, publishers) may be **natural persons or legal persons**. The protection of their subject matter **aims to incite investment**⁴²⁹ and is not subject to any condition of "originality" or "creativity" (and the holders of the traditional neighbouring rights do not have to establish any investment, contrary than the database makers).

At the international level, the phonogram producer is the (legal or natural) person who takes the initiative and has the responsibility for the fixation of sounds.⁴³⁰ The European directives do not define the notion of "phonogram producers".

While the **investment** in such a phonogram might be implied⁴³¹, the European directives do not formally require that a (substantial) investment in the recording of sounds be demonstrated in order to enjoy the producer's related rights. The reference to the initiative and the responsibility for the fixation entails that the person who materially makes the fixation may not be considered the "phonogram producer".⁴³²

Where a user can instantaneously trigger the production of an audio file, of which the composition and the performance are generated by an AI solution, the question arises whether such AI output can be protected under the phonogram producer's related right

⁴²⁹ European Parliament and the Council, Directive 2001/29/EC of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, Rec 10, (O.J. L 167, 22.6.2001, p. 10).

⁴³⁰ Art. 2 (d) of WIPO Performances and Phonograms Treaty (WPPT), 1996.

⁴³¹ European Parliament and the Council, Directive 2006/115/EC of 12 december 2006 on rental right and lending right and on certain rights related to copyright in the field of intellectual property, Rec 5, (O.J. L 27.12.2006, p. 28): "The creative and artistic work of authors and performers necessitates an adequate income as a basis for further creative and artistic work, and the **investments required particularly for the production of phonograms and films are especially high and risky**. The possibility of securing that income and recouping that investment can be effectively guaranteed only through adequate legal protection of the rightsholders concerned" (emphasis added); See also Rec 10 of InfoSoc Directive, European Parliament and the Council, Directive 2001/29/EC of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, (O.J. L 167, 22.6.2001, p. 10), rec. 10: "If authors or performers are to continue their creative and artistic work, they have to receive an appropriate reward for the use of their work, as must producers in order to be able to finance this work. **The investment required to produce products such as phonograms, films or multimedia products, and services such as "on-demand" services, is considerable**. Adequate legal protection of intellectual property rights is necessary in order to guarantee the availability of such a reward and provide the opportunity for satisfactory returns on this investment" (emphasis added); See in this sense: Pelham, *op. cit.*, points 30, 44.

⁴³² F. BRISON, "WPPT, Art. 2", *Concise European Copyright Law*, Brussels, Kluwer, 2016, 194; "IVIR/JIITC report", *op. cit.* p.90.

and, if so, who should be considered the phonogram producer. The user of the AI solution has triggered the production of the AI music output, by pushing a button or by setting certain preferences or parameters, but they could only initiate the production of the audio file because the AI solution offered this possibility. In those cases, it is not so straightforward which person has taken the initiative and the responsibility for fixating the sounds. Contrary to the analogue period in which the phonogram producer's rights originate, when the phonogram referred to an act of "recording" the "performance" of a "work" (or other, unprotected sounds), the fixation in an electronic and now digital context does not necessarily refer to the recording of a sound from an external source. The roles of "authors", "performers" and "producers" are consequently more difficult to distinguish.

Film producers

The "**producer of the first fixations of films**" or "film producer" is not defined as such. The rightholder of these related rights is however the person who has made the "first fixation" of the film. By analogy to the phonogram producer, the film producer is not the person who has materially made the fixation but the natural or legal person who has taken the initiative to make the fixation and under whose responsibility this is done.

Broadcasters

The notion of "**broadcasting organisations**" or "broadcasters" is not defined as such in the Rental and Lending Rights Directive, nor in the InfoSoc Directive nor in Rome Convention. The Rome Convention provides a definition of "broadcasting" as "*the transmission by wireless means for public reception of sounds or of images and sounds*" (art. 3(f) Rome Convention). It covers traditional "wireless" broadcasting (e.g. radio and television).

The Rental and Lending Rights Directive adds that it also protects broadcasts by cable or satellite (art. 7(2) Rental and Lending Rights Directive).

The broadcasters' rights do not have primary relevance for the adoption of AI in the cultural sector.

Database maker

A "**maker of a database**" or "database producer", i.e. "*the person who takes the initiative and the risk of investing*" (recital 41 Database Directive), may be granted a *sui generis* right for the protection of the database under the Database Directive. To this end, they have to establish that "*there was a substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilisation of the whole or a substantial part*" (art. 7(1) Database Directive). Any kind of investment is taken into account and can thus either be of a "qualitative" and/or "quantitative" nature.

The natural or legal person collecting the training data and the person verifying or enriching those data in the preparatory process might be the makers of a database (should their investment result in a protected database).

It may not be so obvious to identify the rightholder of the database rights, to the extent that the different roles may be dissociated.

Training data can be found through platforms where individual users may upload data (e.g. images) and other users might label these data (e.g. identifying cats, bikes or humans in the images), which are then made available to AI developers. The training data, as a whole, may then be exploited as a database, but it is uncertain whether the distinct investments of the different users may be "pooled" to assess the "investment" in the obtaining or

verification of the data and to consider the dataset a protected database to which the rights are (jointly) held by all contributing users.

Press publishers

The new DSM directive has created a new right that protects certain uses of “press publications”. This notion is defined but the holder of this right (art. 2(4) DSM Dir.), the “publisher of press publications” is not.

3.3.3.2. Legal conditions of presumptions of authorship or ownership

According to Art. 5 Enforcement Directive, both the authors of work protected by copyright and the owners of related rights can rely on a **presumption of authorship or ownership**, to enforce their rights.⁴³³

This presumption is rebuttable and holds therefore until proof to the contrary. A defendant can consequently demonstrate that the plaintiff is not the actual author or performer or that they do not hold any rights to the protected creation and that they are not entitled to claim infringement of their rights – but the burden of proof is upon the defendant.

3.3.3.3. AI-related challenges: claims of false/unverifiable authorship

The author holds a central position in the system of copyright protection and the designation of the initial holder of copyright has thus considerable consequences. As explained sub 3.3.2.1, a given output needs to be the result of human creative choices in order to qualify for copyright protection. In the absence of creative decisions, the output cannot enjoy copyright protection. There is no “work” without a (human) author.

For some types of output, it might be difficult or even impossible to determine from the output alone whether it is the result of human and/or non-human decisions. While more complex cultural productions (such as cinematographic productions) that are entirely generated by an AI solution may not trick the public to believe that they were human-made, this might not be so easy to discern for other productions (e.g. music, photographs). For instance, if a piece of music was generated with the help of an AI solution, such as the AI composer AIVA,⁴³⁴ it will from the output alone be almost impossible to determine how the music was created. The producers of such AI-generated output may be tempted to conceal the machine-origin of such production or even falsely attribute the authorship of AI autonomous production to a human and claim copyright protection.

In an attempt to measure the intensity of the phenomena, participants of the Delphi study were asked if, in their opinion, people using AI to generate cultural outputs are attributing themselves authorship of work.

⁴³³ European Parliament and the Council, Directive 2004/48/EC of 29 April 2004 on the enforcement of intellectual property rights, Art. 5, (O.J. L 157, 30.04.2004): “for the author of a literary or artistic work, in the absence of proof to the contrary, to be regarded as such, and consequently to be entitled to institute infringement proceedings, it shall be sufficient for his/her name to appear on the work in the usual manner; the provision (...) shall apply mutatis mutandis to the holders of rights related to copyright with regard to their protected subject matter.”

⁴³⁴ See case study presented in section 3.2.2.3

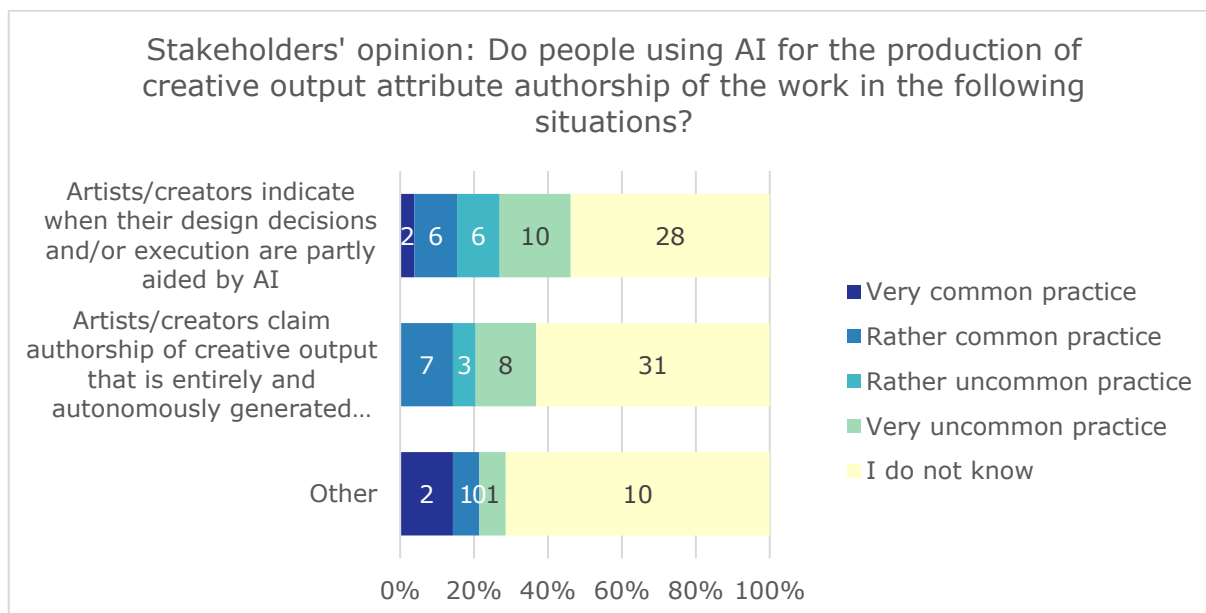


Figure 34: Stakeholders' opinion on the attribution of authorship when using AI for the production of creative output

Source: Technopolis Group Survey

As shown in Figure 40, a significant proportion of the participants in the survey state that they do not know whether or not creators indicate that their artistic pieces are partially generated by AI systems. This incapacity can be analysed as revealing of the difficulty (or even impossibility) from the input alone to determine how the latter was generated. However, it should also be noted that the AI solutions considered within this study are still at their developing stage, which might explain that stakeholders are not fully aware of their (potential) impact. An organisation representing authors and rightholders for instance believes that presently there is always human involvement and therefore an author (not a false author). Hence, it stated that by performing operations that add substantial value to the outcome, one could think of a collaboration or co-ownership. Nevertheless, some of the participants appear on the contrary well-aware of the raised challenge. One of them, for instance, acknowledges that AI tools are used very differently in the creative industry, for instance, from minimal automated post-processing changes (e.g. colour adjustment in photos) to fully AI-generated content. Therefore, for this participant, one should distinguish AI-generated output with artistic significance (created by humans with the aid of AI) and output entirely generated by AI).

This issue of false authorship was addressed by the IVIR/JIPP study, which further insists that the presumption of authorship could further facilitate this type of behaviour.⁴³⁵ By virtue of this presumption, it is indeed sufficient for the author's name to appear on the work in the usual manner, for the author to be entitled to institute infringement proceedings.⁴³⁶ In many jurisdictions, this presumption can also be invoked by legal persons in infringement proceedings (e.g. in Belgium and France). The presumption can be rebutted if it can be demonstrated that the person marked as the author is in fact not

⁴³⁵ C. HARTMANN, J. ALLAN, P.B. HUGENHOLTZ, J.P. QUINTAIS et D. GERVAIS, "JIIP & IVIR report on AI and IP", *op. cit.*, p. 8.

⁴³⁶ European Parliament and the Council, Directive 2004/48/EC of 29 April 2004 on the enforcement of intellectual property rights, Art. 5, (O.J. L 157, 30.04.2004)

the author of the creation. This proof may however be difficult to come by: how can it be demonstrated that a creation is not the result of human creativity, but of machine-led processes if the human senses cannot distinguish between both, especially when the creation process remains unknown to the public?

This impossibility to distinguish between cultural output as the result of creative choices and those automatically generated with AI solutions could become problematic in the future. First, if consumers are not capable of making the difference between human and AI cultural outputs, then it is likely that the latter, which will be cheaper, will capture a substantial part of the revenues of traditional authors and composers. Second, and more directly related to copyright law, if CMOs are also in the impossibility of differentiating between AI autonomously generated outputs and works deserving protection, there is a risk of seeing some actors taking advantage of this deficiency to claim revenues due to human creators.

3.3.4. Economic rights and infringement – Which acts in an AI process have copyright relevance?

The development and the use of AI solutions in the cultural sector rely on complex technical processes, in which data are copied, transmitted and transformed. Some of these acts might be considered relevant acts for copyright or related rights, in particular “reproductions” or “extractions”. Such acts require the right holders’ consent – unless an exception applies – meaning that it is important for the developer or user of an AI solution to know which acts are protected under copyright.

3.3.4.1. Copyright and related rights, economic rights and infringement

The author enjoys exclusive rights under copyright, which can be discussed in the broad categories of right of “reproduction” and of “communication to the public”. The holders of neighbouring rights enjoy the same rights (with some nuances).

3.3.4.1.1. *Copyright: reproduction and communication to the public*

Article 2 of the InfoSoc Directive⁴³⁷ defines the reproduction right as ‘the exclusive right to authorize or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part’ of the authors’ works.

Until now, the CJEU has proposed ‘a consistently broad interpretation of the reproduction right’.⁴³⁸ In *Infopaq I*, the Court considered that several reproductions occur during a data capture process, such as the creation of a TIFF file by scanning, the conversion into a searchable text file (by an OCR or Optical Character Recognition process), the storing of an extract, and the printing of that extract on a paper medium.⁴³⁹ Under *Infopaq I*, partial

⁴³⁷ European Parliament, Directive 2001/29/EC of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, (O.J. No. L 167/10, 22.06.2001).

⁴³⁸ M. LEISTNER, “Europe’s Copyright Law Decade: Recent Case Law of the European Court of Justice and Policy”, *Perspectives, Common Market Law Review* 2014, 51, p. 569.

⁴³⁹ CJEU 16 July 2009, *Infopaq International A/S/ v. Danske Dagblades Forening*, EU:C:2009:465, paragraph 51.

reproductions may be prohibited insofar as the part of the work reproduced fulfils the condition of copyright protection, i.e., reflects the author's own intellectual creation.^{440 441}

Similarly, in *Premier League*, the Court held that 'the reproduction right extends to transient fragments of the works within a satellite decoder and on a television screen, provided that those fragments contain elements which are the expression of the author's own intellectual creation'.⁴⁴² That 'transient fragments' of works are covered shows how encompassing the right of reproduction is. In those two cases, the operations discussed by the Court (i.e., within the data capture and conversion process or within the satellite transmission process) are essentially technical. Therefore, the Court appears to support a technical view of reproduction, but, in reality, it applies a highly conceptual and very broad notion of reproduction to various technical processes and their outcomes: many technical processes could fall under the reproduction umbrella because of the technological neutrality principle embedded in its definition⁴⁴³.

If the application of the reproduction notion to the processes under scrutiny in *Infopaq I* and *Premier League* appears very technical (and if lawyers are expected to analyse complex technical processes to identify each reproduction), this is made possible by the **broad and formalistic notion of reproduction**, completely disconnected from the real context of the process and from the requirement of having acts of (economic) exploitation.⁴⁴⁴ Formalism definitely prevails over substance, and the absence of any requirement to make a substantial analysis of the situation facilitates the extension of the legal concept. This technical reading was confirmed in the CJEU's decisions on the exception for temporary acts of reproduction (art. 5(1) InfoSoc Dir)⁴⁴⁵.

More significantly, in order to delineate the reproduction right, the CJEU so far does **not refer to the presence of a 'public'** neither does it put forward the condition that the use should be commercial or have a for-profit nature. The "public" could however be a proxy for an economic exploitation, even under the reproduction right. In order to bring the "reproduction right" in line with the protection of the exploitation of the work, the reproduction should address a public. This is clear from the analysis of the communication to the public right by the CJEU, but the same reading is arguably applicable to the reproduction. Similarly, the CJEU has examined in *Pelham* whether a sampled musical

⁴⁴⁰ CJEU 16 July 2009, *Infopaq International A/S/ v. Danske Dagblades Forening*, EU:C:2009:465, paragraph 32 and ff.

⁴⁴¹ This criterion for finding partial reproduction under copyright seems to differ from the criterion for finding a partial reproduction under the producer's related right, as interpreted in the CJEU's *Pelham*-decision (see below).

⁴⁴² CJEU 4 October 2011, *Football Association Premier League and Others v. Media Production Services Ltd.*, C-403/08 & C-429/08, EU:C:2011:631, paragraph 159.

⁴⁴³ Even for software, a highly technical item that raises issues when subsumed under copyright, the Software Directive proposes an undefined and neutral notion of reproduction without giving it a technical content. Indeed, Article 4(a) of the Software Directive does not say that loading, displaying, running, transmission, or storage of a computer program are (technical) reproductions, it just states that 'in so far as' those acts of loading, etc. 'necessitate such reproduction, such acts shall be subject to authorisation by the rightholder'.

⁴⁴⁴ The reproduction right appears general and technology-neutral, not only under the Berne Convention, also under the InfoSoc Directive. Thus, InfoSoc Directive, art. 2 does not offer a technical notion. The Court's interpretation remains very conceptual, broad, and synthetic, and therefore all the (technical) processes (whatever their input and output) can be sliced into many highly specific reproductions. See however CJEU 29 July 2019, *Pelham GmbH and Others / Ralf Hütter and Florian Schneider-Esleben*, C-476/17, EU:C:2019:624, *infra* sub 3.3.4.1.2.

⁴⁴⁵ CJEU 4 October 2011, *Football Association Premier League and Others v. Media Production Services Ltd.*, C-403/08 & C-429/08, EU:C:2011:631, paragraph 160 et seq.; CJEU 5 June 2014, *Public Relations Consultants Association Ltd. / WSPaper Licensing Agency Ltd and Others*, C-360/13, ECLI:EU:C:2014:1195.

sequence is recognisable to the ear of a listening *public* in order to find a reproduction of the phonogram.⁴⁴⁶

Furthermore, concerning the related right of the phonogram producers, the Court of Justice proceeded to a shift in its case law regarding the interpretation of the right of reproduction, especially in the *Pelham* decision. It is not sufficient to find a technical copy in order to establish a partial reproduction of the phonogram, it is also required that the sample be recognisable (see also below 3.3.4.1.2 and 3.4.1.1.3). It is however uncertain if this interpretation will be extended to the interpretation of the reproduction right under copyright (see below 3.4.1.1.3).

Concerning infringement, in copyright, it commonly requires proving the following two elements:

- **Substantial similarity:** It entails identifying the boundary of the work, and whether parts of a work qualify as a work (and what parts are original). This assessment is sometimes difficult as the case law does not offer clear guidance as to how the boundary of a work is to be determined and how the unprotected ideas should be filtered and discarded. Secondly, the infringement analysis requires a comparison of the protected work (adequately identified and filtered) with what the defendant has taken.
- **Copying:** In the case of identity or sufficient (substantial) similarity between the two, copying must still be established as independent creation lies outside copyright's scope. This assessment is of course left for the discretion of the courts and has led to complex case law. The copying requirement may be satisfied even though the defendant did not consciously copy. Copying is often difficult to prove directly, and it is usually indirectly established by showing that the defendant could have accessed the work. This derivation or 'causal connection' between the work and what is used by the alleged infringer must be shown, and this is complex.

3.3.4.1.2. *Related rights*

Performers

Performers enjoy economic rights of **different types**: the **exclusive economic rights** allow them to control the exploitation of the performances and **the remuneration rights** ensure a compensation where their exclusive rights are restricted by virtue of exceptions or other limitations. These rights are generally vested in the performers as natural persons.

A summary of the **telos of performers' right** can be found in Recital 5 of the Rental Directive. Accordingly, the recognition of these legal prerogatives to performers is a necessity for ensuring to them an adequate remuneration. Such a remuneration is the "basis for further creative and artistic work (...)" and "the possibility for securing that income and recouping that investment can be effectively guaranteed only through adequate legal protection of the rightholders concerned".

The **Rental** and **InfoSoc Directive harmonise** the performers' economic rights, in particular the rights to authorise or to prohibit: the fixation of their performances⁴⁴⁷, the

⁴⁴⁶ CJEU 29 July 2019, *Pelham GmbH and Others / Ralf Hütter and Florian Schneider-Esleben*, C-476/17, EU:C:2019:624.

⁴⁴⁷ European Parliament and the Council, Directive 2006/115/EC of 12 December 2006 on rental right and lending right and on certain rights related to copyright in the field of intellectual property, Art. 7(1), (O.J. L 27.12.2006, p. 28).

rental or lending of fixations of their performances⁴⁴⁸, the broadcast and communication to the public of their performances⁴⁴⁹, the distribution in respect of fixation of their performances⁴⁵⁰, the reproduction of their performances⁴⁵¹ and the right to make the performance available to the public.⁴⁵²

- With regard to the notion “**communication to the public**”, the CJEU has ruled in *Reha Training* that it should be given the same interpretation under copyright (art. 3 InfoSoc Directive) as under the related rights (art. 8(2) Rental Directive), despite the different context and objectives pursued under both provisions⁴⁵³. Unless otherwise indicated by the EU legislature, the same is true for the other exclusive rights of reproduction and distribution.
- Performers enjoy (unwaivable) **remuneration rights**, such as the right to an “equitable remuneration”, for the broadcasting of recorded performances⁴⁵⁴ or following the transfer.

Phonogram and film producers

The phonogram producer holds economic rights, both exclusive rights and remuneration rights⁴⁵⁵. The film producer holds exclusive rights.

⁴⁴⁸ European Parliament and the Council, Directive 2006/115/EC of 12 December 2006 on rental right and lending right and on certain rights related to copyright in the field of intellectual property, Art. 3 (1)(b), (O.J. L 27.12.2006, p. 28).

⁴⁴⁹ European Parliament and the Council, Directive 2006/115/EC of 12 December 2006 on rental right and lending right and on certain rights related to copyright in the field of intellectual property, Art. 8(1), (O.J. L 27.12.2006, p. 28).

⁴⁵⁰ European Parliament and the Council, Directive 2006/115/EC of 12 December 2006 on rental right and lending right and on certain rights related to copyright in the field of intellectual property, Art. 9(1) (b), (O.J. L 27.12.2006, p. 28).

⁴⁵¹ European Parliament and the Council, Directive 2001/29/EC of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, Art. 2 (b), (O.J. L 167, 22.6.2001, p. 10).

⁴⁵² European Parliament and the Council, Directive 2001/29/EC of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, Art. 2 (2) (a), (O.J. L 167, 22.6.2001, p. 10).

⁴⁵³ CJEU 31 May 2013, *Reha Training Gesellschaft für Sport- und Unfallrehabilitation mbH / Gesellschaft für musikalische Aufführungs- und mechanische Vervielfältigungsrechte eV (GEMA)*, C-117/15, ECLI:EU:C:2016:379, paragraph 28.

⁴⁵⁴ European Parliament and the Council, Directive 2006/115/EC of 12 December 2006 on rental right and lending right and on certain rights related to copyright in the field of intellectual property, Art. 8(2), (O.J. L 27.12.2006, p. 28).

⁴⁵⁵ European Parliament and the Council, Directive 2006/115/EC of 12 December 2006 on rental right and lending right and on certain rights related to copyright in the field of intellectual property, art. 8(2), (O.J. L 27.12.2006, p. 28): “Member States shall provide a right in order to ensure that a single equitable remuneration is paid by the user, if a phonogram published for commercial purposes, or a reproduction of such phonogram, is used for broadcasting by wireless means or for any communication to the public, and to ensure that this remuneration is shared between the relevant performers and phonogram producers. Member States may, in the absence of agreement between the performers and phonogram producers, lay down the conditions as to the sharing of this remuneration between them”.

The **exclusive rights** encompass the rights of **reproduction**⁴⁵⁶, **distribution**⁴⁵⁷, **broadcasting** and **communication** to the public⁴⁵⁸ and **making available** to the public⁴⁵⁹. These notions can be understood in the same way as under copyright, even when the legal consequences for authors and holders of related rights differ (e.g. exclusive right v. remuneration right)⁴⁶⁰.

With regard to the **reproduction right**, the CJEU ruled in *Pelham* that “sampling”, i.e., “taking a sample from a phonogram, most often by means of electronic equipment, and using the sample for the purposes of creating a new work”⁴⁶¹, involves a “reproduction of a sound sample, even if very short” that “must, in principle, be regarded as a reproduction ‘in part’ of that phonogram” and is covered under the exclusive right of the producer (art. 2(c) InfoSoc Dir)⁴⁶².

It is however not sufficient to demonstrate that any portion of a copy of the recording has been used: a “fair balance” should be struck between the interests of the rightholders (a high level of protection, the protection of the producer’s investment) and the protection of the interests and fundamental rights of users of protected subject matter as well as of the public interest⁴⁶³. It follows that “where a user, **in exercising the freedom of the arts**, takes a sound sample from a phonogram in order to use it, **in a modified form unrecognisable to the ear**, in a new work, it must be held that such use does not constitute ‘reproduction’ within the meaning of Article 2(c) of Directive 2001/29”.⁴⁶⁴

With this decision, the CJEU seems to steer away from the literal and technical interpretation of the reproduction right. This new interpretation, according to the Court, would be consistent with the specific objective of protecting the producer’s investment. At least, this case law proposes a stricter test for finding a partial reproduction of a phonogram, which takes into account users’ rights and the interests of the general public, and requires an audibly “recognisable” partial copy in the modified phonogram⁴⁶⁵.

The Court applies a different criterion to the **distribution right**, which allows the phonogram producer to control the making available to the public, by sale or otherwise,

⁴⁵⁶ European Parliament and the Council, Directive 2001/29/EC of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, Art. 2(c), (O.J. L 167, 22.6.2001, p. 10).

⁴⁵⁷ European Parliament and the Council, Directive 2006/115/EC of 12 December 2006 on rental right and lending right and on certain rights related to copyright in the field of intellectual property, Art. 9, (O.J. L 27.12.2006, p. 28).

⁴⁵⁸ European Parliament and the Council, Directive 2006/115/EC of 12 December 2006 on rental right and lending right and on certain rights related to copyright in the field of intellectual property, art. 8, (O.J. L 27.12.2006, p. 28).

⁴⁵⁹ European Parliament and the Council, Directive 2006/115/EC of 12 December 2006 on rental right and lending right and on certain rights related to copyright in the field of intellectual property Art. 3(2), (O.J. L 27.12.2006, p. 28).

⁴⁶⁰ CJEU 31 May 2013, *Reha Training Gesellschaft für Sport- und Unfallrehabilitation mbH / Gesellschaft für musikalische Aufführungs- und mechanische Vervielfältigungsrechte eV (GEMA)*, C-117/15, ECLI:EU:C:2016:379, paragraph 28.

⁴⁶¹ CJEU 29 July 2019, *Pelham GmbH and Others / Ralf Hütter and Florian Schneider-Esleben*, C-476/17, EU:C:2019:624, paragraph 35.

⁴⁶² CJEU 29 July 2019, *Pelham GmbH and Others / Ralf Hütter and Florian Schneider-Esleben*, C-476/17, EU:C:2019:624, paragraph 29.

⁴⁶³ CJEU 29 July 2019, *Pelham GmbH and Others / Ralf Hütter and Florian Schneider-Esleben*, C-476/17, EU:C:2019:624, paragraph 32.

⁴⁶⁴ CJEU 29 July 2019, *Pelham GmbH and Others / Ralf Hütter and Florian Schneider-Esleben*, C-476/17, EU:C:2019:624, paragraph 31 – emphasis added.

⁴⁶⁵ It is however not clear whether this decision also affects the “partial reproduction” test under copyright. There are good reasons to expand this teleological interpretation to the reproduction under copyright (see below 3.4.1.1.3).

the phonograms and “copies thereof”.⁴⁶⁶ The notion of “copy” is not defined and must therefore be interpreted in view of the legislative context and purpose of the Rental Directive, which according to the CJEU is to fight “piracy”, i.e., the production and distribution of “counterfeit copies of phonograms”.⁴⁶⁷ It is then ruled that a “copy of a phonogram” can be “an article which reproduces **all or a substantial part** of the sounds fixed in a phonogram”, such reproductions being the only ones which by their nature are “intended to **replace** lawful copies of that phonogram”.⁴⁶⁸ A phonogram containing a sound sample of another phonogram cannot be considered a “copy”, since it does not reproduce all or a substantial part of that phonogram.

Broadcasters

Broadcasting organisations have **economic rights** over their broadcast, i.e., the rights of **fixation, reproduction, rebroadcasting** and (limited) **communication** to the public, **distribution** and **making available** to the public⁴⁶⁹.

3.3.4.1.3. AI-related challenge: Uncertain infringement of reproduction right to protected AI-input in the training of algorithms

The use of protected cultural content for training AI-algorithms relies on complex technical processes, in which uncountable copies are made. If every single copy is qualified as a “reproduction”, then the AI-developer or -user must clear the right or rely on an exemption to avoid infringing the exclusive rights. The interpretation of the reproduction right seems to be evolving, however, at least for the related rights.

Copyright relevance of data processing

In the offline world, the activities of reading books, listening to songs, or watching films in the private sphere do not fall under the material scope of copyright law. Similarly, the reliance on the experience or inspiration gathered through these activities to generate creative content is also outside of the scope of this right.

Arguably, from a functional point of view, the processing of “data” that includes protected subject matter as part of an AI model's training process could be seen as a functional equivalent of “reading”, “seeing”, “listening” and “learning”, which should then remain outside the scope of copyright. The scale, the impact and the purpose of such data processing may however distinguish these operations from the slow and onerous reading and learning by humans.

The technical reading of the **reproduction** right, *de lege lata* – at least as far as copyright is concerned –, is confirmed in the CJEU’s decisions on the “partial” or “temporary” reproductions and on the exception for temporary acts of reproduction under Article 5(1) InfoSoc Directive. Under that reading of the reproduction right, to the extent that entire protected works are copied or that there is a substantial similarity between the protected

⁴⁶⁶ Rental Directive, art. 9(1)(b).

⁴⁶⁷ CJEU 29 July 2019, Pelham GmbH and Others / Ralf Hütter and Florian Schneider-Esleben, C-476/17, EU:C:2019:624, paragraph 45.

⁴⁶⁸ CJEU 29 July 2019, Pelham GmbH and Others / Ralf Hütter and Florian Schneider-Esleben, C-476/17, EU:C:2019:624, paragraph 46.

⁴⁶⁹ See Art. 7, 8 and 9 Rental Dir. And Art. 2(e), 3nr. 2(d) InfoSoc dir.

works and the training data, used as AI input, a “reproduction” under the copyright sense can be considered as taking place.⁴⁷⁰

This interpretation based on the decade old case law of the CJEU under *Infopaq* and *Premier League*, could however have been seriously challenged by more recent decisions of the CJEU concerning the related rights (in particular *Pelham*), in which the Courts adopts a more purposive interpretation of the economic rights. The “reproduction” might thus require that the intermediate process of mechanically copying a work results in a reflection of the work in the output of the process.

Where a recording is thus included in a dataset with the purpose of training an AI system, it is copied as part of a technical process in order to extract certain information, but it does not necessarily lead to an output showing a “recognisable” portion of the works, performances or phonograms. In that reasoning, it could be argued that such copies are not “reproductions” and consequently do not require the prior authorisation of the rightholders.

The DSM dir. confirms that text and data mining operations are not under the rightholders’ control, even if the training sets contain copyright protected material. Such mining falls outside copyright either because the reproduction right can be interpreted to not cover every technical copy or because existing exceptions apply (such as the exceptions for temporary acts of reproduction⁴⁷¹) or because the newly created **TDM exceptions** exempt these reproductions (see Section 3.3.6). From the research, no evidence has been found that AI developers acquire licences for the use of datasets (containing protected subject matter).

Where the protected content is collected for the purpose of training an AI solution by the AI developer, the entire process is managed by one organisation and arguably the right of **communication to the public** is not applicable. During the expert interviews, one international company raised the issue of making datasets available to other organisations for training their AI solutions. To the extent that such datasets contain labelled data (representing the protected works, performances and recordings), it was uncertain whether such accessibility could be considered an act of communication to the public or making available to the public under copyright and related rights.

Relevance of other IP rights

The use for TDM and training of subject-matter protected under other IP rights might require an equivalent interpretation of those other IP rights so as to ensure that the same operations are treated similarly under other exclusive rights.

Design law could be relevant when it comes to the training of algorithms used afterwards for the production of visual art output. The TDM exceptions in the DSM dir. are not extended to design law, this point could probably be analysed in the pending review of design law, taking into account that design law, contrary to copyright, does not expressly include a right to prohibit reproductions (see the definition of the use of a design under EU design rules).

⁴⁷⁰ In theory the defendant could still establish an “independent creation” for escape the finding of an infringement. This will however not be the case where protected creations are used as AI input, where the AI developer or user is using exactly pre-existing work to extract certain information (such as features or patterns) from the data set, consisting of pre-existing works.

⁴⁷¹ Art. 5(1) InfoSoc Dir.

3.3.4.2. Database rights

Where the training data are contained in a protected database, it should be verified whether, during the process of training the AI-solution, acts of extraction and/or reutilisation of the contents of the database are performed.

3.3.4.2.1. *Extraction and re-utilisation*

The maker of a database holds exclusive rights protecting the exploitation of the database. Unlike authors and performers, the maker of the database does not have any moral rights but only the economic rights of extraction and re-utilization. These rights are given a broad interpretation.⁴⁷²

The purpose of the sui generis database right is to protect the investment in the database, in particular in the obtaining, verification or presentation of the contents of a database.⁴⁷³ So far, the CJEU has interpreted "those concepts of 'extraction' and 're-utilisation' [...] as referring to any act of appropriating and making available to the public, without the consent of the maker of the database, the results of his or her **investment**, thus depriving him or her of revenue which should have enabled him or her to redeem the cost of that investment".⁴⁷⁴

The Database Directive defines the "**extraction**" as the "permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form" (art. 7(2)(a) DBD).

Any act of "appropriation" of the whole or a substantial part of the contents of the database is protected under the extraction right.⁴⁷⁵ The objective of the extraction right is "to guarantee the person who has taken the initiative and assumed the risk of making a substantial investment in terms of human, technical and/or financial resources in the obtaining, verification or presentation of the contents of a database a return on his investment by protecting him against the unauthorised appropriation of the results of that investment by acts which involve in particular the reconstitution by a user or a competitor of that database or a substantial part of it at a fraction of the cost needed to design it independently".⁴⁷⁶ The transferred data may be arranged differently than in the protected database, the notion of extraction covers these "adaptations" of the database as well.⁴⁷⁷ The purpose of the transfer of the content is irrelevant: this might be for competitive or non-competitive, commercial or non-commercial purposes.⁴⁷⁸

The temporary or permanent transfer of the database are protected: a permanent transfer is found "when those materials are stored in a permanent manner on a medium other than

⁴⁷² CJEU 19 December 2013, *Innoweb BV tegen Wegener ICT / Wegener ICT Media BV and Wegener Mediaventions BV*, C-202/12, ECLI:EU:C:2013:850, point 34; CJEU 9 October 2008, *Directmedia Publishing GmbH v Albert-Ludwigs-Universität Freiburg*, C-304/07, ECLI:EU:C:2008:552, point 32.

⁴⁷³ European Parliament and Council, Directive 96/9/EC of 11 March 1996 on the legal protection of databases, rec. 40, (O.J. L 77, 27.03.1996, p. 20).

⁴⁷⁴ CJEU 3 June 2021, *CV-Online Latvia / Melons SIA*, C-762/19, ECLI:EU:C:2021:434, paragraph 51.

⁴⁷⁵ CJEU 9 October 2008, *Directmedia Publishing GmbH / Albert-Ludwigs-Universität Freiburg*, C-304/07, ECLI:EU:C:2008:552, point 34.

⁴⁷⁶ CJEU 9 October 2008, *Directmedia Publishing GmbH / Albert-Ludwigs-Universität Freiburg*, C-304/07, ECLI:EU:C:2008:552, point 33 and references to earlier case law.

⁴⁷⁷ CJEU 9 October 2008, *Directmedia Publishing GmbH / Albert-Ludwigs-Universität Freiburg*, C-304/07, ECLI:EU:C:2008:552, point 39; CJEU 5 March 2009, *Apis-Hristovich EOOD / Lakorda AD.*, C-545/07, ECLI:EU:C:2009:132, point 47.

⁴⁷⁸ CJEU 9 October 2008, *Directmedia Publishing GmbH / Albert-Ludwigs-Universität Freiburg*, C-304/07, ECLI:EU:C:2008:552, point 46.

the original medium", whereas the transfer is temporary "if the materials are stored for a limited period on another medium, such as the operating memory of a computer".⁴⁷⁹ Both the permanent and temporary transfer may be considered an "extraction" but the classification might have an incidence on the gravity of the infringement.

The "**re-utilization**" is defined as "any form of making available to the public all or a substantial part of the contents of a database by the distribution of copies, by renting, by on-line or other forms of transmission. The first sale of a copy of a database within the Community by the rightholder or with his consent shall exhaust the right to control resale of that copy within the Community". It covers every act of making available to the public the results of the investment, thus depriving the maker of the database of revenue which should have enabled them to redeem the cost of the investment.⁴⁸⁰ The nature and form of the process used are not relevant. The operator of a meta-search engine creates a risk for the maker of a database, who may lose income from advertising if their website (giving access to their database) loses traffic. As a consequence, the maker of the database may be deprived of revenue which should have enabled them to redeem the cost of the investment in setting up and operating the database.⁴⁸¹

Nevertheless, the CJEU decided in *CV Online Latvia* that the transfer of the substantial contents of a database, as part of the indexing process of a specialised search engine, and making the data available to the public without the consent of the database maker is an extraction and re-utilisation of the database. For the Court, even if the activities fall under the legal notions of extraction/re-utilisation one should check whether they have the effect of "depriving that person of income intended to enable him or her to redeem the cost of that investment" before concluding to an infringement. Importantly, it is necessary to examine whether the acts under scrutiny "are such as to affect the investment of the maker of the database which has been transferred to another medium and has been made available to the public".⁴⁸² The Court requires that a "fair balance" be struck "between, on the one hand, the legitimate interest of the makers of databases in being able to redeem their substantial investment and, on the other hand, that of users and competitors of those makers in having access to the information contained in those databases and the possibility of creating innovative products based on that".⁴⁸³ The offering of new services (such as the aggregation of specialised search engines) may create value for competitors and users of such search services but, ultimately, the "main criterion for balancing the legitimate interests at stake must be the **potential risk to the substantial investment** of the maker of the database concerned, namely the **risk that that investment may not be redeemed**"⁴⁸⁴ (emphasis added). This CJEU reading of the condition for infringing the database rights puts forward the justification and the purpose of the protection by the sui generis right. This development in the interpretation of the rights of extraction and re-utilization is worth to be flagged in relation to the type of use made by AI tools: so long as they do not lead to the development of competing databases or substitute products that

⁴⁷⁹ CJEU 9 October 2008, *Directmedia Publishing GmbH / Albert-Ludwigs-Universität Freiburg*, C-304/07, ECLI:EU:C:2008:552, point 44.

⁴⁸⁰ CJEU 19 December 2013, *Innoweb BV tegen Wegener ICT / Wegener ICT Media BV and Wegener Mediaventions BV*, C-202/12, ECLI:EU:C:2013:850, point 37.

⁴⁸¹ CJEU 19 December 2013, *Innoweb BV tegen Wegener ICT / Wegener ICT Media BV and Wegener Mediaventions BV*, C-202/12, ECLI:EU:C:2013:850, point 41.

⁴⁸² CJEU 3 June 2021, *CV-Online Latvia / Melons SIA*, C-762/19, ECLI:EU:C:2021:434, point 38.

⁴⁸³ CJEU 3 June 2021, *CV-Online Latvia / Melons SIA*, C-762/19, ECLI:EU:C:2021:434, point 41.

⁴⁸⁴ CJEU 3 June 2021, *CV-Online Latvia / Melons SIA*, C-762/19, ECLI:EU:C:2021:434, point 44.

might impact the investment of the database maker the use would remain outside the scope of the database right.

The rightholder can only prevent the extraction or re-utilization of the **whole** or of a (quantitatively or qualitatively) **substantial part** of the content of the database⁴⁸⁵. The assessment of whether the extracted or re-used part was substantial must refer "to the investment in the creation of the database and the prejudice caused to that investment by the act of extracting or re-utilising that part".⁴⁸⁶ A 'substantial part, evaluated quantitatively', of the contents of a database refers to the "volume of data extracted from the database and/or re-utilised, and must be assessed in relation to the volume of the contents of the whole of that database". A link is made between the quantitative investment required to create the re-used or extracted part of the database: if the resources invested in the creation of that part are substantial, then that portion constitutes a substantial part of the content of the database and its re-utilisation or extraction is protected. The investment in the obtaining, verification or presentation of the contents of the database will also guide the assessment of the qualitatively "substantial" part: the extracted or reutilised part may represent a qualitatively substantial investment, without being a quantitatively substantial part of the database.⁴⁸⁷ The intrinsic value of the data is not relevant for assessing whether a substantial part of the contents of the database is concerned.⁴⁸⁸

The extraction or re-utilization of **insubstantial** parts of the database is not protected, unless they are done in a repeated and systematic manner, which imply acts that conflict with a normal exploitation of that database or which unreasonably prejudice the legitimate interests of the maker of the database.⁴⁸⁹ The repeated and systematic acts may not lead, through their cumulative effect, to the reconstitution of the database as a whole or of a substantial part of it, regardless of whether the purpose of the user was to create another database.⁴⁹⁰ Such acts would seriously prejudice the investment of the maker of the database.

The **consultation** of a database is not protected under the exclusive rights (although contractual conditions may apply)⁴⁹¹, except "when on-screen display of the contents of that database necessitates the permanent or temporary transfer of all or a substantial part of such contents to another medium".⁴⁹²

That the content of two databases present the same physical and technical characteristics might be evidence of an extraction, unless this coincidence can be explained by factors other than a transfer between the two databases concerned (such as the use of a common source).⁴⁹³

⁴⁸⁵ CJEU 9 Nov. 2004, *The British Horseracing Board*, C-203/02, ECLI:EU:C:2004, point 68 et seq.; Study in support of the evaluation of Directive 96/9/EC on the legal protection of databases, Annex 1, In-depth analysis of the Database Directive, article by article, <https://op.europa.eu/s/oUKD>, (accessed in July 2021).

⁴⁸⁶ CJEU 9 Nov. 2004, *The British Horseracing Board*, C-203/02, ECLI:EU:C:2004, point 69.

⁴⁸⁷ CJEU 9 Nov. 2004, *The British Horseracing Board*, C-203/02, ECLI:EU:C:2004, point 71.

⁴⁸⁸ *Idem*, point 78.

⁴⁸⁹ European Parliament and Council, Directive 96/9/EC of 11 March 1996 on the legal protection of databases, Art. 7(5), (O.J. L 77, 27.03.1996, p. 20).

⁴⁹⁰ CJEU 9 Nov. 2004, *The British Horseracing Board*, C-203/02, ECLI:EU:C:2004, point 87-89.

⁴⁹¹ *Idem*, point 54; CJEU 9 October 2008, *Directmedia Publishing GmbH / Albert-Ludwigs-Universität Freiburg*, C-304/07, ECLI:EU:C:2008:552, point 51.

⁴⁹² CJEU 9 October 2008, *Directmedia Publishing GmbH / Albert-Ludwigs-Universität Freiburg*, C-304/07, ECLI:EU:C:2008:552, point 53.

⁴⁹³ CJEU 5 March 2009, *Apis-Hristovich EOOD / Lakorda AD.*, C-545/07, ECLI:EU:C:2009:132, point 51.

3.3.4.2.2. *AI-related challenge: acts of extraction to generate AI training data sets.*

The exclusive rights protecting the use of the content of databases is mainly relevant for the **input phase**, i.e. when the vast quantities of data (possibly collected in a protected database) are used to train an AI algorithm.

The collecting of data from various sources, e.g. by web scraping or by using a protected access to a database, in another environment, even on a temporary basis, may be considered a protected act of extraction (considering the large interpretation of the economic rights that the CJEU has adopted so far). This might be the case, even if the purpose of the different collections is different and if the second collection is not a substitute for the first database.

For instance, a well-curated music catalogue makes it easy for the users to search and find music tracks, thanks to the investments in searchability of its contents and the presentation of the search results. The purpose of this database is to facilitate the access to and the use of the music catalogue. The content of this catalogue may be scraped for competing purposes, e.g. to build a competing music catalogue (enriched with content from other source), or for other purposes (e.g. mining the meta-data of the tracks to spot historic evolutions or for training an AI-algorithm for creating AI-music). The transfer of the content of a protected database to a training set, used for developing a non-commercial AI solution (e.g. automated classification of images in the catalogue of a museum or public library) may also be a protected act, despite the lack of a commercial purpose.

It is uncertain whether the technical operations to optimise the data for training the AI algorithm (conversion to a different format, enriching the data with meta-data) amounts to a distinct protected act, considering that the Database Directive does not provide for an adaptation right as such.

At the same time, the curation of the content in the database may amount to a sufficient investment for the resulting dataset to be protected under the sui generis database right.

Importantly, finding an act of extraction or re-utilisation of the content of the database is independent of any infringement of the economic rights protecting the work, performances or recordings that may be included as protected elements in the database. The database rights, copyright and related rights apply in a cumulative way.

Moreover, the same cultural content (protected works, performances, recordings) may be included in several databases (e.g. music catalogues commonly carry the same albums). The mere presence of such cultural content in a training dataset cannot suffice to conclude that this training set infringes an earlier database in which the same works are included. The database makers may indeed make independent efforts to obtain the same content for different databases, without violating each other's rights and earning their own sui generis rights due to their individual efforts. Where the same cultural content, contained in several databases, is used for training purposes, the user may face complex issues of management of rights (also for the exercise of the opt-out of the TDM exception – see below). They will have to obtain the authorisation to use the individual works, performance and/or recordings in the database. If in addition the cultural content is also part of one or more protected databases (which may have overlapping content if they share the same sources), of which at least a "substantial part" is extracted or reutilised, then the authorization of the makers of those databases is also required.

3.3.5. Moral rights – Which actions in an AI process are relevant for the moral rights under copyright and the performers' rights?

The author and performer enjoy certain moral rights on their works and performances. The use of protected works and performances as AI input to develop an AI solution is mostly hidden, as part of a technical process. Nevertheless, considering the impressive AI applications that are being developed and the impact they may have on the human author or performer, who may have a strong personal connection with their work or performance, the question arises whether the moral rights have a role to play when AI solutions develop in the cultural sector.

3.3.5.1. Authors and performers' moral rights

3.3.5.1.1. Copyright

In addition to their economic rights, authors enjoy moral rights upon their work, which allow their owner to defend their personality as expressed within the work. As the result of the divergences existing between civil law and common law traditions concerning the purpose of copyright law, these rights benefit only for very limited harmonisation both at the international and European level.

At the international level first, **art. 6bis of the Berne Convention** recognised authors a right to claim attribution and a right of integrity:

Independently of the author's economic rights, and even after the transfer of the said rights, the author shall have the right to claim authorship of the work and to object to any distortion, mutilation or other modification of, or other derogatory action in relation to, the said work, which would be prejudicial to his honour or reputation.

On the European scale, it should first be noted that moral rights are **not part of the *acquis communautaire***.⁴⁹⁴ Recital 19 of the InfoSoc Directive is in this regard explicit by stating that:

The moral rights of rightholders should be exercised according to the legislation of the Member States and the provisions of the Berne Convention for the Protection of Literary and Artistic Works, of the WIPO Copyright Treaty and of the WIPO Performances and Phonograms Treaty. Such moral rights remain outside the scope of this Directive.

However, the Court of Justice started to approach this area of copyright law. Hence in the *Phil Collins* case, the Court spoke of the aims of both moral and economic rights under copyright and neighbouring rights, but remained very general.⁴⁹⁵

⁴⁹⁴ European Parliament and the Council, Directive 2001/29/EC of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, rec. 19, (O.J. L 167, 22.6.2001, p. 10).

⁴⁹⁵ CJEU 20 October 1993, *Phil Collins / Imtrat Handelsgesellschaft mbH*, C-92/92 and C-326/92, ECLI:EU:C:1993:847, [1993] ECR, I-5145, paragraph 20; See also CJEU 20 January 1981, *Musik-Vertrieb Membran GmbH and K-tel International / GEMA*, 55/80 and 57/80, ECLI:EU:C:1981:10, [1981] ECR 147, paragraphs 12-13 (discussed in § 2[2][b][i] *infra*) (also noting "... that copyright comprises moral rights of the kind indicated by the French Government").

The absence of any EU harmonisation results in **divergences between Member States'** legal orders. Most Continental EU countries recognise three moral rights of authors: the right of disclosure, the right of attribution, and the right of integrity.

- According to the **right of disclosure**, the actual, flesh-and-blood author is alone empowered to decide when their work is finished and how, when, and under which circumstances it should be made accessible.
- The **right of attribution** as recognised in the Copyright Acts provides that an author has the right to have authorship of their works attributed to them, to prevent others from attributing authorship of their works falsely to themselves or others, and to refuse to have their authorship of any given work recognised. Thus, an author may choose to remain anonymous or to hide their identity under a pseudonym but may later change their mind and identify as the author of the work at issue.
- The right of integrity protects the **integrity of a work** against any alteration or other impairment of the work that might prejudice the honour or reputation of the author.

Some EU Member States recognize a fourth moral right **to retract** a work in the event of changes in the author's opinion, known in France as the *droit de repentir*.

Above this enumeration, divergences exist concerning the conditions of exercise of these rights. Two approaches can in this regard be distinguished: a **subjective** and an **objective** one. Following the subjective approach in some Member states, such as France or Belgium, it is – a least in theory – left to the author to determine if a given use of its work is constitutive of an alternation incompatible with the integrity of that work.⁴⁹⁶ The objective approach, on the contrary, subordinate the characterisation of the violation to the demonstration of an objective prejudice for the honour or reputation of its author. This approach is in line with the minimal requirement of the Bern convention, which allows the author to “object to any distortion, mutilation or other modification of, or other derogatory action in relation to, the said work, *which would be prejudicial to his honor or reputation* »⁴⁹⁷. In Europe, some Member states such as Spain⁴⁹⁸ or Germany⁴⁹⁹ are following this approach and by hence requiring the demonstration of an objective damage for the authors.

Lastly, divergences also exist between the Member States concerning the possibility for the rightholder to waive their moral rights. In France, this possibility is barred by art. L. 121-1 Code of Intellectual Property, which provides that moral rights are “perpetual, inalienable and imprescriptible.” Other civil law countries, such as Germany, are adopting a more nuanced approach and permit to waive the moral rights in certain circumstances (for instance to allow the publisher to alter the work within reasonable limits⁵⁰⁰). On the other side, the common law countries fully authorise the waiving of moral rights.⁵⁰¹

⁴⁹⁶ Under Belgian law, the author can even object to “non-material” modifications” if such the intervention affects the “spirit” of the work: Cass. 8 May 2008, AM 2009, 1-2, 102, note F. GOTEZEN; See in France Cour de Cassation, 17 december 1991, 89-22.035.

⁴⁹⁷ Art. 6bis BC, emphasis added.

⁴⁹⁸ See Art. 14(iv) Spain Intellectual Property Act: “Exigir el respeto a la integridad de la obra e impedir cualquier deformación, modificación, alteración o atentado contra ella que suponga perjuicio a sus legítimos intereses o menoscabo a su reputación”.

⁴⁹⁹ §14 German copyright Act.

⁵⁰⁰ § 39(2) German copyright Act.

⁵⁰¹ GOLDSTEIN and HUGENHOLTZ, *supra* note 89 at 346.

3.3.5.1.2. Performers' moral rights and personality rights

Performers also enjoy moral rights. These rights are usually more limited and less robust. These moral rights are partially harmonised at the international level, but not at the European level.⁵⁰² **Art. 5(1) WPPT** hence provides that performers also enjoy moral rights allowing them "to claim to be identified as the performer of his performances, except where omission is dictated by the manner of the use of the performance, and to object to any distortion, mutilation or other modification of his performances that would be prejudicial to his reputation." It could be noted that concerning Performers moral rights, Art. 5 Beijing Treaty specify that these rights are recognised "taking due account of the nature of audiovisual fixations". These rights of attribution and integrity can be reinforced by the member states.

For completeness, performers, as any natural persons, may also enjoy **personality rights** under national law, which might allow them to act against certain uses of their portrait or other aspects of their appearance. These rights, not harmonised at the European level, are eventually outside the scope of performers' moral right since they are recognised to any natural persons irrespective of the existence of a performance. However, the relationship between these personality rights and the prerogative arising from the moral rights recognised over the performance are often overlapping, hence creating a further degree of complexity when it comes to performances in comparison to copyright moral rights.

3.3.5.2. AI-related challenges: the relevance of the right of integrity for AI-training data

In the AI context, moral rights, and particularly the right of integrity, could play a role when works or performances are utilised as an AI solution input since they could be used by right owners **to oppose the use of their work or performance as AI inputs**.

Authors and performers enjoy the protection of the economic rights when their works or performances can be recognised in the AI-output. The integrity right could add a further layer of protection, even if the work or performance is merely used as input data but cannot be identified in the AI output or if the economic rights have been transferred, have expired, or are subject to an exception. As noted in one study "the processing by an AI system can be quite different from mere digitisation. For instance, the authors of a novel may not want their works to be processed by an AI system, even if the user does not imply communication of the work per se, where such processing could be perceived by them as a derogatory treatment of the work".⁵⁰³

The results gathered in the frame of the Delphi study are in this regard interesting. Experts participating in the survey were asked their opinion on whether the author of creative works could make use of moral rights to oppose the use of their work to train AI. As it can be seen in Figure 35, the majority of them (59%) believe that every use of the creative work to train AI against the will of the author infringes the moral right of integrity. However, 7 experts argued that this depends on the purpose of the AI processing. In their opinion, using creative works just to train AI does not necessarily infringe the moral integrity, but if this AI is used to generate AI-made creative output based on the trained dataset (e.g. mimics of style) - especially if it affects the honour and/or reputation of the author - it can

⁵⁰² European Parliament and the Council, Directive 2001/29/EC of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, Rec. 19, (O.J. L 167, 22.6.2001, p. 10).

⁵⁰³ J. DREXL et al., "Artificial Intelligence and Intellectual Property Law-Position Statement of the Max Planck Institute for Innovation and Competition of 9 April 2021 on the Current Debate", *Max Planck Institute for Innovation Competition Research Paper 2021 No. 21-10*, p. 12.

be subject to moral rights of the authors whose works are used to train the AI. Therefore, they conclude that the possibility to oppose the processing of creative works must be examined on a case by case basis to determine if the conditions for a moral infringement are in place.

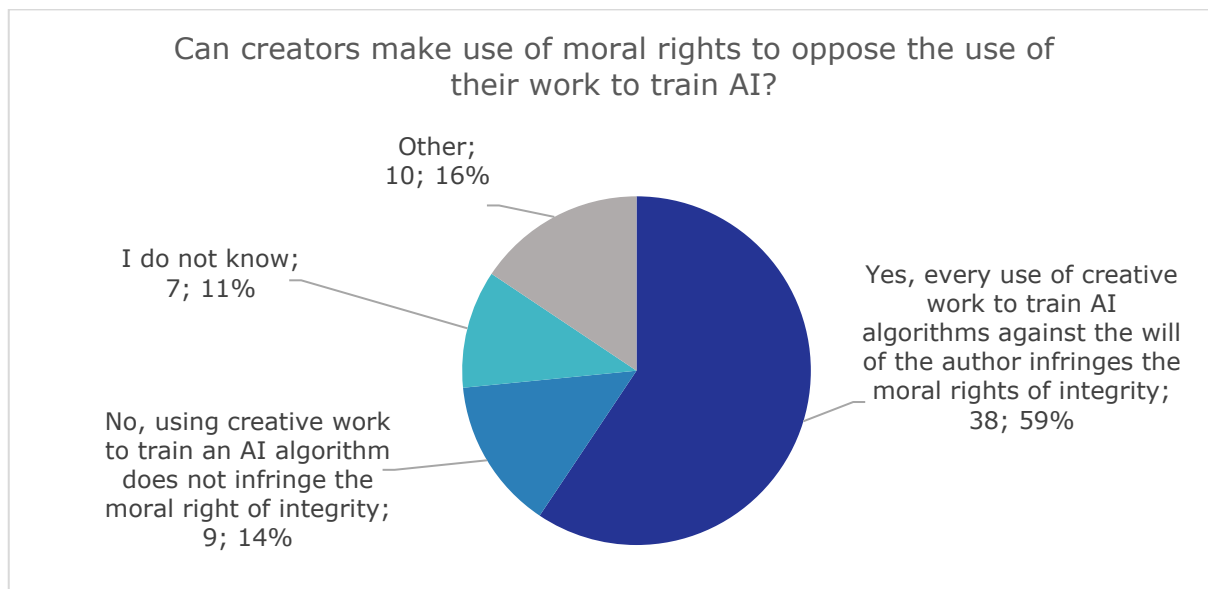


Figure 35: Experts' opinion on the use of moral rights to oppose the use of works to train AI

Source: Technopolis Group Survey

The experts' divergence of opinions might first be analysed as the result of the legal order in which they place themselves to answer this question. Indeed, the divergence between the national legal orders explained in the previous paragraph will automatically lead to different results. Hence, in country offering a strong protection of moral rights by relying upon a subjective approach and without providing further limitations, the rightholders could well be in a position to prohibit the use of its work as AI input, even if the latter are not recognisable at the output stage. On the contrary, in countries relying on an objective approach, a prejudice to the *honour or reputation of the author seems to be less likely if the work is only use in the preparatory phase*.

The **cultural sector** with which the experts are the most familiar might also have been an important factor. Indeed, the relevance of the moral rights of attribution and integrity might vary depending on the field of creation. As pointed out by one interviewee, for video games, the role played by moral rights is for instance more limited. As many video games integrate several elements that derive from previous works, it is quite difficult to identify the many authors of a single production, and this reduces the effect that moral rights, and in particular the right to claim authorship, can have.⁵⁰⁴

⁵⁰⁴ See also J. STEIN, *Authorship and Moral Rights in Video Games*, Glasgow, Press Start, 2015 and more generally J. DREXL et al., *Artificial Intelligence and Intellectual Property Law-Position Statement of the Max Planck Institute for Innovation and Competition of 9 April 2021 on the Current Debate*, *Max Planck Institute for Innovation Competition Research Paper 2021 No. 21-10*, p. 12: "The extent to which this issue may arise in practice would depend on how many intermediate steps would be necessary for the ML model training."

Depending on the AI developments in the cultural sector, the possible application of the moral rights to the training of AI solutions will be addressed (see section 3.4.1.3).

3.3.6. Exceptions (text and data mining) – To which extent do the TDM exceptions apply to AI relevant operations?

The use of protected content as AI-training data may involve certain protected acts, which require the rightholders' prior consent – unless they are exempted under one of the copyright exceptions. The newly introduced exceptions for "text and data mining" (TDM) may relieve the developers and users of AI solutions in the cultural sector of this burden.

3.3.6.1. Presentation of the TDM exceptions

The DSM Directive has introduced a **definition** of TDM in the following terms: "any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations" (Article 2.2 DSM dir.). The definition is sufficiently broad to embrace the current TDM application panorama.⁵⁰⁵

TDM activities **may involve copyright-relevant acts** if a work, performance, fixation or database qualify for protection. Depending on the technique used, TDM may involve: 1) the reproduction of copyright-protected content; 2) the extraction of a substantial part of the database⁵⁰⁶; 3) the reproduction and adaptation of a computer program.⁵⁰⁷

To solve the legal uncertainty concerning TDM activities and improve the attractiveness of the system, compared to systems offering a favourable legal framework (e.g., Japan or the US), the European legislator introduced two *ad hoc* TDM exceptions in the DSM Directive.

3.3.6.1.1. The TDM exception for scientific research

Article 3 DSM Directive contains the TDM exception for the purpose of scientific research. Such an exception was included in the 2016 Commission's proposal for the Directive on copyright in the DSM.⁵⁰⁸

Four components of this TDM exception can be distinguished: 1) the rights affected; 2) the beneficiaries; 3) the scope; 4) the pre-existing condition.

As to the **first element**, Article 3 provides for an exception to:

⁵⁰⁵ R. DUCATO AND A. STROWEL, "Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to Machine Legibility", *CRIDES Working Paper Series*, 2018.

⁵⁰⁶ See in this respect art. 1(6) of the Open Data Directive N° 2019/1024, which provides that "The right for the maker of a database provided for in Article 7(1) of Directive 96/9/EC shall not be exercised by public sector bodies in order to prevent the re-use of documents or to restrict re-use beyond the limits set by this Directive".

⁵⁰⁷ The definition is sufficiently broad to embrace the current TDM application panorama. For a technical definition of TDM, see Marti A. Hearst, 'Text Data Mining' in Ruslan Mitkov (ed), *The Oxford Handbook of Computational Linguistics* (Oxford University Press 2003). Specifically on text mining, Ronen Feldman and James Sanger, *The text mining handbook: advanced approaches in analyzing unstructured data* (Cambridge university press 2007). For an extensive analysis of the definition of TDM, see Triaille, de Meeûs d'Argenteuil and de Francquen, *Study on the legal framework of text and data mining (TDM)*.

⁵⁰⁸ Proposal of COM/2016/0593 final for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market - 2016/0280 (COD).

- the right of reproduction of whole or part of databases protected by copyright (Article 5(a) Database Directive);
- the right of extraction of whole or a substantial part of databases covered by the sui generis right (Article 7(1) Database Directive);
- the right of reproduction in whole or part of works, fixations of performances, phonograms, fixations of broadcasts, the original and copies of films (Article 2 InfoSoc Directive);
- the right of reproduction of on-demand press publications⁵⁰⁹ (the new press publisher rights established by Article 15(1) CDMSD).⁵¹⁰

With regard to its beneficiaries, the exception is granted to research organisations and cultural heritage institutions only. **Research organisations** are defined as “a university, including its libraries, a research institute or any other entity, the primary goal of which is to conduct scientific research or to carry out educational activities involving also the conduct of scientific research: (a) on a not-for-profit basis or by reinvesting all the profits in its scientific research; or (b) pursuant to a public interest mission recognised by a Member State; in such a way that the access to the results generated by such scientific research cannot be enjoyed on a preferential basis by an undertaking that exercises a decisive influence upon such organisation” (Article 2(1) DSM dir.). While the notion of **cultural heritage institution** refers to “a publicly accessible library or museum, an archive or a film or audio heritage institution”.⁵¹¹

Those who can benefit from the TDM exception are essentially institutions that provide a cultural or public service in the interest of society on a non-for-profit basis.⁵¹² The reality of research is however more complex. It is not unusual for a public university to be involved in a consortium with industry and SMEs (it is actually encouraged by many research programs supported by the European Commission). The issue is partially addressed in the Recitals of the DSM Dir. where it is stated that research organisations and cultural heritage institutions “should be able to rely on their private partners for carrying out text and data

⁵⁰⁹ For the purpose of the DSM dir., press publication means “a collection composed mainly of literary works of a journalistic nature, but which can also include other works or other subject matter, and which: (a) constitutes an individual item within a periodical or regularly updated publication under a single title, such as a newspaper or a general or special interest magazine; (b) has the purpose of providing the general public with information related to news or other topics; and (c) is published in any media under the initiative, editorial responsibility and control of a service provider” (Article 2(4) DSM dir.). Recital 56 adds that the new right exists for “journalistic publications, published in any media, including on paper, in the context of an economic activity that constitutes a provision of services under Union law”. For instance, the notion includes “daily newspapers, weekly or monthly magazines of general or special interest, including subscription-based magazines, and news websites” (Recital 56). Press publications include the article, as literary work, but also other subject matter accompanying the article, such as photo and videos. The definition does not extend to scientific journals and blogs “that provide information as part of an activity that is not carried out under the initiative, editorial responsibility and control of a service provider, such as a news publisher” (Recital 56 and Article 2(4) *in fine* DSM dir.). The TDM exception under Article 3 arguably does not impact the new press publisher right insofar the right does not exist in publications for scientific or academic purposes.

⁵¹⁰ Article 15(1) DSM dir. confers to press publishers not only the exclusive right recognised at Article 2 of the InfoSoc Directive, but also the right at Article 3(2) InfoSoc Directive. However, Article 3 DSM dir. refers to the acts of reproduction and extraction only, and the TDM exception does not extend to the right of making available press publications to the public. The drafting of Article 3 referring to Article 15(1) confirms that TDM is limited to the analysis of text and data in order to generate something different from the original corpus subject to mining.

⁵¹¹ Article 2(3) DSM dir.

⁵¹² Hospitals carrying out research may be included in such a definition, as expressly mentioned in Recital 12 DSM dir.

mining, including by using their technological tools" (Recital 11). Therefore, in the context of public-private partnerships, the DSM dir. leaves some room for the private actors to benefit from the exception at Article 3 DSM dir., if required by the needs of the project. At the same time, this implies that such condition will not extend beyond the scope of the collaborative project or after its conclusion.

The third element of the TDM exception concerns **its scope**: TDM activities shall be directed to research purposes only. By "scientific research" it is meant research both in natural and human sciences.⁵¹³ The content of such a notion appears to be quite vague, especially if compared to the corresponding definition of scientific research in the General Data Protection Regulation (GDPR).⁵¹⁴ Statistical studies or research for technological development are very probably covered, although such areas do not fall *stricto sensu* under the umbrella of human or natural sciences. For instance, statistical purposes are conceptually distinct from scientific research in the GDPR.⁵¹⁵ However, despite the formulation of Recital 12, statistics must be seen as scientific research for the purpose of Article 3 CDMSD, considering that several TDM applications are aimed at obtaining statistical results.

Fourthly, Article 3 requires from the TDM beneficiary to have "**lawful access**" to the work or other protected subject matter to be mined, i.e. the "pre-existing" condition. In other words, the exception only works under the condition that research organisations and cultural heritage institutions have already lawful access to the resource. By lawful access the DSM dir. means "access to content based on an open access policy or through contractual arrangements between rightholders and research organisations or cultural heritage institutions, such as subscriptions, or through other lawful means [...] Lawful access should also cover access to content that is freely available online".⁵¹⁶ Therefore, if the content is protected and the user does not enjoy a right to access and use it, TDM cannot be performed for research purposes under the exception. This could be seen as an important limitation to the exception that will be addressed below.

If all the conditions listed in Article 3(1) are met, the research organisation or the cultural heritage **institution is allowed to perform the TDM activity** and to retain the copies of the works and subject matter made according to the exception.⁵¹⁷ This is a relevant addition that recognises the importance of maintaining the copies of the protected work and of the data to fulfil the scientific rationale, for instance by allowing the peer review and verification of the results.⁵¹⁸ The TDM exception, though, comes with the obligation to store the copies "with an appropriate level of security".

Furthermore, Article 3 DSM dir. contains a provision about "**security measures**" that must be applied to the storage of the copies generated via TDM (Article 3(2) DSM dir.). Additionally rightholders can apply measures "to ensure the security and integrity of the networks and databases where the works or other subject matter are hosted" (Article 3(3) DSM dir.). In particular, such security or integrity measures on the files stored should in no way limit the possibility of applying the TDM tools. The Directive leaves the exact

⁵¹³ Recital 12 DSM dir.

⁵¹⁴ Recital 156 GDPR defines scientific research "in a broad manner including for example technological development and demonstration, fundamental research, applied research and privately funded research. In addition, it should take into account the Union's objective under Article 179(1) TFEU of achieving a European Research Area. Scientific research purposes should also include studies conducted in the public interest in the area of public health".

⁵¹⁵ For the definition of statistical purposes, see Recital 162 GDPR.

⁵¹⁶ Recital 14 DSM dir.

⁵¹⁷ See Article 3(2) DSM dir.

⁵¹⁸ Recital 15 DSM dir.

definition of the measures required from the lawful miner under Article 3(2) and of those applicable by the rightholder (Article 3(3)) to the determination of the parties. Member States should encourage stakeholders to define best practices concerning the application of the above-mentioned measures (i.e., rightholders, research organisations and cultural heritage institutions).⁵¹⁹ In practice, it is likely that those measures will make it more complex or cumbersome to conduct TDM on the corpuses, and it will be very difficult to disentangle the measures objectively justified by the security or integrity of the corpuses and those that go beyond (and hamper TDM).

Finally, the Directive establishes an important principle with the prohibition of contractual provisions overriding the TDM exception for research purposes.⁵²⁰ Rightholders are not entitled to prohibit TDM on contractual grounds. However, they do have the possibility to limit the “lawful access” to their works, performances, recordings or databases. Measures to control access to the content can thus be used to (indirectly) limit the possibility to use the content for TDM.

3.3.6.1.2. General TDM exceptions and limitations

Article 4 DSM dir. contains another provision favourable to TDM activities that was specifically added during the legislative process leading to the adoption of the Directive.⁵²¹ Analogously to Article 3, the provision here commented establishes that the miner must have **lawful access** to the resource as a *sine qua non* condition.⁵²² Apart from this common ground, all the other elements of the exception differ.

Notably, Article 4 is **broader in terms of the plethora of beneficiaries**. There are in fact no limitations or qualifications to comply with: everyone is in principle entitled to the exceptions or limitations that will be implemented by the Member States under Article 4.

The **scope** of the exception is wider as well: not only research purposes are covered, but any TDM activity, whether non-profit or for profit, as long as it falls under the definition of TDM in Article 2(2) DSM dir.

Another difference concerns the **object** covered by the exception here at stake. In addition to the list of rights that can be limited by TDM for research purposes, Article 4 includes: the right to reproduction and the right to adaptation of computer programs.⁵²³ Indeed, TDM activity may also concern software code and, if the legislative intent is to lower the barriers for TDM, then a limitation to the exclusive rights protecting computer programs is welcome.

It should yet be noted that, contrary to the TDM exception for scientific research, Article 4(3) foresees the **possibility for rights-owners to reserve their rights**. Accordingly: “The exception or limitation provided for in paragraph 1 shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means

⁵¹⁹ See, Article 3(4) DSM dir.

⁵²⁰ See, Article 7(1) DSM dir.

⁵²¹ In particular, see the version of the text dated 25 May 2018 (Council of the EU, Interinstitutional File: 2016/0280(COD), doc. 9134/18, available here: <https://www.consilium.europa.eu/media/35373/st09134-en18.pdf>.

⁵²² The language used in Article 4 slightly varies as it refers to “lawfully accessible” works or other subject matter, while Article 3 refers to the works or other subject matter “to which [the beneficiaries, i.e. the research organisations and cultural heritage institutions] have lawful access”, but this does not affect the pre-condition.

⁵²³ European Parliament and the Council, Directive 2009/24/EC of 23 April 2009 on the legal protection of computer programs (codified version), Article 4(1)(a), (O.J. No. L 111/16, 05.05.2009).

in the case of content made publicly available online". As to the manner in which the opt-out can be expressed, a distinction is made between content that is publicly available online and "other cases"⁵²⁴. For online accessible content, only the opt-out "by the use of machine-readable means, including metadata and terms and conditions of a website or a service" should be considered "appropriate". In "other cases", the opt-out can be expressed by "other means", such as contractual agreements or a unilateral declaration. No further specifications (e.g. by way of standards) are provided in the DSM dir.

Furthermore, a rightholder may want to opt-out of the TDM processing for certain purposes (e.g. producing competing cultural creations) but not for other purposes (e.g. training algorithms used for identifying deepfakes). Such granularity is not provided in the DSM dir, but contractual regulation of the authorised/prohibited uses could help to solve this (in favour of the large users who can afford to go through this contractual process).

3.3.6.2. AI-related challenges: the uncertainty concerning the exception's scope and modalities for reserving the rights (opt-out)

Both TDM exceptions are relevant for the use of AI solutions in the cultural sector. Many applications of AI will indeed rely on TDM techniques to train AI algorithms. Some applications may not use protected content as such, e.g. many recommendations systems are based on an analysis of user data and meta-data linked to music or videos in catalogues. Other systems will process the protected works and, in case of recorded versions, performances and fixations, e.g. applications to generate photos or apply filters imitating the style of a well-known artist (e.g. the application Prisma⁵²⁵). Moreover, all training data might be collected in protected databases. It can indeed be expected that access to sufficient quantities of data will be another important limitation to the possibility to perform TDM processes and to train AI solutions, especially for smaller players. The uneven access to big data sets may reinforce the existing inequalities between market players. The newly introduced TDM exceptions may not solve the issue of access to the data but they may be relevant for the use of data for developing these applications in all cultural sectors under examination, in addition to existing exceptions under the InfoSoc Directive (such as the exceptions for temporary acts of reproduction or the exceptions for research).

While the TDM exceptions are not yet transposed in all national laws and have not been tested in practice, the TDM exceptions already appear to raise several questions. When asked during the Delphi survey about their knowledge of TDM exceptions, 70% (44) of the stakeholders answered they were aware of the TDM exceptions either from a theoretical perspective or because they knew a real case, but those exceptions were unknown for 29% (19) of the respondents (see following figure)

⁵²⁴ Recital 18 DSM dir.

⁵²⁵ Prisma Ai, <https://prisma-ai.com/>, (accessed in July 2021): The application is presented more in details in Chapter 3.1 of this report.

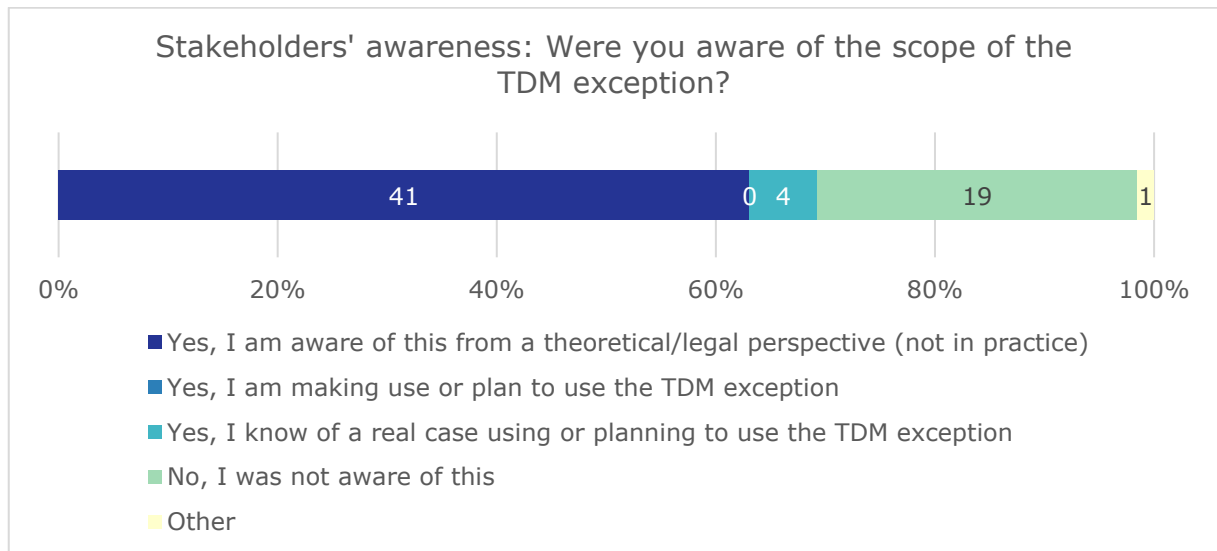


Figure 36: Stakeholders' awareness on the TDM exception

Source: Technopolis Group Survey

On the other side, their opinion on whether the **use of creative works for AI training may lead to some kind of copyright-related obstacle** was quite restrictive: 70% (45 respondents, among which 23 are legal practitioners/academics, 8 artists/rightholders, 14 representatives of CMOs, and 1 producer in the creative sector⁵²⁶) of them indeed stated that the use of creative works for AI training needs to be covered by licences. In this regard, an organisation representing creators and authors for instance asserted that when the AI produces, on the basis of a copyright-protected input, a recognisable competing product, there is reproduction in the AI process and therefore an obligation to have a licence to make such a reproduction. Furthermore, this organisation commented that, if the AI user derives value from the exploitation of the output, whether recognisable or not, they should acquire a licence and pay remuneration for the author. On this other side, only 18% (11 respondents, of which 5 are legal practitioners/academics, 2 trade associations, 1 artist/rightholder, 1 CMO, 1 producer and 1 developer of AI solutions⁵²⁷) believed AI practice to be covered by an exception (and 5% (3) or that it has no copyright relevance (see Figure 37).

⁵²⁶ This categorisation of respondents is based on a multiple-choice question in the Delphi survey. The categories are not mutually exclusive.

⁵²⁷ Ibid.

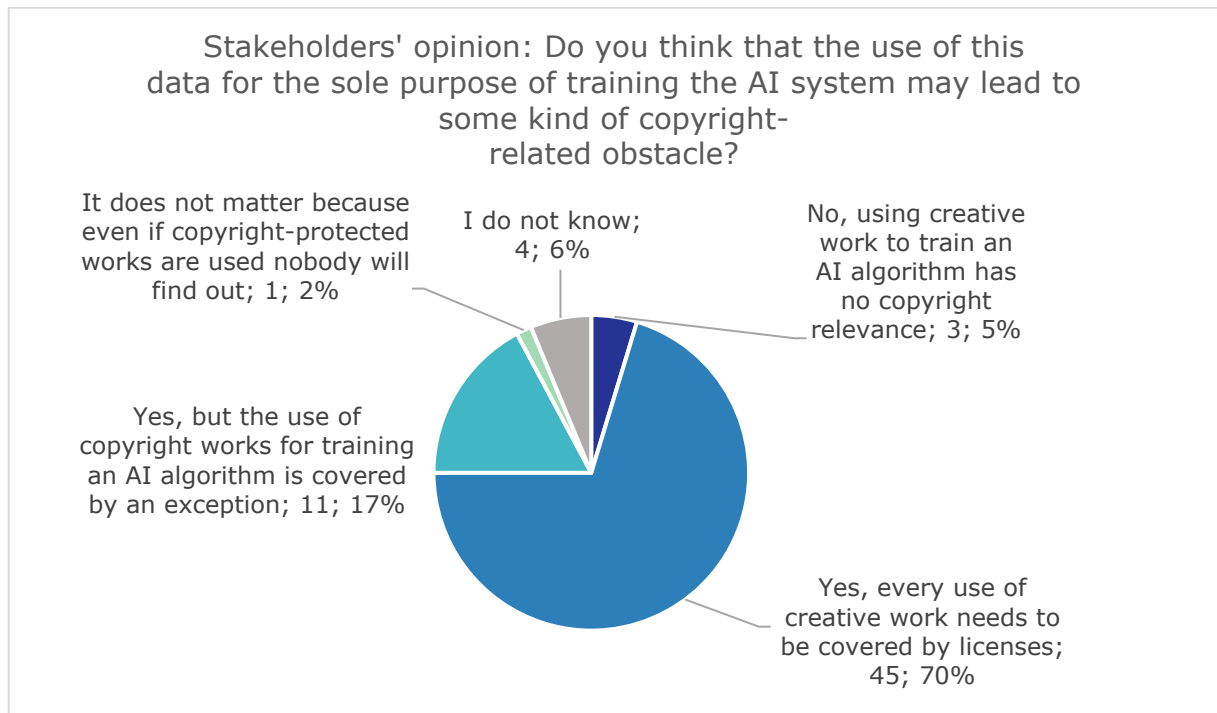


Figure 37: Stakeholders' opinion on whether the training of AI leads to some copyright-related obstacle (N=64)

Source: Technopolis Group Survey

By contrast, the majority of respondents to the in-depth interviews (stakeholders and legal scholars) clearly considered the processing of data to train AI systems as an act that could fall under the TDM exceptions (indeed TDM refers to the processing of data “with a view to gaining new knowledge and discovering new trends”⁵²⁸, and for “the development of new applications or technologies”⁵²⁹).

Furthermore, the Directive does not specify how the different exceptions apply – except for recital 9 of the DSM dir. which indicates that the exception for temporary reproductions (under Art. 5(1) InfoSoc dir.) remains available. Several interviewees mentioned the uncertainty concerning the cumulative application of exceptions. Is it possible to exempt a TDM process under other exceptions, such as the exception for temporary acts of reproduction or the research exception, if the TDM exception does not apply, in particular in case of opt-out by one or several rightholders?

⁵²⁸ See recital 8.

⁵²⁹ See recital 18.

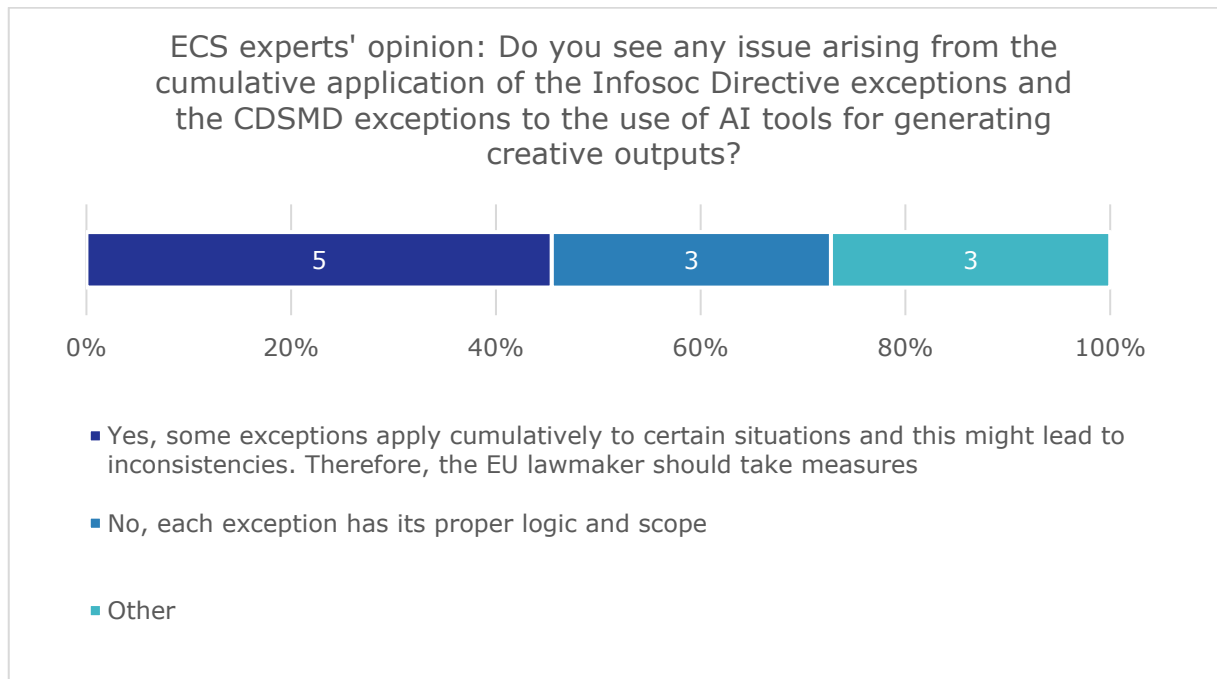


Figure 38: Opinion of ECS experts on the cumulative application of exceptions (Total votes=11)

Source: Technopolis Group Survey

Obviously, no general assessment of the TDM exceptions can be made – considering that the TDM exceptions have not yet entered into force in the Member States. Yet, some important questions in the application of the TDM exceptions to AI-related activities may already be expected.

Although questions about the **adequacy and current scope of TDM exceptions** were not addressed in the Delphi survey (considering the unfinished transposition of the DSM Directive), the in-depth interviews revealed that many respondents are dissatisfied with them.

The representatives of one large technology company stated that articles 3 and 4 DSM are key legislative developments and a positive step in terms of access and exploitation of data. However, they did raise concerns related to the TDM exemptions. In their view, Article 3 of the DSM reflects a research environment that does not exist since 'pure' public research institutions do not exist in reality (there is always a commercial component). However, the respondents to the Delphi survey might have overlooked that Article 3 applies as well in the case of research conducted within a Public-Private Partnership. With regard to article 4 of the DSM, they identified a challenge in that it only deals with reproduction, but does not contain any rights to distribute or communicate to the public the input dataset for further training the AI system. Thus, according to them, article 4 does not accommodate their needs because it only gets them halfway through.

Another multinational technology company held the opinion that, although at this stage it is still too early to tell, there will be legal uncertainty once the TDM exceptions will be implemented in all national legislations. For instance, they raised a question on how these TDM exemptions will practically interact with existing exceptions. They preferred that the interplay between the TDM exemptions and existing exceptions be clearer (the circumstance that rec. 9 clarifies the relation with art. 5(1) InfoSoc Directive does not bear

any teachings on the relation with the research exception provided in 5(3)(e) Infosoc Directive).⁵³⁰ Moreover, they mentioned the lack of harmonisation between EU member states' copyright legislation (and a fortiori TDM exemption) as a legal obstacle.

A visual media company indicated that they do not feel at ease with the existing TDM exemptions under the DSM dir. Regarding the scientific research exemption, they would like to see a clearer line between what constitutes scientific research and commercial exploitation. They pointed out that research is often funded by private companies and that the result of the research eventually ends up in a black box where it is no longer possible to distinguish whether a licence is needed or not. Lastly, one public institution in the cultural sector mentioned that these exemptions were very much needed in the cultural sector.

The possibility given to the rightholders to reserve their rights may limit the impact of the general exception for TDM provided in Art. 4. By opting out of the TDM exception, the rightholder (copyright, related right, *sui generis* database right) can regain exclusive control over the use of their work or other subject matter.

A first issue that was repeatedly flagged in the interviews concerns the conditions for **exercising and communicating the reservation of rights**. According to Art. 4(3) DSM dir., the reservation by the rightholder should be made "expressly" and "in an appropriate manner, such as machine-readable means in the case of content made publicly available online". As could be observed in the expert interviews conducted for the present study, legal experts' opinions diverge with regard to the interpretation of this condition. Some argue that the rightholder should be compelled to justify their choice, whereas for others only minimal efforts (for instance in the form of implementation of machine-readable code on their website) should be required from the rightholder side. Divergent opinions were noted, ranging from a single exercise of the opt-out (seemingly assuming the existence of some central collecting point for opt-outs) that a user ought to consult prior to TDM processes, to the exercise of the opt-out per publication (to the analogy of robot.txt files⁵³¹). Several interviewees among the stakeholders have expressed their concern about the implementation in national law, which could lead to fragmentation in a field where they require scale and consequently a harmonised approach among the Member States.

Another issue is the accumulation of exclusive rights: any set of "data" can be protected under copyright and several related and *sui generis* rights. The opt-out of the TDM exception by one rightholder may entail the prohibition of processing the dataset altogether, even if other rightholders do not object. Interestingly, the holders of database rights may thus prohibit the use of their datasets, which may include protected works and other subject matter of rightholders who do not object to the TDM process. Data traders or companies offering training data may thus systematically exercise their opt-out rights to their protected databases (if any) in order to safeguard existing monetisation models, thus overruling preferences of other rightholders and ending up controlling the information contained in the database and the works. Where the same works, performances and recordings can be found in other databases, the opt-out of the maker of one database can

⁵³⁰ As a matter of fact, it should however be noted that Art. 25 DSM dir. explained that "Member States may adopt or maintain in force broader provisions, compatible with the exceptions and limitations provided for in Directives 96/9/EC and 2001/29/EC, for uses or fields covered by the exceptions or limitations provided for in this Directive."

⁵³¹ Such practice would require that the administrator of the page (such as a webpage) have the right to exercise the opt-out for all elements of content on the page. This might be the case for a webpage that is administrated by the author of all content on the page. Mostly datasets will contain content from different rightholders, meaning that the expression of the opt-out at the level of the dataset may not reflect the position of all such rightholders.

be lawfully circumvented by using an alternative database in which the same or similar material is included.⁵³²

Finally, in practice, it may prove difficult to verify compliance with the opt-out. TDM processes are mostly invisible to the public and carried out without prior information to the rightholder: the rightholder is not necessarily aware or notified that their material is used in a TDM process – especially because the protected features of their content will not necessarily appear in the AI output. When one or more rightholders have reserved their exclusive rights, their works, performance or recordings may not be used for training AI algorithms. As a matter of fact, it will be impossible for them to verify whether their exclusive rights are actually observed. Even if they are aware of the TDM process and they suspect that their works are concerned, the rightholder has no obvious legal basis to request access to the process (e.g. by means of audit rights) or to force the AI solution provider to demonstrate that the protected content has *not* been used.⁵³³ This additional difficulty in the verification process may hinder the effective opt-out of the TDM exception and, in case of opt-out, the enforcement of the (revived) exclusive rights.

Beyond this question of interpretation, several experts and stakeholders (especially developers of AI solutions) interviewed expected that the absence of standardisation of this opt-out would lead to legal uncertainty. Without more specific instructions on the exercise of the opt-out, interviewees expressed the concern that different practices will arise in the Member States or among different (larger) players, be it on the rightholders' or on the AI solution provider's sides.

Depending on the technical developments in the field of AI, the actual implementation in the national copyright laws and the resulting practice, these questions may need to be addressed (see section 3.4.1.2).

3.3.7. Transversal issues – How do the copyright rules relate to other AI-relevant legal regimes?

3.3.7.1. Information obligation – transparency?

Several studies demonstrate that knowing that a given piece was written or composed by an AI solution **affects consumers' experience**. The studies reveal that users are **negatively biased** against algorithmic creation.⁵³⁴

⁵³² This concern may require the AI-developer to precisely document their training data: a database maker may conclude to the infringement of their rights if the training set contains the same content as their database. The AI developer must then be able to refute the copying of that database by demonstrating that the content of the training set might correspond to the content of the protected database but has not been transferred from that database.

⁵³³ In addition, the AI solution provider may invoke the protection of their trade secrets to oppose any audit or forced transparency on the training process. The allocation of the burden of proof is a predictable point of conflict in any future litigation. See however EP (Committee on legal affairs), Report on intellectual property rights for the development of artificial intelligence technologies, https://www.europarl.europa.eu/doceo/document/A-9-2020-0176_EN.html (accessed in July 2021), par. 18: "considers that non-personal auditable records of data used throughout the life cycles of AI-enabled technologies in compliance with data protection rules could facilitate the tracing of the use of copyright-protected works and thereby better protect right-holders and contribute to the protection of privacy, if the requirement to keep auditable records were extended to cover data containing or deriving from images and/or videos containing biometric data".

⁵³⁴ In this regard, one of our interviewees mentions that this impact might, in reality, depend upon the public targeted. All studies cited are indeed dealing with audio or writing works. No study seems to have been done on video-gamers, who might, on the contrary, be positively biased toward the implementation of AI technologies.

The **transparency** of AI output providers regarding the reliance on AI solution is difficult to ascertain. On the one hand, certain AI output providers voluntarily and explicitly flag that a given output was generated automatically (e.g. DeepL). On the other side, the issue of false authorship proves that other providers might attempt to hide the AI-origin of the creations.

In this context, the question is whether a **legal obligation to inform** the public of the AI / human provenance of a given output could be useful from the copyright perspective. Similarly, the draft regulation on AI provides a transparency obligation to public, in the sense that “providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that **natural persons are informed that they are interacting with an AI system**, unless this is obvious from the circumstances and the context of use”⁵³⁵ (emphasis added).

This increased transparency could allow consumers to make **informed choices** regarding the type of content they choose to consume. If the results of the behavioural studies mentioned above are correct, such increased transparency should also **favour human creation over AI-based creations**.

However, the adoption of such transparency obligation is not without **practical concerns**. At which point should an output be stamped as “AI-generated”? If such an obligation applies as soon as an AI solution is used, then all outputs in which AI has played a minor or major role should be labelled as “AI-generated”, even if a human still makes major creative choices and the output is protected under copyright. Such large obligation would be disproportionate, from a copyright perspective.

The transparency obligation could then be limited to the output that is autonomously generated by AI, without human intervention of any major importance. This means that the absence of human intervention and consequently the absence of copyright protection would be communicated. In practice, this obligation may not be feasible, neither effective, since it may not be verifiable whether any given output is created entirely or partially by an AI system.

If this practical obstacle can be solved (e.g. by means of AI-driven detection of non-human creations), compliance with the obligation would however be possible and this could address the false authorship issue.

The potential of this transparency requirement should hence be further investigated, and the scope of the obligation should be specified more precisely.

3.3.7.2. Coexistence of various data regulations

While the European Commission is taking various initiatives to stimulate the sharing of “data” and the development of innovative data applications and/or infrastructure, other types of regulation might restrict the sharing and processing of the same data. While, in theory, these sets of rules can apply to different portions of “data”, in practice, seemingly contradictory rules will apply to the same processes and the same data set.

For example, the copyright rules may contain a generous carve-out for TDM, but the same processes may be subject to stringent data protection rules in which other exemptions are provided. Similarly, both copyright and data protection may provide exceptions for

⁵³⁵ Art. 52(1) Proposal for a Regulation laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts, {SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}.

“research” but the notions of “research” and the conditions for the exceptions may not be harmonised.⁵³⁶

This leads to a complex whole of applicable rules, which are not always straightforward to articulate and which may entail a paralysing legal insecurity (considering the prohibitive sanctions that might be associated with a violation of the applicable rules).

3.4. Policy scenarios

Several challenges to the legal framework of copyright and the related rights have been identified in the light of development and use of AI-solutions in the selected cultural sectors. Although much will depend on the development of AI technologies, their applications and the business models, this section presents and discusses certain “policy scenarios”, i.e. potential forward-looking changes to the copyright system which may be considered to address the AI-related challenges described in section 3.3. The different scenarios have been drafted based on the information gathered through desk research and semi-structured interviews conducted with legal experts from academia and targeted stakeholders. For each section, a recap of the issue and the questions are presented. Questions in relation to these policy scenarios have been submitted to the stakeholders and experts in the frame of a Delphi survey.⁵³⁷ The results obtained in the frame of the latter and through in-depth interviews for each scenario are presented below. These scenarios serve as a starting point for further reflections on the evolution of copyright and for future discussions and analysis, under the subsection entitled “discussion”.

The current expectation is that, in some sectors (music, visual arts), autonomously generated AI output (music performed by non-human actors, photo-like images automatically generated) without human creators may replace part of the existing practices and impact human creators, while in other sectors, more complex creations (such as audiovisual content or video games) will continue to require a significant human creative effort – albeit supported by AI tools. It is also expected that the business models will partially shift to a service-based model, through which a user acquires unlimited rights to automatically generated content. The AI-content provider does not monetise its AI solution and AI-generated content by licensing the same content to selected licensees (in function of territory, duration, exclusivity) but by constantly generating new AI output that feeds the activities and business operations of the contracting party. It is, however, too early to predict the evolution of the business models and the role that exclusive rights will play as a legal basis for such business models (especially concerning contractual protection).

The policy scenarios are developed and assessed in light of the purposes of copyright and related protections⁵³⁸. Such purposes may be:

- ensuring a high level of protection for (human) authors and performers and other rightholders and, consequently, safeguarding a fair and reasonable remuneration.

⁵³⁶ See also J.-P. TRIAILLE, “The exception for scientific research under Eu copyright law and Eu privacy law”, *Law, norms and freedom in cyberspace. Liber amicorum Yves Pouillet*, Brussels, Larcier, 269 et seq.

⁵³⁷ As indicated in the methodological remarks in section 3.1.4, participants of the Delphi survey included both individual stakeholders and organisations. Therefore, answers may reflect the opinion of a single individual or the several individuals/organisations represented by an organisation taking part in the Delphi survey (e.g. CMOs, cultural associations, etc.).

⁵³⁸ EP (Committee on legal affairs), Report on intellectual property rights for the development of artificial intelligence technologies, https://www.europarl.europa.eu/doceo/document/A-9-2020-0176_EN.html, (accessed in July 2021) par. 6. See also rec. 2 DSM dir.

- creating an incentive for authors and other producers of AI-generated output to innovate, create and disseminate the creations to the public (correcting market failures).
- taking into account the public interest in the development of innovative technologies, the availability of high-quality creative content and the unrestricted use of information.

The identified policy scenarios are meant to address the challenges identified in section 3.3 and presented as relating either to the AI input or the AI output issues.

In the opinion of most legal experts participating in the Delphi survey implemented for this study, AI practices raise copyright uncertainties or challenges. Regarding the use of protected works or other subject matter as an input for developing AI tools, 48 out of 65 experts (74%) find considerable or some uncertainties or even perceive some impediments to the use of protected creative works as input to train AI systems. Concerning the protection of AI output, 52 experts (80%) find considerable or some uncertainties regarding the IP protection of the output generated by AI in the creative sector (see Figure 39).

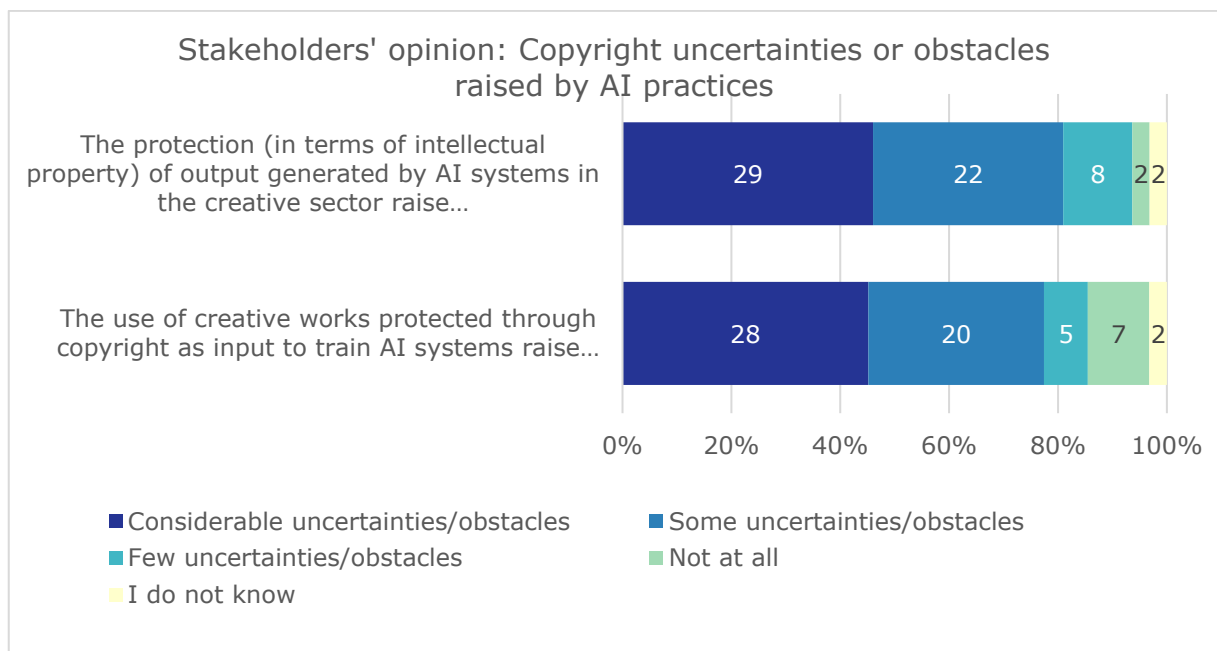


Figure 39: Stakeholders' opinion on copyright uncertainties raised by AI practices (N=64)

Source: Technopolis Group Survey

In the following section, the assessment of the policy scenarios is presented in further detail.

3.4.1. Considering the copyright status of the creative input used for AI training

Before an AI solution can autonomously generate cultural output (such as music, videos, images, ...), the AI solution is trained using datasets that often contain protected content

(works, performances and/or recordings) which may or not be contained in a database protected under the sui generis rights.

The participants in the Delphi survey, the stakeholders and the expert interviewees were asked to express their opinion on the following points (identified as the most relevant), related to the use of protected contents as input for training AI solutions:

1. Should the material acts of using the protected content in the process of training an AI solution be protected under the reproduction right?
2. Should such acts be exempted under the TDM exceptions and under which modalities?
3. Should the authors and performers be able to rely on their moral rights to oppose the use of their protected works or performances for training AI solutions?

For each topic, the issue will be outlined with the different policy scenarios, the results of the empirical research (the in-depth interviews and the Delphi study) will be presented and their potential impacts discussed.

3.4.1.1. Should the training of an AI solution with copyright protected works require the authorisation of the copyright owner?

If the TDM exception was developed for the purpose of stimulating the use of data to develop innovative AI-driven applications, then the current copyright framework might potentially pose some challenges (as identified in 3.3.4, 3.3.5 and 3.3.6). If that would be the case then these might be addressed by re-examining the scope of protection of the reproduction right, the application of the TDM exceptions and/or the exercise of the moral rights.

3.4.1.1.1. *The scope of protection of the reproduction right and the TDM exceptions and policy scenarios*

The use of datasets that contain protected works, performances and/or recordings, for the purpose of training AI algorithms can be assessed from several copyright angles, to conclude whether or not such use requires the rightholders' prior consent.

The "processing" of "data" arguably involves material acts of copying of a technical nature, which could be considered as not falling under the reproduction right. In that case, the rightholders' prior consent is not required for the use of their creations to train AI solutions (see above 3.3.4.1.3). Some may however argue that the introduction of the TDM exceptions in the DSM dir. implies that the technical acts needed for mining some content are falling under the reproduction right: where they are not covered by the notion of reproduction, an express exception in legislation would not be required.

The scope of the TDM exceptions is the next issue. Arguably the use of all protected subject matter as input for training an AI solution is covered under the definition of "text and data mining". However, a literal and narrow reading of the TDM exceptions could restrict their application to the *analysis* of information, considering the wording of the definition of TDM as an "automated *analytical* technique aimed at *analysing* text and data in digital form in order to generate information"⁵³⁹. This explicit reference to the *analytical* purpose would then cast doubt on the possibility to use TDM processes to subsequently *generate* cultural output under this exception. While text and data analysis technologies aim at recognising patterns and extracting insights from existing data sets and are used mostly outside the

⁵³⁹ Art. 2(2) DSM dir.

creative sectors, AI technologies can also be used to create cultural content, that can sometimes function as a substitute to the content used to train the AI solution. Some stakeholders may feel that the TDM exception must not allow such competing uses, while others may hold the opinion that all such technical processes may be exempted – regardless of the purpose for which the outcome is used.

In the interviews and the Delphi survey, it was examined whether the scope of protection of the reproduction right and the TDM exceptions may pose any particular problems. It is not yet clear whether the implementation of the TDM exceptions will lead to practical issues and litigation, but it appears that the following policy scenarios were worth to be investigated:

- **status quo**
- legal clarification that the processing of data for AI algorithms' training **is not covered under the scope of protection offered by the reproduction right**.
- legal clarification that the processing of data for the training of AI algorithms **falls within the scope of the TDM exceptions** (for instance through guidances).
- introducing **a distinct copyright exception** for the use of works and other protected subject matter as data for training AI applications for the ultimate purpose of generating cultural creations.
- legal clarification that TDM exceptions **apply when data are processed for AI training for "generating information", not to "create" new creative output**.
- legal clarification that TDM **exceptions encompass AI training for non-commercial use of the technology, not for commercial purposes**.

3.4.1.1.2. Stakeholders' and scholars' opinions

As can be seen in Figure 40, the stakeholders participating in the Delphi survey expressed different views as to the need to clarify that the training of AI is or not covered by the TDM exceptions. Out of the 22 experts participating in the Delphi survey (and strongly agreeing with the need to clarify that the training of AI is not covered by the TDM exceptions), 18 are rightholders or organisations representing rightholders in the creative sector.⁵⁴⁰ Five participants stated that for them it is clear that the TDM exception was intended to cover AI practices. However, they believe that a clarification would be useful for the sake of legal certainty.

⁵⁴⁰ The last two policy scenarios displayed in Figure 40 were suggested by survey participants in the first round and added to the second round of the Delphi survey.

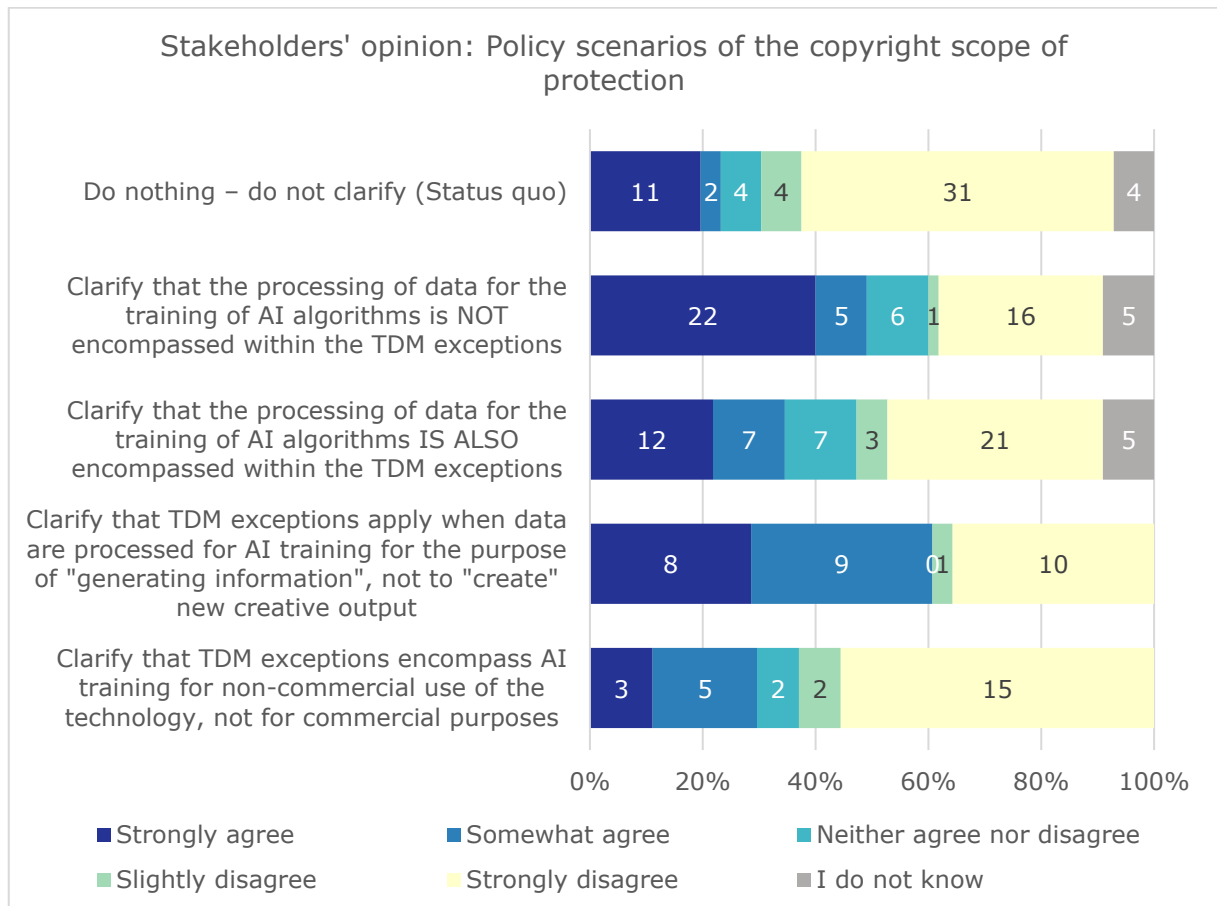


Figure 40: Stakeholders' opinion on scenarios of the copyright scope of protection (N=56 for the first three scenarios, N=28 for the fourth and N=27 for the last one)

Source: Technopolis Group Survey

Stakeholders *supporting the **status quo*** (6 legal academics/lawyers, 2 CMOs and 1 producer and 1 artist/rightholder) argued that the current legislation is an adequate basis for the time being. They argue that nothing would indicate that the balance between copyright and exceptions needs to be recalibrated. Therefore, it would be wise to let the Directive plays out and see to what extent the framework in the EU holds lessons when applied in practice, rather than starting to regulate, tilting in one direction or the other.

Experts participating in the Delphi survey who were in favour of clarifying **that the use of data for the training of AI is not encompassed** by the TDM exceptions, provided the following arguments for their opinion:

- 1) **Exploitation of creative works and remuneration:** They argued that AI training is a form of exploitation of creative works, and therefore, the use of copyright-protected works need the consent of the rightholders. AI-generated art can only be produced by using human-created art (protected under copyright). Under this view, the role of creators should be acknowledged in the copyright law.
 - One IP expert from the publishing sector argued that the TDM exception without the right to remuneration for the rightholder is opening the door wide for companies to create competing products that imitate and partially replace human creative works, without the aim of providing good for the community (as it is the case for

scientific research). This expert has also argued that the mass use of works increases the risk of illegal acquisition, and the creation of libraries of copies that are unlikely to be controlled and where deletion cannot be guaranteed. This would lead to disruption of the European creative and cultural market.

- Moreover, some experts express concern toward the fact that AI products would be largely produced using European creative works but benefit companies based **outside the EU**, thus reinforcing geo-political imbalances in the data economy.
- Another IP expert from the book sector argued that the main **beneficiaries** of the TDM exception are not academics, applying these TDM processes for research purposes, but rather **tech companies**, such as software manufacturers and other tech & communication companies.
- **Interpretation and application of the TDM exception:** In the opinion of two experts (an IT law expert and an IP expert working at a CMO), the training of AI algorithms serves a different purpose than the one pursued by the TDM exception. Since the copyright exceptions should be interpreted strictly, the use of creative content to train AI would then be justified in the context of scientific research, but not for other purposes. One stakeholder from the music sector also mentioned the importance of the “non-commercial research” requirement for the application of the TDM exception,⁵⁴¹ as a significant part of the AI technology has a commercial purpose.
- In the opinion of two participants (one IP expert in the music sector and one expert in IT law), copyright exceptions should meet **the three-step test** under the Berne Convention. In their view, the use of creative works for any training of AI is contrary to the three-step test.
- Another stakeholder with expertise in IP and consumer law observed that, in case protected content used to train AI is identifiable in the output of the AI, the training is subject to the licence of the rightholder (e.g., the training would be considered infringement). The *rationale* is that in such a case, the act is a reproduction, arrangement or alteration.
- **Establishing limits for a fast-evolving technology:** Two other participants (one IP expert working at a CMO and an individual artist from the music sector) defended a more cautious approach, in the light of the uncertain impact of AI on IP. At present it is unknown to what extent AI will conflict with or disturb traditional IP management, hence some limits of AI use should be established. Early intervention could protect creators’ rights and support the industry overall.

Participants in the Delphi survey who favoured the clarification that **AI training is included by the TDM exceptions** provided the following arguments:

- **Development of AI and investments in European AI:** A producer and developer of AI from the cultural heritage sector stated that the TDM exception is key to the development of web applications and AI solutions, as copyright law tends to block the development of such innovations. For them, access to data is not only in the interest of big technology and communication companies, but it is also important for small IT companies that develop new solutions. Another stakeholder

⁵⁴¹ Even if some stakeholders seem to read it that way, it should be specified that the text of the Directive does not mention such a ‘commercial purpose’ condition.

participating in the survey argued that clarifying that the TDM exceptions extend to AI would encourage investment in AI, which is believed to generate growth and socio-economic benefits. In the opinion of this participant, it is sufficient that rightholders can opt out of the TDM exception, while the application of the TDM exception operates as a default position.

- **Transactions costs linked to the TDM licences.** Another participant with expertise in IT law added that obtaining TDM licences for commercial use may not be affordable for many start-ups and small businesses. Such companies would then outsource their TDM activities in order to be competitive in the AI sector, which may lead them to set up directly in jurisdictions with more flexible access to and use of materials for training AI systems. For them, the legal uncertainty and legal complexity could also lead to brain drain, and to decisions to divest in the EU by companies from outside the EU.
- **AI as a fundamental technology and the risk of AI bias:** A stakeholder specialised in IP and IT-related law argued that AI is a fundamental technology used in many applications and devices, which should not be burdened with costs or uncertainties within the EU, especially considering the recent decisions of the US Supreme Court on fair use (including in the *Oracle v. Google* case). Another expert (a developer of AI solutions) shared the opinion of AI bringing benefits for the progress of many fields. This expert also argued that it does not make sense to restrict the access of AI to creative works, and used the analogy of an individual going to a library, getting a book, reading and understanding it, and this process resulting in a copyright issue.
- **Largest access to data.** Another stakeholder from the visual arts and design sector explained that, in the absence of a TDM exception, training AI with only licence-free data would result in a too one-sided AI model, and therefore, a bias in the system. Therefore, AI should be trained with the widest possible range of data.
- **AI training vs. AI creation:** One of the stakeholders in the field of visual arts and design argued that, from a copyright point of view and to protect rightholders, it is necessary that data processing be limited to the training itself. In their opinion, the generation of creative output by AI is not covered by the exceptions of the Directive, and may result in a copyright-relevant act.

Delphi participants had a **divided opinion** on whether there should be a **clarification** that TDM exceptions apply when data are processed **for AI training to generate information but not to generate new cultural output**.

- For some participants, this policy scenario would address the needs of the AI and technology sectors to be able to innovate and develop new products, while human IP creators would have their IP rights protected and rightholders would be able to exploit new opportunities in licensing their works for AI creation.
- Two other stakeholders noted that AI should be treated uniformly from a legal perspective, and that the differentiation in terms of the purpose of the activity (e.g. "for general information" and "for creation") would introduce a layer of legal uncertainty. Moreover, given that a trained AI model is a black box, developers re-using available trained models might not be able to determine the individual elements of the training dataset, and it could be impossible to ensure that a given trained dataset will only be used to generate information instead of generating creative output.

- The remaining participants disagreed, either because they suggest that licences should always be necessary for AI processing, or because they interpret that licences are not needed due to the TDM exceptions.

The policy scenario that TDM should **only apply to the AI training for non-commercial uses** of the technology was rejected by a higher proportion of participants. Some participants explained that non-commercial use is not a decisive aspect and its borders might not always be clear. For instance, in the case of subsequent commercialisations of research efforts, or in case the commercialisation is carried out by a different legal entity, or in the case that the output of AI is not commercialised but is used in a commercial environment (e.g. AI used in recommender systems of platforms offering human-made creative works).

Stakeholders participating in **the second round** of the survey were asked to assess the impact of three policy scenarios: the status quo, the scenario in which a legal clarification that data processing for AI training is encompassed by the TDM exceptions, and the scenario in which data processing for AI training is not encompassed.

As it can be observed in the Figure below, most participants agree that continuing with the **status quo** will give time to observe how different transpositions of the TDM exceptions influence the creative market in relation to AI. This is followed by the agreement of 58% of the respondents on the view that, with the status quo, AI companies will assume that AI training is also encompassed by the TDM exceptions. Participants showed less consensus on the impact of the status quo on the flourishing of AI and the attractiveness of the EU for IT companies.

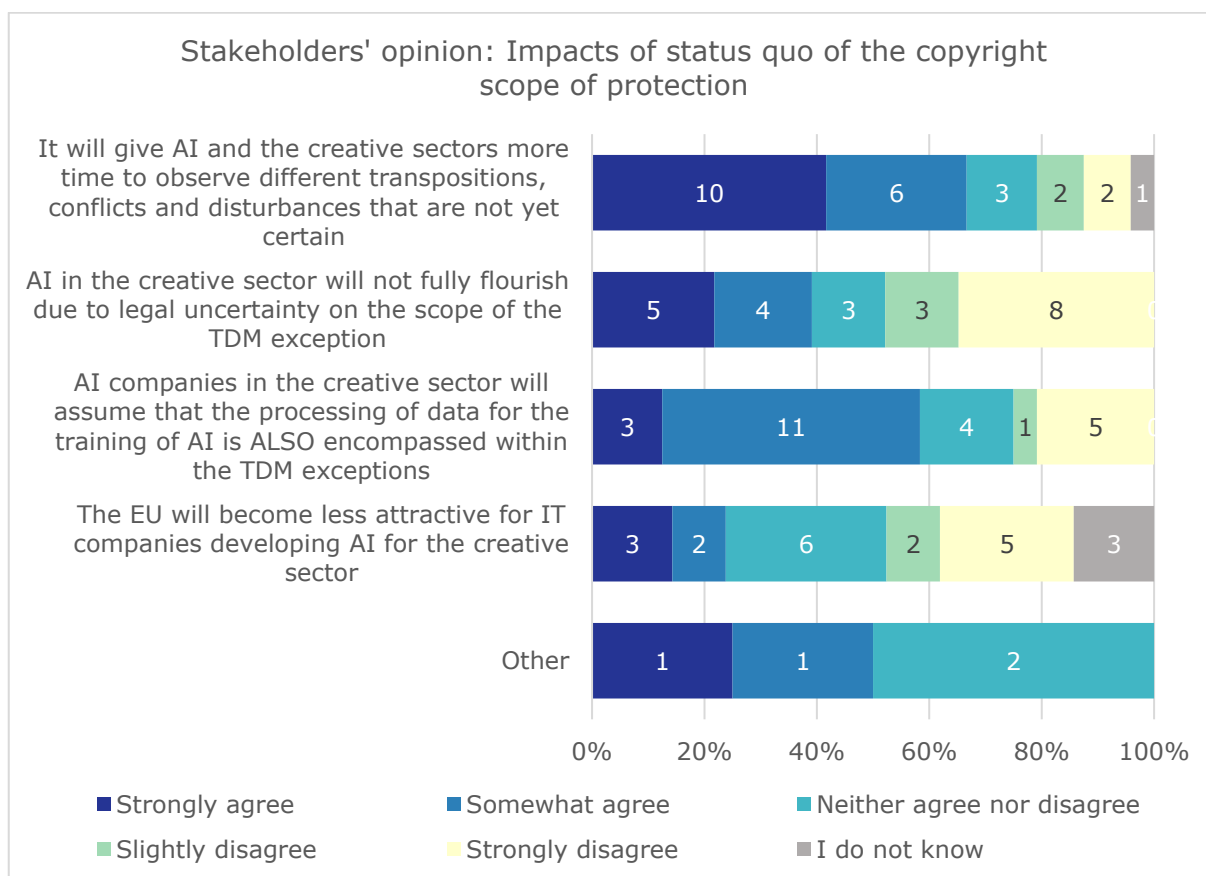


Figure 41: Stakeholders' opinion on the impacts of the status quo of the copyright scope of protection

Source: Technopolis Group Survey

The stakeholders have divergent views on the impact of a clarification that **AI training is encompassed by the TDM exceptions**. Most think that this training is indeed not covered by those exceptions (see Section 3.3.6.2). In their opinion, the most negative impacts of having AI training encompassed by the TDM exceptions affect the value of the human-made works and the revenue of artists. They also stated that the cost to protect copyright works against unauthorised use would be negatively affected if AI training is encompassed by the TDM exception – presumably due to their investment in technological measures to prevent access to AI robots or due to their reticence to deal with the opt-out. It can also be deduced from their answers that the impact of such contrasting policy scenarios has, in their opinion, more impact for authors/rightsholders than for IT companies. In this regard, it is worth remembering that the largest group were rightsholders or organisations representing rightsholders in the creative sector. By contrast, as a recall, the majority of respondents to the in-depth interviews (stakeholders and legal scholars) clearly considered the processing of data to train AI systems as an act that could fall under the TDM exceptions

Stakeholders' opinion: Impacts of policy scenarios of the copyright scope of protection

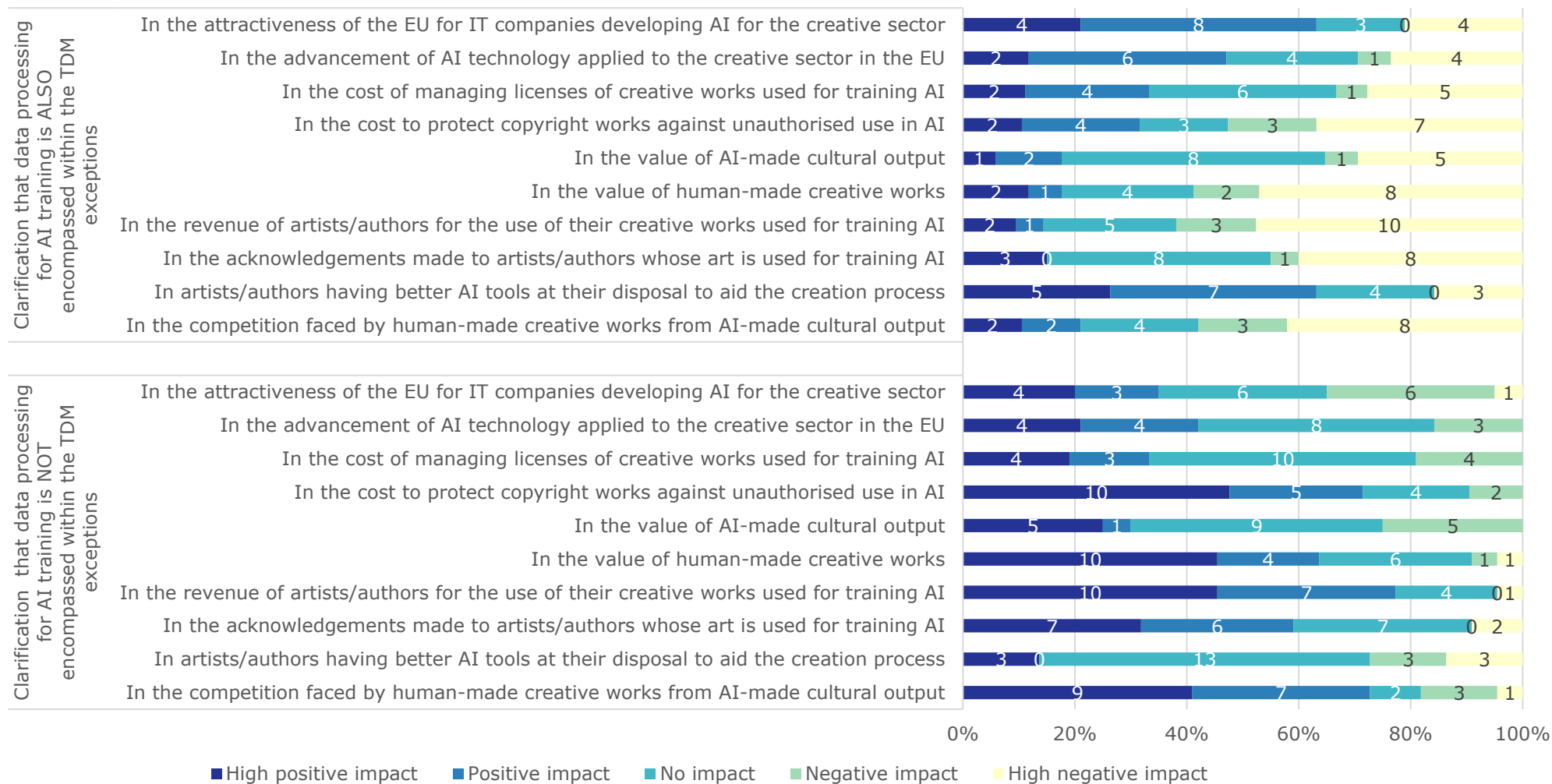


Figure 42: Stakeholders' opinion on the impacts of policy scenarios of the copyright scope of protection

Source: Technopolis Group Survey

3.4.1.1.3. Discussion

Scope of the reproduction right

The vast scope of protection of the reproduction, covering ephemeral, invisible copies of a purely technical nature (see Section 3.3.4.1) can be explained by historic reasons and has been criticised in scholarly literature.⁵⁴² Especially for the ephemeral and partial reproductions, the literal and technical reading equates every “copy” (in the technical sense) to a “reproduction” (in the legal sense), without considering the importance of the “copy” for the exploitation of the “work”. Arguably, the notion of “reproduction” could be revisited to align its content with the purpose of the economic rights, i.e. protect the exploitation of the work or other protected subject matter. To the extent that the technical copies made in the course of a TDM process serve to extract information, meta-data or other insights from the processed training data set, the protected creations in the data set are used for the sake of the information they contain, not for the sake of commercialising, exploiting or putting forward the protected works in their original expression.

It could be argued then that the contours of the reproduction right should be revisited to focus its protection on the protection of the works (and other subject matter) and their exploitation of the protected aspects (i.e. the original expression, the recognisable performance, the recognisable recordings).

Although the decisions of the CJEU on the reproduction right have oftentimes put forward a technical and literal reading of the reproduction right, the Court has demonstrated an openness to a more flexible and more balanced approach of the reproduction right in the *Pelham* decision. In *Pelham*, the Court contrasted a **literal** interpretation (which would be warranted by the purposes of providing a high level of protection and protecting the producer’s investment) to a **contextual** interpretation (which would seek a fair balance with other fundamental rights) of the notion “reproduction in part” of a phonogram⁵⁴³.

Interestingly, the Court opens the door for a flexible interpretation of the reproduction right, depending on the context of each case: when a literal interpretation of the partial reproduction may conflict with the fundamental rights of the users or with the public interest, the court adapts the contours of the partial reproduction to find a balance between both. Hence, the protection of intellectual property must not automatically prevail over other fundamental rights: when the protection of the producer’s right conflicts with the freedom of the arts (as part of the freedom of expression), the use of a 2-second recording, unrecognisable to the human ear, in a new expression would not be a “reproduction in part” under the InfoSoc Directive. The benchmark for the court is the “usual meaning” that is given to the notion of “reproduction” and the fair balance between the fundamental rights. Importantly, the Court notes that the literal interpretation of the reproduction right would allow “the phonogram producer to prevent another person from taking a sound sample - even if very short - from his or her phonogram for the purposes of artistic creation in such a case, despite the fact that **such sampling would not interfere with the opportunity which the producer has of realising satisfactory returns on his or her**

⁵⁴² See inter alia A. STROWEL, “Reconstructing the reproduction and the communication the public rights: how to align copyright with its fundamentals”, *Copyright reconstructed*, 2018, p. 203 and S. DEPREUW, *The Variable Scope of the Exclusive Economic Rights in Copyright*, Brussels, Kluwer, 2014, 189 et seq. and references there.

⁵⁴³ Regarding the right of communication to the public, the CJEU has repeatedly decided that this concept requires an individual assessment, taking into account several complementary, non-autonomous and interdependent criteria (see e.g. Judgment in VG Bild-Kunst, C-392/19, paragraph 33-34).

investment"⁵⁴⁴ (emphasis added). Finally, the Court opted for **asymmetrical criteria**: it acknowledges that the ratio for protection of phonograms is the investment by the producer, but the criterion for finding a "partial reproduction" is not the investment (unlike for the sui generis database rights) but the possibility to recognise the part of the recording used in a new creation.

The *Pelham* decision was issued in a case relating to the phonogram producer's rights. The question is then whether this remains an isolated case or if the Court will continue this line of reasoning in other cases on the related rights and even extend it to copyright cases.

Considering that the reproduction right is harmonised for copyright and the related rights in Article 2 of the InfoSoc Directive, it should be interpreted in a consistent way under copyright and the related rights.⁵⁴⁵ Moreover, the same narrower interpretation of the reproduction right for authors and performers could arguably be justified by the need to ensure them "an appropriate reward" (not the maximum remuneration) for their creative input (recital 10 InfoSoc Dir.). On the other hand, in its decision in *Pelham*, the Court insists on the "second" objective i.e. "the specific objective of the exclusive right of the phonogram producer, referred to in recital 10, which is to protect a phonogram producer's investment".⁵⁴⁶

Similarly, the CJEU has decided in *CV Online Latvia*, concerning the database directive, that the infringement of the rights of extraction and re-utilisation should be assessed with regard to (i) the material acts of copying the content of a database and making it available to the public and (ii) the impact of such material acts on the investment by the database maker. In this case, it should indeed be verified whether a specialised search engine, which copies the content of freely accessible databases and which gives its users access to the same, **affects the investment of the database maker**.⁵⁴⁷ The courts should thus strike a **fair balance** between the legitimate interest of the makers of databases (i.e. redeem their substantial investment) and the interests of users and competitors (i.e. having access to the information contained in those databases and the possibility of creating innovative products based on that information).

The Court summarises that the "substantial investment" in the obtaining, verifying or presenting of the contents of the database is central for (i) determining whether any given database can be protected under the sui generis right and (ii) deciding whether acts of copying and making available infringe the rights of extraction or re-utilisation, i.e. when such acts constitute a risk to the possibility of redeeming that investment.⁵⁴⁸

With the decisions in *Pelham* and *CV Online Latvia*, the Court seems to adopt a more flexible approach to the exclusive rights. In addition to the detriment to the investment, it is willing to consider the **interests of other stakeholders** (users, competitors, general public) at the stage of **finding an act protected under the respective exclusive rights** (rather

⁵⁴⁴ CJEU 29 July 2019, *Pelham GmbH and Others / Ralf Hütter and Florian Schneider-Esleben*, C-476/17, EU:C:2019:624, paragraph 38.

⁵⁴⁵ See on this point Judgment in *FAPL*, paragraph 188 and in *Reha Training*, paragraph 33 – where in both cases the right of communication to the public in different directives was at under consideration and in *Reha Training* the Court decided that the notion of "communication to the public" should be given the same meaning, even if the nature of the right (exclusive or compensatory) may vary in different directives. A fortiori should the reproduction right, harmonised in the same article of the InfoSoc Directive, be given the same meaning.

⁵⁴⁶ CJEU 29 July 2019, *Pelham GmbH and Others / Ralf Hütter and Florian Schneider-Esleben*, C-476/17, EU:C:2019:624, paragraph 30.

⁵⁴⁷ CJEU 3 June 2021, *CV-Online Latvia / Melons SIA*, C-762/19, ECLI:EU:C:2021:434, paragraph 38.

⁵⁴⁸ CJEU 3 June 2021, *CV-Online Latvia / Melons SIA*, C-762/19, ECLI:EU:C:2021:434, paragraph 46.

than taking such interests into account for the application of the exceptions or other restrictions to the exclusive rights).

In the same vein, the European Parliament has recalled, in relation to the copyright-implications of AI technologies, that “any approach must strike the right balance between the need to protect investments of both resources and effort and the need to incentivise creation and sharing”.⁵⁴⁹

Lastly, the argument according to which, the introduction of the TDM exceptions in the DSM dir. implies that the technical acts needed for mining some content are falling under the reproduction right can be challenged. Some legal exceptions do only confirm the boundaries of the rights or the state of the law, and are expressly recognized in the law to increase the security for the market participants and to codify and clarify the law as it applied. Even if the exception can be set aside by an opt-out, this does not entail that all technical copies made in the TDM process are to be considered as reproductions: the opt-out is meant to reserve the right, where the right is applicable, i.e. where a “reproduction” can be found (in particular determining whether the work is recognisable and thus exploited in the course of the TDM process or afterwards). The exception for parody is a good example of an exception that was in practice already recognized and applied by the courts before its official adoption by a legislator: the courts indeed had always to balance copyright with freedom of expression. Therefore, the express recognition of the parody exception did not mean that before this moment true parodies were prohibited by copyright law.

Along this reasoning, a developer or user of an AI solution could argue that the technical copies made of protected cultural content in the course of a TDM process (while training an AI solution) are not sufficient for finding an infringement of the reproduction right. Instead, they could put forward the conflict between the protection of the copyright or related rights and other fundamental rights (such as the freedom of expression or the freedom to conduct a business) and argue that the copy in the input phase does not qualify as a “reproduction” if (i) the work or other subject matter is not recognisable in the output and (ii) to the extent that such copy does not prevent the author/performer from receiving an appropriate reward for the use of their work or performance and the producer from realising a satisfactory return on investment (rec. 10 InfoSoc Dir). Such interpretation would preserve the possibility for the rightholders to control the exploitation of their creations, while bringing the notion of “reproduction” in line with what is commonly understood by this notion.

In this interpretation, any copy that does not amount to an exploitation of the work itself should not be considered an act of reproduction: technical copies in the process of training an AI are arguably not exploiting the work itself but aim at extracting other information would then not be covered.

As far as the database rights are concerned, the developer of the AI solution who uses the content of a database for training the algorithms could require the courts to consider their interests while assessing the infringements of the exclusive rights. Moreover, the database maker will have to demonstrate that the use of the content of their database curbs the

⁵⁴⁹ EP (Committee on legal affairs), Report on intellectual property rights for the development of artificial intelligence technologies, https://www.europarl.europa.eu/doceo/document/A-9-2020-0176_EN.html, (accessed in July 2021), par. 10.

possibility to redeem their investment. This will depend on the business model adopted by the database maker and their possibilities to monetise the database.

Application of the TDM exception

In the hypothesis that the reproduction right is interpreted as a tool to control the true exploitation of the work, performance or recording (as described in the previous section), the application of the TDM exceptions might appear redundant. Should this functional approach of the reproduction right not be fully retained, then the **TDM exceptions would create legal certainty** as to whether the user of AI solutions should acquire the rightholders' consent for processing their works as AI input.

Considering the origin, the wording and the purpose of the TDM exceptions, the **status quo should suffice** to demonstrate that the use of protected works, performances and recordings contained **in training data sets** are exempted, if this use is made for the purpose of analysing the data, extracting insights (such as patterns) and using those insights for ulterior creations, including AI-generated output in the musical, visual, audiovisual and games sectors – to the extent that the AI output does not result in a (partial) reproduction of protected subject matter.

Nevertheless, it might be useful to verify that the transposition of the TDM exceptions in the Member States' national copyright laws respects these contours. The Delphi survey and interviews have demonstrated that the TDM exceptions are a cause for concern and that the different stakeholders and experts understand the scope of the TDM exceptions differently, which entails that a **risk** exists that **different interpretations may also prevail after the national transpositions**.

The discussion on the TDM exceptions offers a good opportunity to clarify the **relationship** between the **different exceptions** that may apply to the same data extraction processes (especially the exceptions for temporary acts of reproduction, for research and for TDM). It is explicitly confirmed that the TDM exceptions may apply in addition to the exception for temporary act of reproduction in art. 5(1) InfoSoc Dir. (see rec. 9 DMS dir.) but this is not stated as a general principle for other exceptions. There is no reason to deny the simultaneous application of several exceptions, provided that the conditions of each are met. Inversely, the circumstance that the conditions of one exception are not met, does not entail that other exceptions cannot apply. It might be useful to confirm – for the sake of clarity – that several exceptions may apply to the processes implemented by AI tools. This is particularly relevant in case the rightholders have reserved the TDM use by using the opt-out option under art. 4 DSM Dir.: if the conditions of another exception are met (in particular the exceptions for research in art. 5(3)(a) and (n) InfoSoc Dir.), the AI user does not need the rightholders' prior consent.

At some point, it might be useful to make the various exceptions covering research-related reproductions or communications to the public more coherent, within copyright (in particular in art. 5(3)(a) and (n) InfoSoc Dir. and art. 3 DSM Dir.) and between copyright and other legal instruments (in particular the various provisions relating to scientific and historical research in the GDPR). For the massive processing of data sets, which may contain data whose use is restricted under several legal regimes, it should be avoided that different notions of "research" apply and trigger different, incompatible or overly restrictive limitations. A lack of coherence between the relevant provisions permitting "research" would have the de facto effect of harmonising the permitted data processing operation at the level of the strictest provision. The different provisions will indeed be applied to the same data mining process, in order to comply with all legal requirements, the conditions of the strictest exception will prevail.

3.4.1.2. How to exercise and communicate the TDM opt-out?

The possibility to reserve the exclusive rights in relation to text and data mining process exempted under art. 4 DSM dir. raises some legal and practical challenges, , Several policy options addressing these challenges have been discussed with experts and stakeholders and will be framed in a broader copyright discussion.

3.4.1.2.1. Legal uncertainty on the communication of the opt-out and alternative options

The training of AI solutions in the creative sector often requires relying upon data, the use of which might be restricted by copyright or related rights. Therefore, an AI developer will legally need to identify these rights and their owners carefully to obtain, when necessary, their clearance. This task might be burdensome (even prohibitive) due to the plurality of rights and rights-owners concerned for a single training data set (high transaction costs).

The TDM exceptions could solve these issues. However, the TDM exception for other purposes than research offers the rights owners the opportunity to opt out. The conditions for and the management of these opting-out decisions could, in practice, also be challenging both for rightholders and for users, especially since the requirements for exercising and communicating an opt-out decision, according to Art. 4(3) DSM dir., appear to lack sufficient clarity (supra 3.3.6.2).

The participants to the Delphi study were asked about their opinion on the conditions of exercise and the standardisation of opt-out decisions and the uncertainties that follow from this possibility to reserve the exclusive rights. In order to analyse the potential use of the opt-out, the following policy scenarios have been considered:

- Status quo: do nothing, no guidelines or best practices given and no centralised register, relying on industry practices/cooperation.
- A clarification of the conditions for exercising and communicating an opt-out decision according to Art. 4(3) DSM dir. through guidelines / best practices
- Create a centralised register of works with opt-out decisions
- Other, please specify.

3.4.1.2.2. Stakeholders' and scholars' opinions

According to 58% of the experts consulted in the Delphi survey, the opt-out is likely to become an extended practice among artists and other stakeholders, while 22% believe it will not be a common practice (see Figure 43).

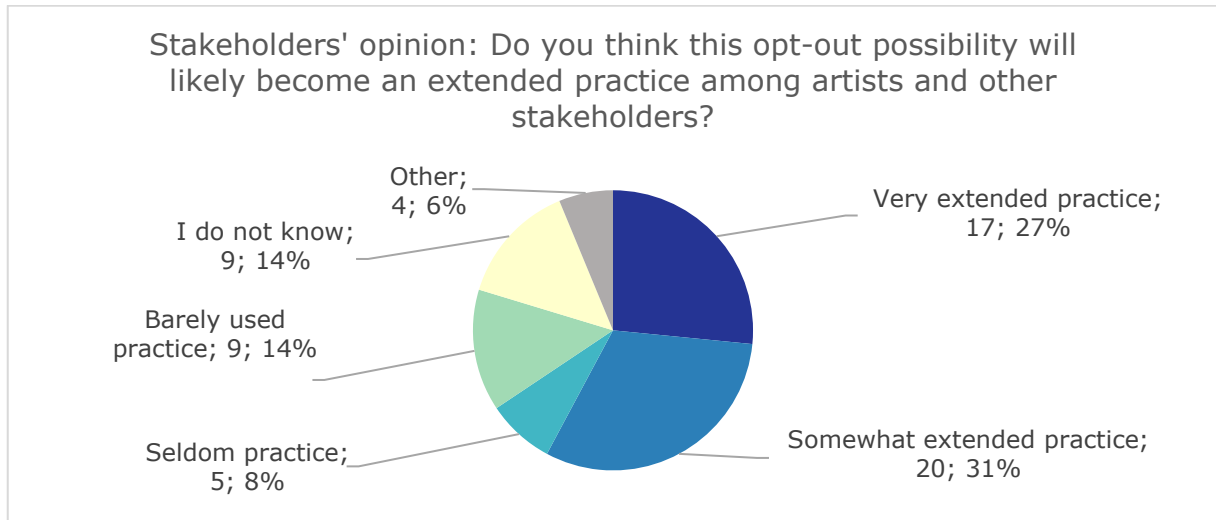


Figure 43: Stakeholders' opinion on whether the opt-out possibility will become an extended practice (N=64)

Source: Technopolis Group Survey

Some scholars participating in the in-depths interviews expect that this practice will remain a dead letter, except for the press industry. According to them, most rightholders are unlikely to use this mechanism because they are not really concerned about the analysis of their work by AI systems. Nevertheless, they believe that rightholders could use this mechanism as a bargaining chip to negotiate a “fair” share in the value generated by the output.

Similarly, the in-depths interviews revealed overall a great deal of legal uncertainty regarding the expression of the opt-out.

Figure 44 presents the opinions of the participating stakeholders about the policy scenarios. As it can be observed, there is a large agreement for clarifying the conditions for the opt-out. Two participants elaborated on this by arguing that guidelines and best practices could be a tool to raise awareness and provide visual authors with an efficient right to opt-out and agree on a licence contract. Furthermore, it was noted that the guidelines and best practices should take into account the specificities for the different sectors. However, another participant argued that even if guidelines/best practices are defined, this does not necessarily lead to their implementation in practice.

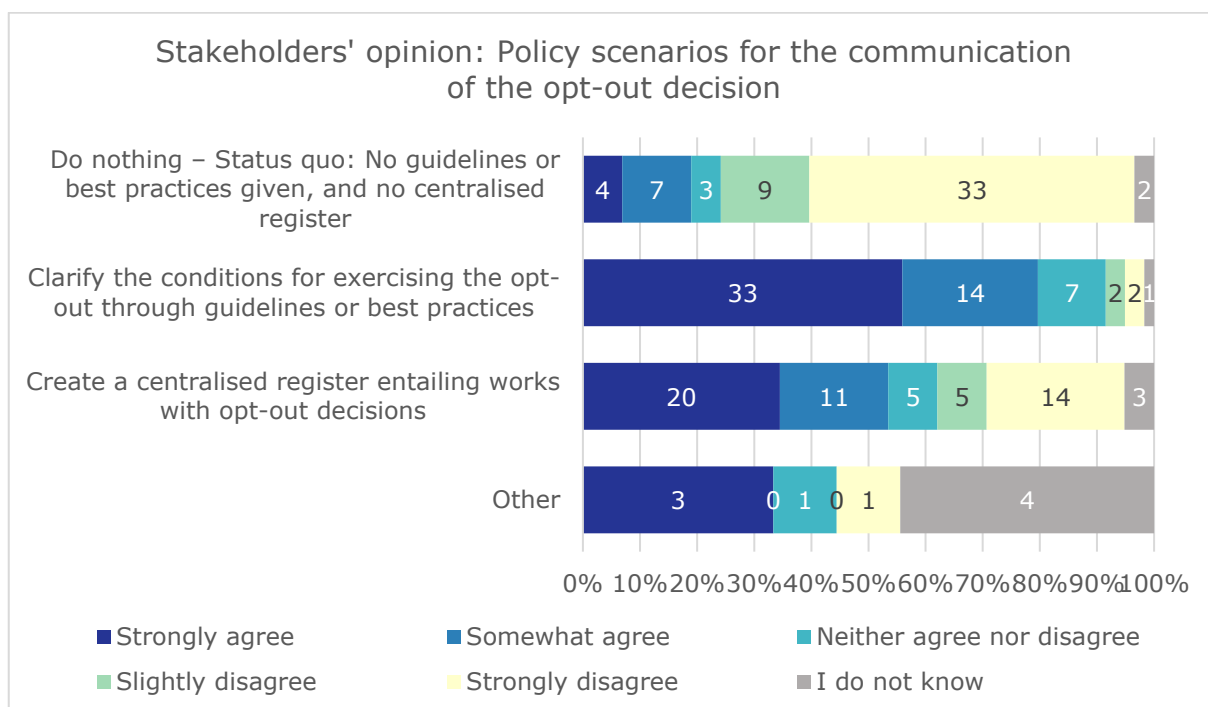


Figure 44: Stakeholders' opinion on scenarios for the communication of the opt-out decision (N=59)

Source: Technopolis Group Survey

As regards the technical means of expressing the opt-out, some scholars participating in the in-depth interviews stressed that TDM exceptions constitute a point of equilibrium between the interests of market players. In their view, imposing **additional requirements** on rightholders could create unjustified asymmetries (e.g. respecting a particular format to express the opt-out, defined by the user). Other scholars felt that minimal efforts will be necessary on the part of rightholders, e.g. the **implementation of machine-readable solutions or a contractual notice**. Two experts in the music sector suggested the opt-out could be registered with CMOs responsible for administering music creators right. The establishment of a **centralised, digital opt-out register** was also mentioned as a potential and appropriate solution. Almost half of the stakeholders who participated in the Delphi survey (48%; 31) strongly or somewhat agree to the creation of a centralised register containing the works with opt-out decisions. Of them, 15 are legal practitioners/academics, 7 artists/rightholders and 8 representatives of CMOs, 3 trade associations and 1 publishing company.⁵⁵⁰ Some of them argued that this would be the best option, as it would ensure legal certainty for those that want to use the protected content as well as for the rightholders. On the operational side, an IP expert in the publishing sector suggested the use of existing agencies managing standards such as the International Standard Text Code or Editeur.

Interestingly, several participants disagreed with the creation of a centralised register, for a variety of (very specific) reasons. It was argued that (i) the creation of a dedicated database would be time-consuming and expensive; (ii) it would be impracticable; (iii) it

⁵⁵⁰ This categorisation of respondents is based on a multiple-choice question in the Delphi survey. The categories are not mutually exclusive.

would be incomplete and quickly out of date; (iv) it would undermine the essential elements of the copyright system by introducing an additional hurdle for rightholders to protect their works; (v) that performing an opt-out should not be dependent on a registration obligation by the rightholders involved; (vi) it would not facilitate that rightholder chooses whether or not to opt-out depending on the medium/website on which the work(s) are made available; (vii) it would inevitably and unintentionally be modelled on certain preconceptions about licensing practices of one or two sectors, therefore, it would not be practical for each creative/cultural sector.

In these open answers, survey participants argued that **machine-readable solutions** would be more practical than a centralised register, as they are scalable and adaptable to the needs of each sector. In their opinion, such solutions should be rightholder-friendly, publicly accessible, interoperable, as well as open to change and discussion between owners and AI agents. In the opinion of these participants, the opt-out could be communicated in different ways. For instance, the opt-out could be expressed in the terms of use of a service, which would be readable for humans but complicated for automated processing. Such automated processing would be easier if the opt-out were indicated in the metadata of the works or based on web standard communication protocols. Stakeholders pointed that perhaps several strategies could be used simultaneously in order to effectively communicate the opt-out, as it might need to be stated in different websites or platforms managed by different stakeholders (e.g. rightholders' website, online platforms where the rightholder has an account, websites of stakeholders' associations, websites of CMOs, etc.).

According to an IP expert representing an organisation in the audiovisual sector that uses AI to generate creative output and that also develops AI, the implementation of an opt-out system that efficiently fulfils all needs of the creative and cultural sectors for rightholders and users seems unachievable. In their opinion, it is expected that a significant number of rightholders will opt-out; for a developer/user of AI, it will be difficult to verify with certainty that a rightholder has not opted out (e.g. because the information is provided in an obscure location). Therefore, AI developers/users might not feel there is sufficient certainty to rely on the exception. This constitutes a problem because vast quantities of data are required in order to train AI algorithms and it would be incredibly time-consuming for users to search for opt-out information on a work by work basis.

For these reasons, a few Delphi survey participants noted that it is likely that each creative/cultural sector will develop its own strategy or standard to declare the opt-out notice.

This contrasts with the views of a multinational technology company which participated to the in-depth interviews. They reported that existing instruments are not designed to serve this opt-out possibility and that, in any event, new mechanisms yet to be developed will have to be deployed and used universally.

Regarding the suitability of different implementations of the opt-out, participants were asked to suggest solutions in the first round of the survey and they assessed them in the second round. Their opinions are shown in the Figure below. **The most favoured solution is the creation of a protocol for machine-readable statements** that can be used by the stakeholders to express and read the opt-out decision (58%; 14 participants, among which 6 are legal academics/practitioners, 2 individual artists/rightholders, 2 CMOs, 2 producers, and 1 trade association). Solutions that require **manual intervention** by the author/rightholder (e.g. in a written letter, direct communication/request) were considered as least adequate. The creation of **registers** of works with an opt-out at the EU or national level was regarded as suitable by 29% and 45% of the participants, respectively. However,

the registration of the opt-out was considered more suitable when done with organisations responsible for administering creators right (e.g., CMOs in the case of music) or through agencies managing standards (e.g., istc-international) and creating free, public, and accessible online platforms where third parties can check if a rightholder authorises the use of works for training AI. It should also be noted that 41% of the participants also regarded a **“no formalities, no registration” approach** as suitable , implying that the parties will find some agreements based on best practices and common standards through self-regulation.

Stakeholders' opinion: Suitability of different implementations of the opt-out

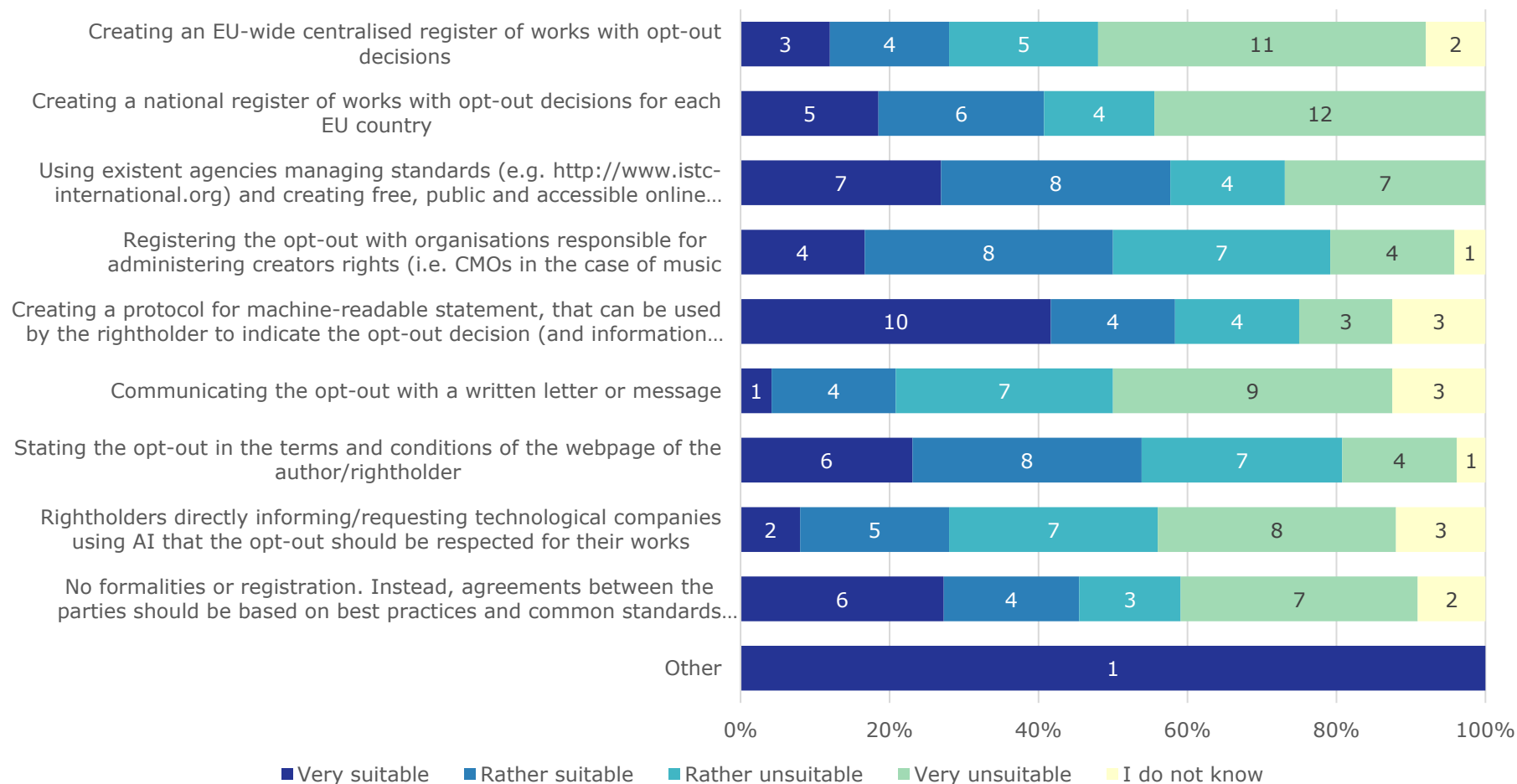


Figure 45: Stakeholders' opinion on the suitability of different implementation of the opt-out (N=24)

Source: Technopolis Group Survey

As can be observed in the following figure, among the scholars consulted at the ECS conference, the solution of implementing a standardised protocol is also regarded as the most adequate, followed by the opinion that the opt-out should be registered by CMOs.

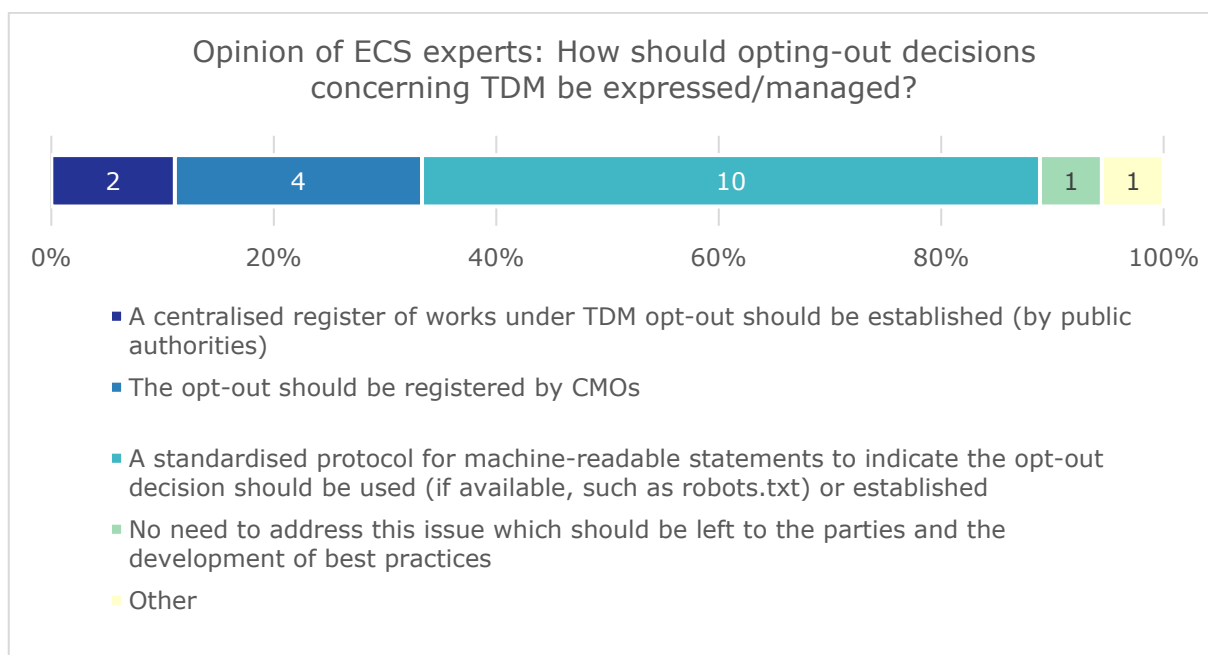


Figure 46: Opinion of ECS experts on the expression of the opt-out decision (Total votes = 17)

Source: Technopolis Group Survey

Some interviewees raised the issue that the rightholders are not informed about the TDM operations exercised on their protected subject matter. This, in their opinion, obscures the exercise and effectivity of the opt-out (e.g. whether the communication opt-out has been seen and respected).

3.4.1.2.3. Discussion

Most of the consulted stakeholders (predominantly legal practitioners/academics, followed by CMOs and artists/rightholders) expect the reservation of rights to become a “somewhat extended” or “very extended” practice, but at this stage it is not yet clear how the opt-out will function in practice. It will be important to monitor the application in practice of the TDM exceptions before assessing the need for further intervention.

An efficient opt-out system should be straightforward to use for rightholders and AI developers/users alike. The rightholders should be able to exercise the opt-out with regard to the content and for the rights they can exercise on that content (or authorise another person to do so). AI developers and users need access to clear, reliable and non-contradictory information on the exercise of the opt-out so that they know, with a reasonable degree of certainty, that they are not infringing any exclusive rights when using a particular data set for training an AI system.

Each of the propositions suggested by the consulted stakeholders presents difficulties, either in specific sectors or digital contexts/use cases. Much will depend on the **access to the data**. Some databases (such as film catalogues) contain a large quantity of data and metadata but access is restricted and controlled by the organisation exploiting that content, which then gets a headstart if it engages in the development of AI solutions. It can be expected that the use of the protected content for training AI is discussed with the rightholders and that no doubt exists regarding the opt-out. Developers of AI solutions who use data from other, multiple, publicly accessible sources will have more difficulty establishing with certainty whether they are entitled to use the protected content as AI input or if the various rightholders have reserved this right. Furthermore, the rights in relation to TDM of the same protected content may be exercised in a fragmented way, along the lines of the national territories of the Member States.

Two aspects may be distinguished: the exercise of the opt-out per se and the communication of such decision to the stakeholders. Any system should make sure, firstly, that the opt-out is exercised by the actual holders of the applicable rights and, secondly, that the decision is communicated in a non-equivocal, non-contradictory and reliable way.

Different *modi operandi* of distributing and acquiring creative content (i.e., mechanisms, platforms, intermediaries, revenue streams) are used from **sector to sector**, and are unlikely to be homogenised. Consequently, stakeholders from different sectors have contrasting suggestions on how the opt-out should be operationalised. For instance, a participant from the publishing sector proposed to entrust agencies managing standards with (the standards for) the opt-outs, while a participant in the music sector suggested a registry with the collective management system.

Moreover, even the **interest** of rightholders to opt-out might differ from one industry to another, as discussed in a study conducted in France: the press sector would be the most interested in making use of the opt-out, whereas literary publishers, rightholders of the music, image and audio-visual sectors appear less concerned.⁵⁵¹ It can also be expected that data traders may opt-out of the TDM exception by virtue of (rightly or wrongly) claimed database rights.

Each sector might consequently develop their own way to communicate the opt-out, and even in case a general opt-out mechanism for all sectors is implemented, variations may be necessary in order to cater for the peculiarities of each sector.

However, the difference between creative sectors is not the only obstacle in finding an adequate manner to implement the opt-out system in a way that is efficient and manageable for both rightholders and for AI developers. The diversity of **digital contexts** for distributing/accessing works adds another layer of difficulty. Examples of such digital contexts are as varied as the webpage of an author, online selling platforms where the rightholder has an account, digital libraries of creative works, websites of stakeholders' organisations or websites of CMOs. An adequate opt-out in one digital context might not be useful in a different one. An opt-out statement in the **terms of use of a website** could, for instance, be efficient when all the creative content of the website is subject to the opt-out following the same terms (e.g. all the creative pieces of the same artist). Such statement may not correspond to the rightholder's intentions when such website contains creative pieces of different artists, who may prefer to express diverging opt-out

⁵⁵¹ A. BENSAMOUN and Y. BOUQUEUREL, Transposition des exceptions de fouille de textes et de données: enjeux et propositions, *Ministère de la culture*, 2020.

statements. It would then be for the administrator of that website to “clear” the opt-out rights for TDM purposes and to integrate the rightholders’ intentions in the terms of use.

An opt-out notification at the level of a **digital file** (e.g., indicated in the metadata of the file) would offer more flexibility. However, if the creative piece is stored by a third party as part of a collection of works (e.g. a selling platform for creative content or a digital library of creative works), the metadata might reflect the choices of the administrator of the collection of works, rather than the specific choices of the rightholder. Also the same work, performance or recording might be included in different digital files with contradictory information, leaving the AI developer without certain information on the opt-out.

The opt-out notification may get lost when the creative piece is **further distributed** by third parties (see also part 1 of this study on copyright and data management, subsection 2.3.1.1 where we refer to absence of metadata and potential reasons for this, among them so-called data-stripping, i.e. the removal of metadata). In the absence of registers where opt-out notifications can be consulted, AI producers would need to establish direct contact with the rightholder in order to have certainty on the permitted use, which may be an insurmountable obstacle (given the quantity of data required to train AI systems).

A **central registry** of works and other protected objects (performances, recordings) could be an alternative that would provide a certain level of certainty to rightholders and AI developers – even in cases where the metadata of the creative piece has been lost. However, such a registry would need to be free, public, easy to use and accessible to robots (e.g. via an API). It should be maintained in a proper and reliable way as a (quasi-) authoritative source, to make sure that the opt-outs duly reflect the rightholders’ intentions. Despite this, some of the consulted stakeholders warned that it could still be considered an impracticable formality that introduces an additional hurdle for rightholders to protect their works.

For the communication of the opt-out, an **interoperable standard** would offer the most effective solution, such as a protocol for machine-readable statements. Such standards would offer flexibility in that they could efficiently be read and processed by AI robots and be adapted for every sector. Such standard would streamline the communication of the TDM rights for rightholders, but also reduce the risks of legal uncertainty for AI developers/users.

At present, no technical standards are set, the discussion being a technical and a complex one. There are efforts to design such a machine-readable solution, i.e. a W3C commission called “Text and Data Mining Reservation Protocol Community Group”.⁵⁵²

⁵⁵² At the moment, this group is assessing alternative technical solutions for expressing the reservation of TDM rights, such as one based on HTTP headers, another based on a file hosted on the origin server and a third based on HTML metatags. The group has also discussed how to define a machine-readable TDM policy that details how a rightholder can be contacted and conditions in which a TDM licence can be acquired, when rightholders opt out. See W3C Community group, <https://www.w3.org/community/tdmrep/>, (accessed in July 2021).

One solution for this would be to have TDM policies defined as a profile of ODRL 2 (Open Digital Rights Language)⁵⁵², which is “a policy expression language and model that provides a flexible and interoperable information model, vocabulary and encoding mechanisms for representing statements about the usage of content and services.” In other words, it is a standardised way to express in a machine-readable format the permitted and prohibited actions over digital resources. See: W3C, <https://www.w3.org/TR/odrl-model/>, (accessed in July 2021).

There are already tools built upon ODRL to operationalise its implementation on websites distributing creative works.⁵⁵³ This effort is in its early stage, and no evidence of its implementation by the creative sector could be found. It remains to be seen whether this solution would be implemented both in the particular websites of individual authors as well as in platforms storing the creative works of several rightholders, and whether it would be flexible enough to be applied to all creative sectors. Therefore, it is unclear whether the implementation of one approach to communicating the opt-out would suffice (i.e. only the implementation of the machine-readable protocol), or whether a combination of more than one of different approaches would be needed (e.g. a machine-readable protocol and a centralised register).

Whichever technical solution is adopted, the challenge will always be to verify that the person expressing the opt-out actually executes the decision of the different rightholders (who may be different persons exercising cumulative rights to the same content in the dataset) and to avoid contradictions at that level.

3.4.1.3. Can moral rights be exercised to block the use of works as AI inputs?

Not only the economic rights but also the moral rights may be relevant for the use of protected works or performances for training AI solutions. Several policy scenarios are presented to address this issue, followed by the response of experts and stakeholders. The options will then be discussed in a broader copyright framework.

3.4.1.3.1. Issue and policy scenarios under consideration

As explained in section 3.3.5.2, moral rights are not harmonised at the EU level. This results in potentially diverging situations with regard to the possibility for rightholders to oppose the use of their works or performances as AI input.

During the interviews and the Delphi study, the following policy scenarios were investigated:

- **Status quo.** The moral rights remain regulated at the Member states' level, allowing each of them to follow its own legal tradition, but with the result that potentially diverging solutions emerge in the internal market.
- **Legal clarification that moral rights can be invoked against the processing of the work or performance for AI training.** This scenario reflects the most far-reaching moral right form of protection. This subjective approach to copyright protects the rightholders without the latter being required to demonstrate that the use of their work or performance is objectively prejudicial to their honour or reputation.
- **Legal clarification that moral rights cannot be invoked against the processing of the work for AI training if TDM exception is allowed.** Under this scenario, the scope of moral rights is partially harmonised and limited to allow its alignment with the existing exceptions for economic rights to avoid a circumvention of the policy trade-off operated by the EU legislators at the level of the economic rights.

⁵⁵³ An example is RightsML by IPTC: <https://iptc.org/standards/rightsml/>, (accessed in July 2021). This tool extends the ODRL framework to meet the specific needs of the media industry.

- **Legal clarification that moral rights cannot be invoked against the processing of the work for AI training, if the work or performance is not recognisable in the output.** This scenario proposes to restrict the scope of moral rights by adopting a more objective approach and requiring at least the protected subject matter to be recognisable in the output to trigger the application of moral rights.

3.4.1.3.2. Stakeholders' and scholars' opinions

Although there seems to be a consensus among the stakeholders and academics on the need to solve this issue, positions are much more divergent when it comes to the solution itself.

Hence, the experts participating in the Delphi survey showed a high level of consensus on the policy scenario that they considered to be more adequate (see Figure 47): a clarification that rightholders can oppose the processing of the work for AI training based on the moral rights (67% strongly or somewhat agreed). For some, such a clarification would be needed because the usage of creative works for AI purposes is novel for the rightholders, and it would establish clarity on whether or not the TDM exception concerns only the relevant economic rights, leaving the moral rights intact.

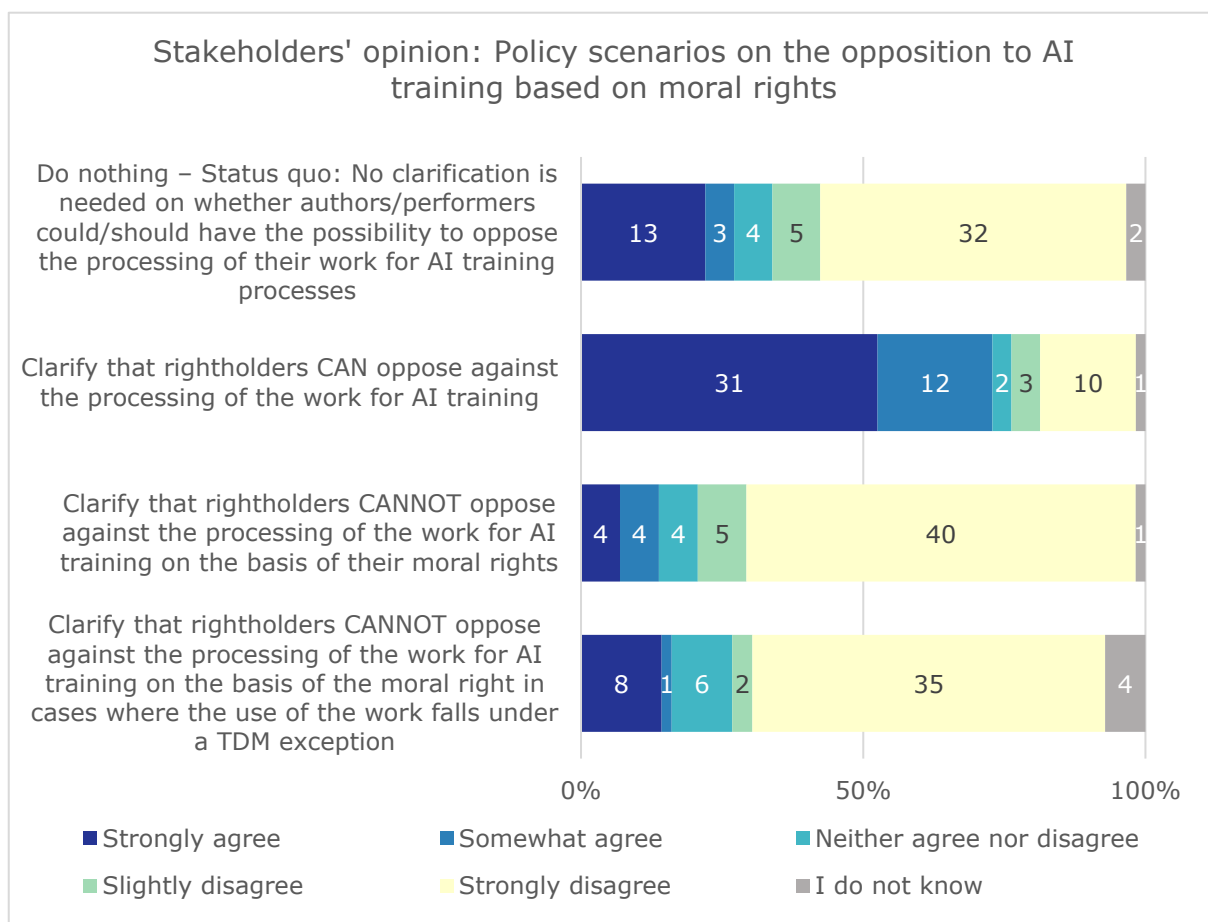


Figure 47: Stakeholders' opinion on scenarios on the opposition to AI training based on moral rights (N=59)

Source: Technopolis Group Survey

In contrast, the vast majority of ECS legal experts are in favour of limiting the exercise of the moral rights to the cases where the performance/work is recognisable at the output stage and its usage is prejudicial for the rightholder. This position also reflects the outcome of the majority opinion in the interviews conducted in the framework of this study.

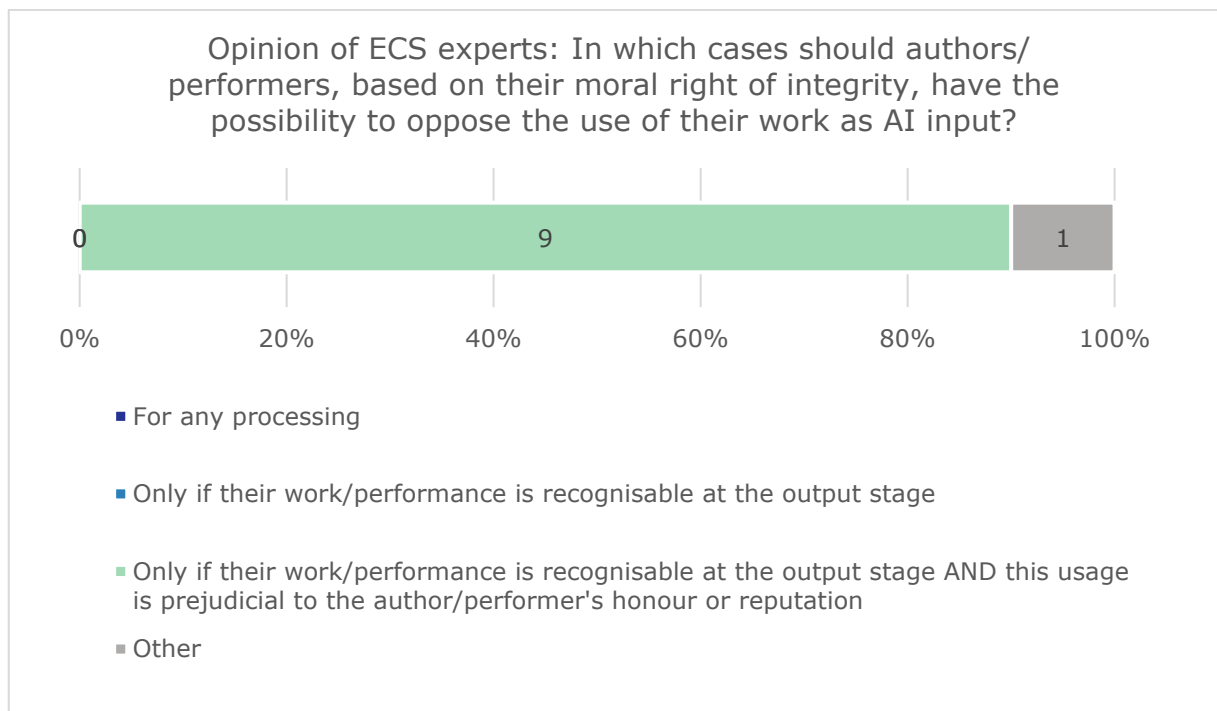


Figure 48: Opinion of ECS experts on the conditions to opposed AI training based on moral rights (Total votes = 10)

Source: Technopolis Group Survey

One IP expert specialised in IT law who participated in the Delphi survey argued that authors should have the right to oppose the processing of their work for AI training irrespective of whether or not it is based on moral rights or economic rights. Two participants believed that the applicability of moral rights should be evaluated on a case-by-case basis, where aspects such as the purpose of the AI training, the commercialisation of the AI output, among others, would need to be analysed. For instance, one of them explained that the moral right of integrity is more likely to apply when an AI was used to mimic or slightly modify an artistic work, and less likely to apply when AI was used, for example, to make the work available in a higher resolution. Another example referred to writers agreeing to the use of their texts to train AI systems that performed spelling and grammar checks or improved style, but feeling that their moral rights were violated when the AI was trained to make political speeches more emotionally powerful.

Among those participants who thought that there should be a clarification that rightholders could not oppose AI training based on moral rights, one argued that copyright exceptions also applied to moral rights, and that national implementations of the TDM exception would

clarify the extent to which the authors were allowed to use their moral rights to limit the use of their copyright-protected works for training AI.

A stakeholder with IP expertise, who indicated no strong agreement nor disagreement with any of the policy scenarios, expressed doubts on whether the use of a work to train AI would have any impact on the public perception of the work or the author. In their opinion, given that moral rights appear aimed at public perception of the work and author, it should be considered whether the AI output has any impact on them. Thus, for cases without impact on public perception, it should be clarified that moral rights cannot be used to prevent AI training. A similar opinion is shared by two other participants who argue that the moral rights are not affected, in the case the output of the AI is not a creative piece.

Furthermore, five participants stated that there was no need for the harmonisation of moral rights at EU level, whereas one argued that the use of moral rights in relation to AI should be fully harmonised, as different national rules fragment the digital single market area.

Although moral rights are not based on economic motives, exercising them may have an economic impact on stakeholders. Therefore, those participating in the survey were asked to assess the economic impact of three scenarios: one in which authors/rightholders can oppose every use of their works to train AI based on their moral rights, another in which they can oppose when the work is visible/recognisable in the output of the AI, and one in which they can oppose when the work is visible/recognisable and prejudicial to the author's honour or reputation. As it can be observed in Figure 49, the scenario with the most favourable impact on the revenue of authors/rightholders is that where they can oppose every use of their work to train AI on the grounds of moral rights. It is not surprising that this option is considered less positive for producers of AI, as well as for the attractiveness of the EU for businesses developing AI in the creative sector.

One stakeholder added that the rightholders should have the right to be informed about the processing of their work for training an AI system. Otherwise, there is the risk that, by the time the author/rightholder becomes aware that their work has been used to train an AI system, the work has already been widely used. In the opinion of this stakeholder, timely opposition to the use of the work (i.e. before it is used to train AI) would prevent very negative consequences, particularly because of the difficulty of tracking which AI systems have used the model trained with certain works, and which output has been generated using such trained models. In contrast, another participant argued that it was impossible to notify all the copyright owners of every possible AI output that had been generated using a model trained with their works and could be considered prejudicial to their honour.

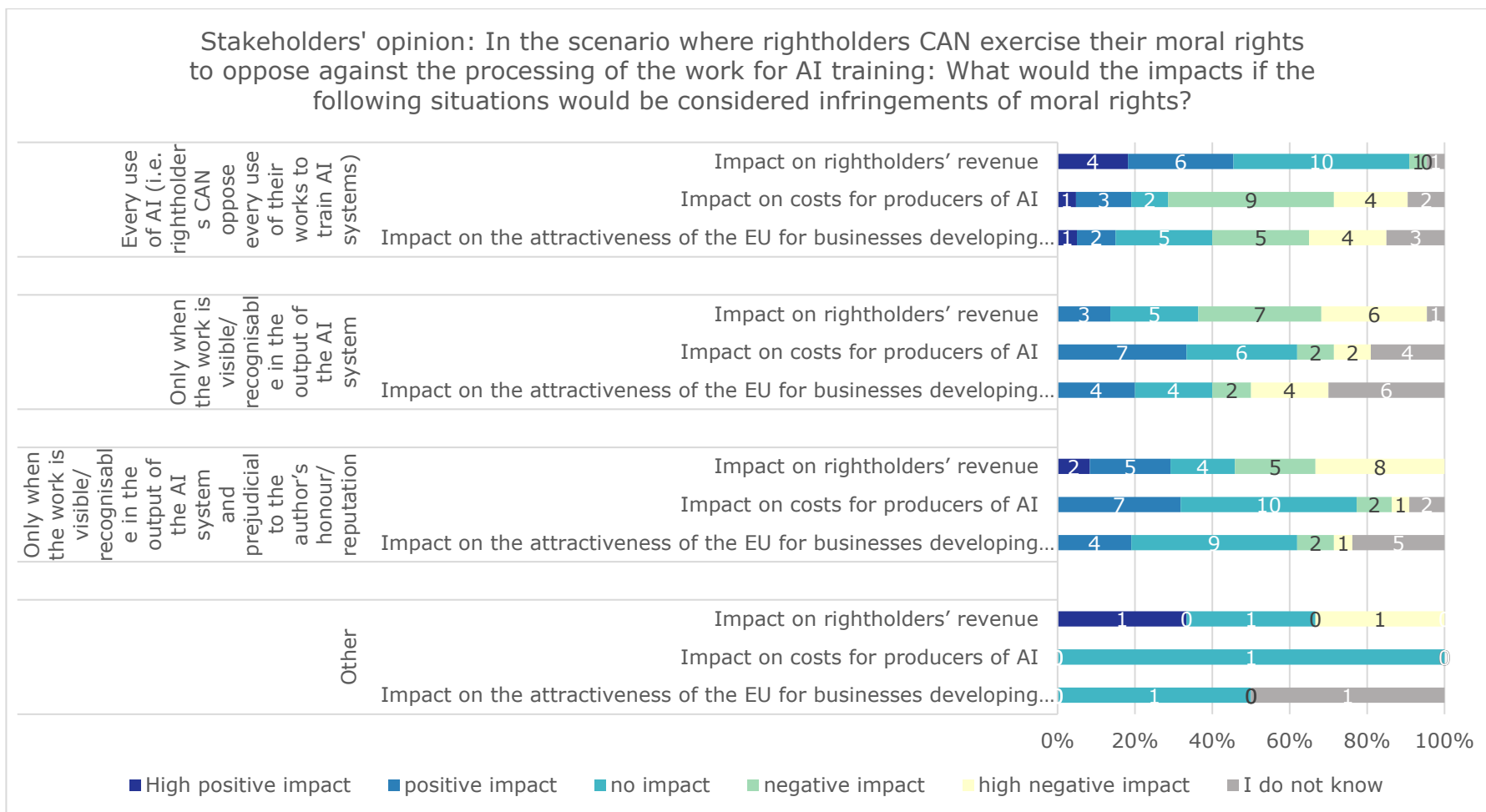


Figure 49: Stakeholders' opinion on the impact of scenarios of the exercise of moral rights

Source: Technopolis Group

3.4.1.3.3. Discussion

The status quo concerning moral rights appears to be **an unsatisfactory scenario in the context of the use of works as input to AI**. Although a study commissioned by the European Commission on *moral rights in the context of the exploitation of works through digital technology* at the end of the 1990s concluded, on the basis of a consultation with interested parties, that harmonisation of moral rights should not be put on the agenda of the Commission⁵⁵⁴, this study shows that the **position of stakeholders might have evolved**, when asked in the context of AI. Regarding the need for harmonisation, the opinion of the stakeholders consulted now seems to be in line with the view expressed by academia. Examined in relation to AI-solutions, harmonising the economic rights while leaving moral rights aside might increasingly be seen as an incongruity. Apart from the fact that it could lead to a fragmentation of the internal market, the exclusion of these rights from the *acquis communautaire* could lead to contradictions, for instance, if the moral rights are used to circumvent limitations foreseen for the economic rights. In this regard, even if the EU legislator, until now, refrained from harmonising moral rights, nothing in the treaties appears to stand in the way of such an action. Hence Art. 114 TFUE generally grants the competency of the European legislator to “adopt the measures (...) which have as their object the establishment and functioning of the internal market”, without a derogating rule for the harmonisation of moral rights. It is therefore important to consider whether a **clarification and a harmonisation of the conditions** under which a rightholder should - based on their integrity right - have the possibility to oppose the use of their work or performance as an AI input is needed. Several options could be explored by the legislators.

At a lower level, the introduction of **an exception** for moral rights at the EU level is a possible option.⁵⁵⁵ The result would be that, if an exception applies to economic rights, it also applies to the moral rights. It would at least bring the protection of the moral rights in line with the exceptions for economic rights. Hence, when TDM practices are allowed, the rightholders should not have the possibility (by exercising their moral right) to circumvent this policy trade-off made by the legislators. Furthermore, this option offers the advantage that the legislators do not have to engage in an in-depth harmonisation of the different national legal orders.

⁵⁵⁴ The outcome of this consultation and the responses sent to representative organisations was summarized as follows: “The analysis of the answers to the questionnaire sent to some representative organisations ... indicates that most interested parties are very cautious about any initiative which would be taken in order to harmonise the level protection of moral rights in Europe. This attitude seems motivated by the fear that a compromise on the moral right issue, if reached at Community level, would jeopardise the high level of protection already enjoyed in some countries. Indeed, the harmonization process could end up for instance with the adoption of waivable moral rights. Different rights holders organisations have pointed out that energy should preferably be devoted to other, more important problems” (see M. SALOKANNEL, A. STROWEL and E. DERCLAYE, *Study contract concerning moral rights in the context of the exploitation of works through digital technology. Final Report*, 1999, Nr ETD/99/B5-3000/E°28). Among others, the following groups responded: Pyramide, European Federation of Journalists, European Federation of Newspaper Publishers, European Federation of Publishers, AIDAA /SACD, IFPI, FIA, FIM, European Writers’ Congress, European Visual Artists, GESAC, Swedish Association of Journalists, Finnish Association of Journalists, Directors’ and Producers’ Society DPRS.

⁵⁵⁵ Comp. J. Drexel et al., “Artificial Intelligence and Intellectual Property Law Position Statement of the Max Planck Institute for Innovation and Competition of 9 April 2021 on the Current Debate”, *Max Planck Institute for Innovation and Competition Research Paper*, Nr 21-10, 2021 p. 12: “In this context, the right to object to any change, disfigurement, mutilation or other impairment of the work might become particularly relevant, and it might necessitate the introduction of an exception or limitation at EU level”.

A more targeted approach could be to specify that, to safeguard the application of certain harmonised exceptions (such as the TDM-exceptions or the exception for temporary acts of reproduction), the moral rights can only be exercised to oppose the use of a work or a performance if it is perceptible to the public. Such exception could make the link with the protection of the reputation of the author, as provided in the Berne Convention.

The legislators could also consider a **more ambitious harmonisation**. A holistic copyright framework indeed requires taking into account moral rights. Rather than maintaining a piecemeal approach, harmonisation at the EU level of moral rights could therefore be considered. Such a solution would require the legislators to decide which approach (objective/subjective) should prevail. If legal scholars seem to agree on the former, the opinion of the stakeholders is much more divided.

3.4.2. Considering the copyright status of output generated by artificial intelligence

This section examines how copyright law affects the production of AI output. It will be examined how various stakeholders see the protection of AI output under copyright, under performer's rights and related rights, how the legal presumption of authorship might face challenges with regard to AI output and whether additional information obligations should be adopted. For each matter, the issue will be briefly set out, together with the policy options, the opinion of the Delphi participants and interviewees and, finally, considerations framing the issue in the copyright context.

3.4.2.1. Should the artificial intelligence autonomous output be protected under copyright?

AI output can be very similar to human-created cultural content, which raises a number of challenges. Several policy scenarios are presented, based on the opinions of experts and stakeholders before embarking in the discussion of the options in a broader context.

3.4.2.1.1. Issue and policy scenarios under consideration

It was explained in section 3.3.2.1 that AI autonomously generated outputs cannot benefit from copyright protection because of the absence of human creative choices. This outcome may be called into question for negatively discriminating against AI cultural outputs.

During the interviews and the Delphi study, the following policy scenarios were investigated:

- **Status quo.** AI autonomously-generated outputs remain unprotected.
- **Creation of a *sui-generis* right for AI autonomously-generated outputs.** Under this scenario, a new *sui-generis* right is introduced over AI cultural outputs resulting from certain investments. In such a case, the right owner and the intensity of the protection should be defined.
- **Amendment of copyright law to allow copyright protection for AI autonomously-generated outputs.** Under this scenario, the conditions of the copyright subject matter are modified to enable the encompassment of cultural outputs not resulting from human creative choices.

3.4.2.1.2. Stakeholders' and scholars' opinions

When asked about the policy scenarios, survey participants indicated a low level of consensus in each of the policy scenarios. Furthermore, none of the policy scenarios was significantly more favoured than the other scenarios (see Figure 50).

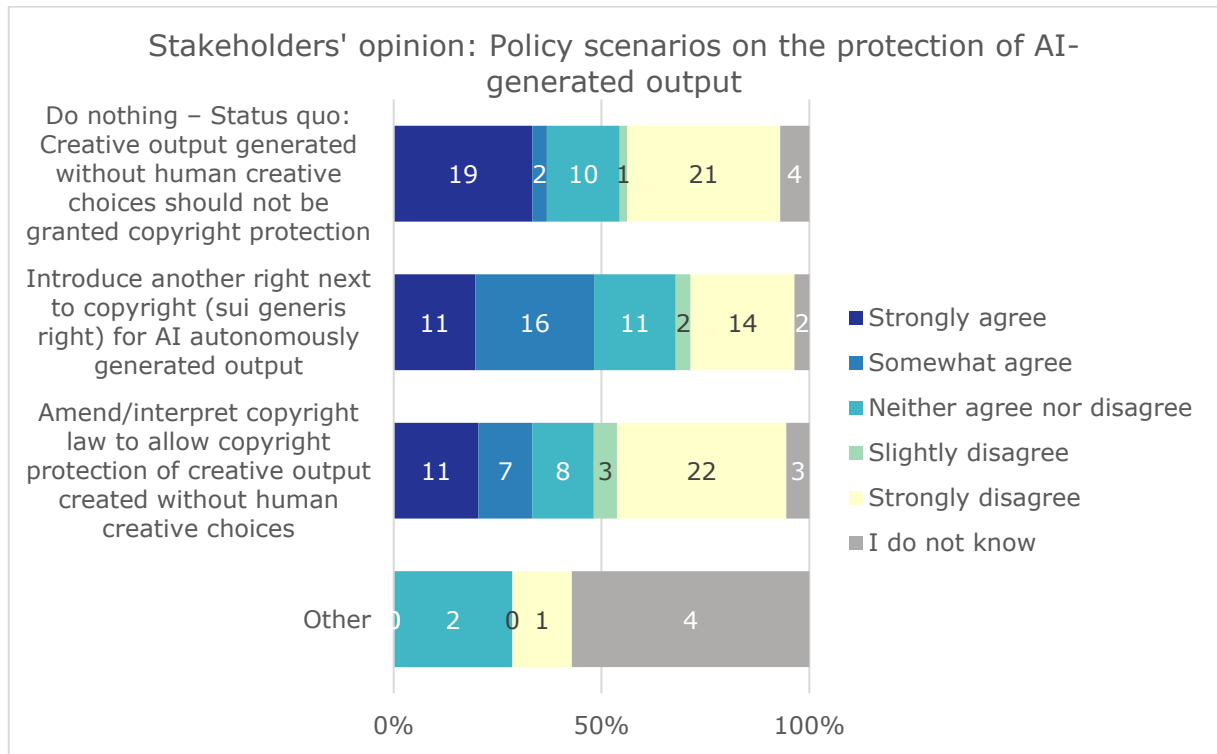


Figure 50: Stakeholders' opinion on scenarios on the protection of AI-generated output (N=57)

Source: Technopolis Group Survey

Those agreeing with maintaining the status quo argued that it was premature for any additional layer of protection for AI-generated output or any legislative amendment. It was also argued that the present copyright, related rights and sui generis database rights framework already offered adequate and effective mechanisms to protect the current use of AI inventions or their generated output. Several respondents stated that, at the present stage, there was not enough evidence on the current and future implications of the use of AI in the creative sector, and thus, careful analyses still needed to be conducted, particularly given the increasingly important role of AI and its potentially significant impact in the value of human creation.

Three participants argued that the protection of intellectual rights was and should be reserved for humans. Along this line of argumentation, two other participants added that it is not clear whether human input into the creation of AI can suffice to meet the test for originality for copyright protection. Their argument is that AI neither produces anything new nor anything of their own, as they are re-using the features of existing originals. One of the participants gave several reasons of how human artistic processes cannot be compared to the algorithmic processes executed by AI; for instance, the role of the psyche, cognitive and emotional abilities, the allusions from culture and experience, the quality

that grows with experience, the change and ageing of the artist, the influence of values and social coexistence.

This position toward the status quo was also shared by the majority of interviewees consulted in this study, including legal scholars, for which copyright law offers adequate protection as is, it is premature to update copyright law to take into account the development in AI.

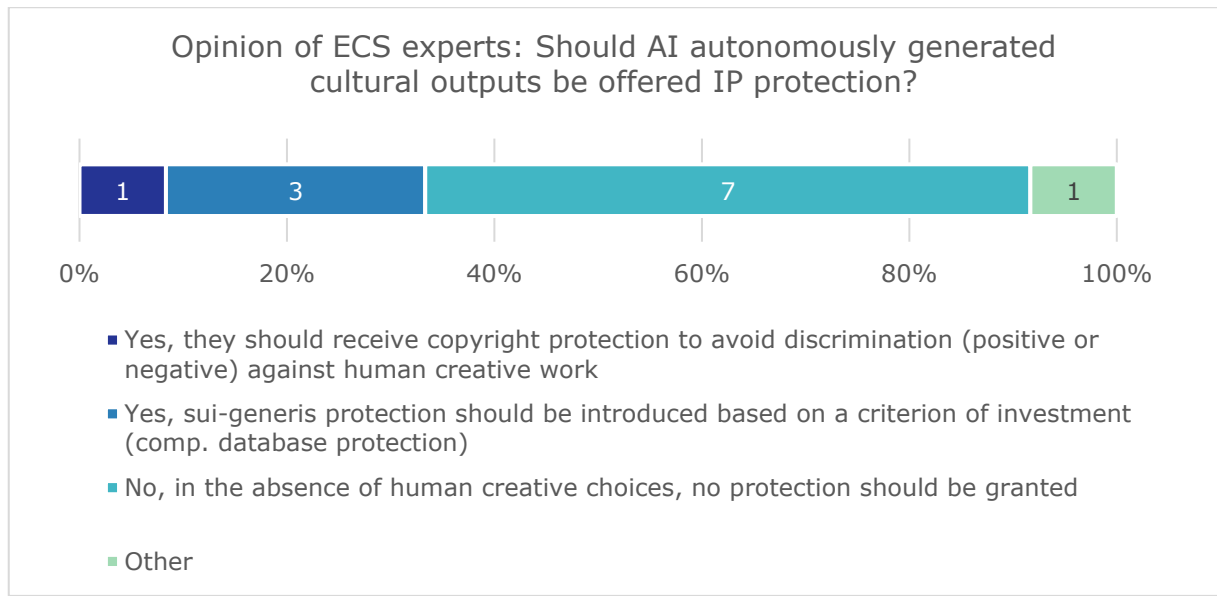


Figure 51: Opinion of ECS on whether AI-generated output should be offered IP protection

Source: Technopolis Group Survey

Those agreeing with the introduction of a sui generis right argued that it could be a means of offering protection to AI output and securing investment in research and development on AI-generated content in the long run. In their opinion, this would also mean that AI-generated output would not enjoy the same protection as human intellectual output, and therefore, this approach would not incentivise the substitution of human-created works with AI-generated output.

A major player in the music industry who was interviewed argued that 'creative' AI output should be protected only if there was a human intervention (which might have already occurred during the selection of the data set to train the algorithm).

Among those who strongly or somewhat agreed with an **amendment of copyright law** to include protection for AI-generated output, three participants suggested that either the producer/developer of the AI would ultimately be the author of the output of the AI system, and the output of their work (the output of their software) should be covered and protected. One participant also argued that protection for AI-generated output should be in place if developers of AI solutions invest in licensing material to train the AI.

Delphi participants were also asked to provide their opinion on the economic consequences of maintaining the status quo and of amending copyright law to allow the protection of AI-generated output. They generally favoured the status quo, as in their opinion granting protection to AI-generated output would be detrimental to authors/rightholders. As it can be observed in the Figures below, the most significant differences between the two scenarios relate to:

- Revenue from human-made works: In the opinion of the stakeholders, with the status quo, consumers will pay less for creative material that can be generated by AI. Furthermore, it is believed that with an amendment that protects AI-generated output, there will be a lower revenue for rightholders of human-made works. A participant explained that a major advantage of AI-generated cultural material is that it is currently IP rights-free, and therefore, it can be offered for commercial use very cheaply. Moreover, where AI succeeds in conducting a creative human task, humans may become secondary, which may lead to job losses. However, a stakeholder argued that AI-generated works are currently generally of inferior quality, and therefore, humans still have a competitive advantage.
- Litigation cases due to similarity: According to the answers of the participants, granting protection to AI-generated output would lead to more litigation costs, not only due to the similarity between human-made and AI-made creative material but also among AI-made creative material.

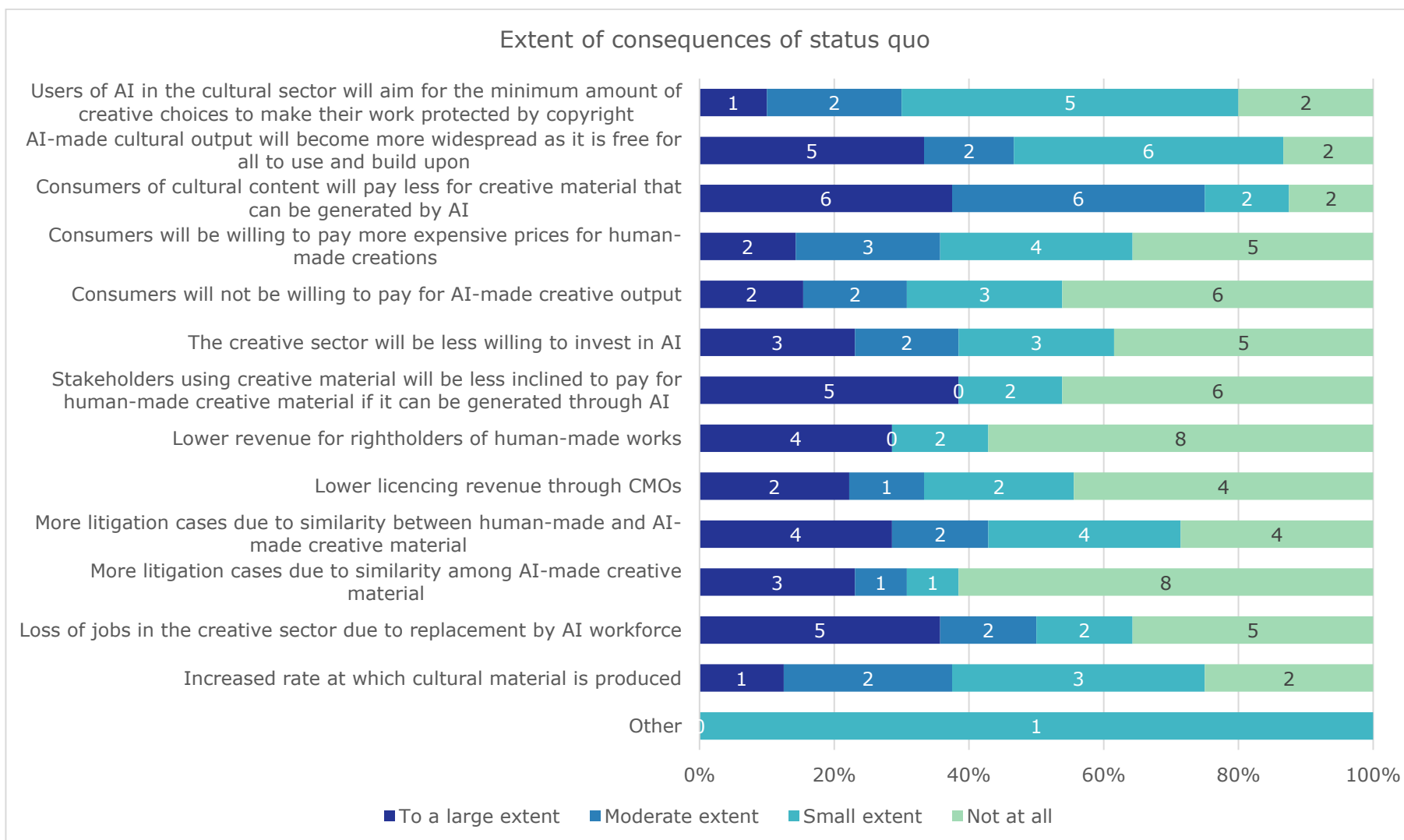


Figure 52: Stakeholders' opinion on the consequences of the status quo of the protection of AI output (N=14)

Source: Technopolis Group Survey

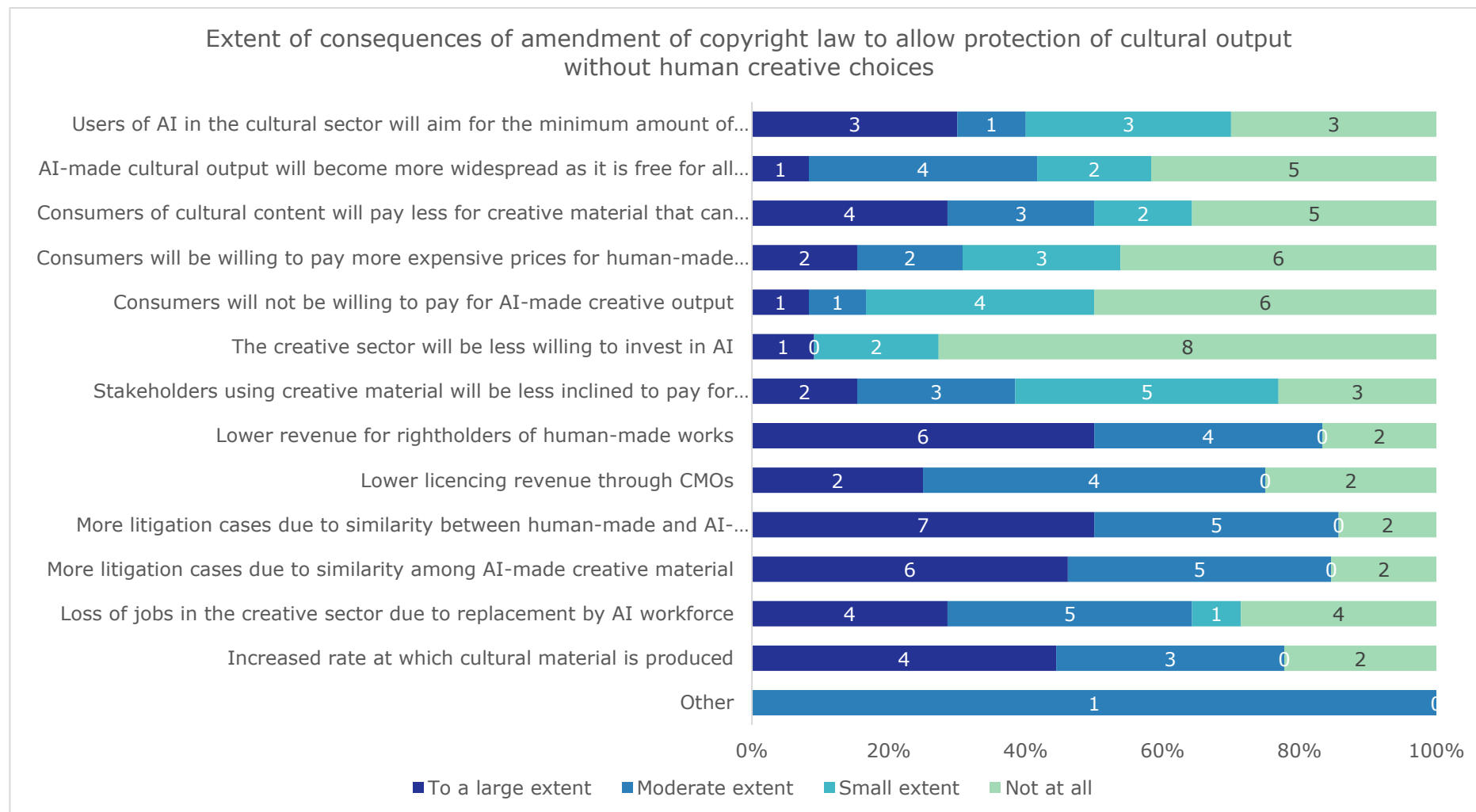


Figure 53: Stakeholders' opinion on the consequences of protecting AI-generated cultural output (N=14)



Source: Technopolis Group Survey

Regarding the consequences of introducing another right (in addition to copyright) for AI-generated output, there was a high level of consensus that this would facilitate the licensing of AI-generated cultural output (see Figure 54). More than half of the participants also believed that it would result in a higher volume of AI-generated cultural output (58%; 12 participants), although not necessarily of a higher quality. In any case, it should be noted that three stakeholders warned that it was too soon to have a clear picture of how AI fits in the creative and cultural sectors, and therefore, it was not possible to determine what a sui generis right for AI-made cultural output would entail.

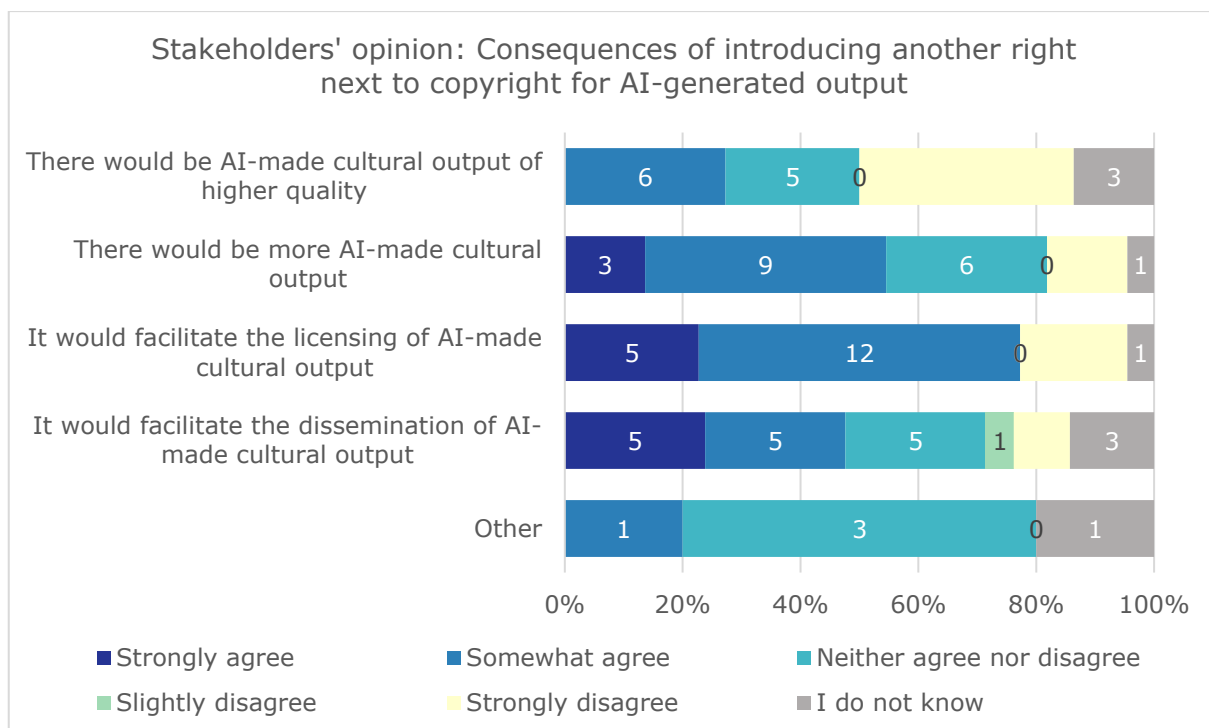


Figure 54: Stakeholders' opinion on the consequences of introducing another right next to copyright for AI-generated output (N=22)

Source: Technopolis Group Survey

If copyright law was amended to allow protection of AI-generated cultural output, or if a sui-generis right were introduced, **the question of the attribution of rights** seems to be open for discussion. Stakeholders participating in the Delphi survey showed low levels of consensus in their answers (see Figure 55): 25% (7) pointed to the user of the AI system; 21% (6) to the producer of the AI system. The developers were also suggested as potential rightholders of the AI-generated output by 11% (3) of the participants, as they are the ones that have designed the behaviour of the AI system. Only one participant (4%) argued that the attribution should be for the AI system itself. Another participant suggested an alternative option, in which the authors on which works the AI-generated output is based should be attributed the rights of the AI output.

Three participants commented that it is difficult to determine as all the above-mentioned parties could make a case for the stake in the product. Two suggested that it should be the party who has the closest link to the creation of the final output, for instance, the person by conducting the arrangements necessary for the creation of the output.

Stakeholders' opinion: If copyright law would be amended to allow protection AI-made cultural output, or a sui-generis right be introduced to whom should the exclusive right be attributed?

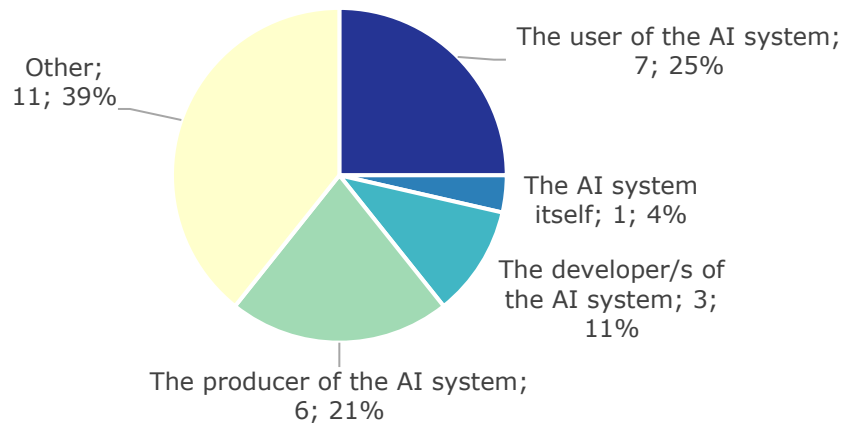


Figure 55: Stakeholders' opinion on the attribution of rights in case AI-generated output would be protected (N=28)

Source: Technopolis Group Survey

3.4.2.1.3. Discussion

The investigation conducted in this study finds that the current policy trade-off existing within copyright law, according to which in the absence of human creative choice no copyright should be granted, should be maintained. AI autonomously generated output should therefore fall outside the scope of copyright protection. Similarly, at this stage, there is no evidence of any need to create a *sui generis* right for AI autonomously generated output.

Based on **personality-related justifications**, copyright is conceived as a human-centric protection system. If the AI output is autonomously generated, i.e. without relevant human intervention in the expression of the AI output, there is no justification for copyright protection for the AI output.⁵⁵⁶

Furthermore, there is currently **no economic justifications** for recognising an exclusive right to AI autonomously generated outputs. From an empirical or theoretical point of view, no market failure could be identified that could justify the legal protection of AI autonomously generated output under copyright.

Empirically, AI solutions are still emerging and AI markets are in their infancy. Despite the absence of copyright protection for AI output, it has been observed that several companies are deploying AI solutions for automatically creating images, music, parts of video games, software and audiovisual productions. The market actors have not awaited the protection under copyright to invest and monetise the product of their services.

⁵⁵⁶ The AI solution may be (partially) protected under copyright for computer programs (and/or patent law), so the human creativity invested in the development of the software solution may be rewarded by exclusive rights.

Because of this absence of maturity of the markets, it is currently too early to establish that these markets are suffering a market failure due to the lack of legal protection. It can be empirically observed that research and investments in these fields are ongoing, suggesting that economic incentives are sufficient to stimulate the production and dissemination of AI output.

From a **theoretical** perspective, the existence of potential market failure could be based on the public good nature of AI outputs once released. While such an analysis is correct in terms of economic classification of the output as a public good, it would be incorrect to automatically infer the existence of a market failure from this characteristic. Indeed, producers of AI solutions do not need to control the output to monetise their investments, which are directed to the AI solution itself and not to the outputs. By controlling the access to and the use of the AI solution by contractual means, developers can monetise their AI solutions, using a service-based business model ("AI solution as a service").

3.4.2.2. Should AI outputs mimicking the "style" of AI inputs be considered a copyright infringement?

3.4.2.2.1. Issue and policy scenarios under consideration

As explained in section 3.3.2.1, generating output mimicking an author's style, a compositor or an artist's performance does currently not fall within the scope of copyright protection. This absence of protection could be a problem for some stakeholders since increasingly advanced AI applications are capable of approximating the style of human-made works or performances.

During the interviews and the Delphi survey, the following policy scenarios were investigated:

- **Status quo.** No legal action is undertaken. Artistic styles remain unprotected under copyright law, and no harmonisation of the unfair commercial practice claim of parasitism is undertaken.
- **Creation of a related (remuneration) right in relation to style.** Under this scenario, creators and artists have recognised an economic interest over their style, which is legally transposed in a remuneration right. Work or performances can hence be used as AI input, but their rightholders enjoy a claim to fair remuneration against users.
- **Protection of style under copyright law.** Under this scenario, the scope of copyright law would be extended to encompass an author's style. The generation of outputs mimicking such a style would hence no longer be legally possible without the rightholder's consent.
- **Harmonisation of unfair commercial practice claims of parasitism.** This last scenario, outside the scope of copyright law, proposes to harmonise within the EU the conditions of an unfair commercial practice claim for "parasitisms". Such a claim would allow under defined circumstances (for instance, the existence of competitive relationship) a claim for unfair commercial practice against undertakings, which propose on the market AI cultural outputs mimicking the work of human creators and artists, with the purpose to take benefit from their talent and reputation.

3.4.2.2.2. *Stakeholders' and scholars' opinions*

A certain **consensus** emerged between the different actors interviewed and the participants of the Delphi survey **against an extension of copyright law scope to style**. A legal scholar even considered this as an abusive extension of copyright law while others stated that "it must remain possible to get inspiration from someone else" and that copyright does not protect generally the "work" of a creator, but each of his work separately.

In the same way, the scenario of extending the scope of copyright to encompass protection for style was deemed less adequate by the participants, as 41% (22) of them slightly or strongly disagreed with it. Even among the participants who strongly or somewhat agreed with the broadening of the scope of copyright, one warned that such an extension would represent a considerable change to intellectual property law irrespective of AI. However, in the opinion, of this participant, it seems morally appealing that artists with a distinctive style should be able to prevent (or license) an AI system that mimics their style. One stakeholder advised that although the scope of copyright should remain unchanged, there should be special attention that AI mimicking does not become copying. Two other participants warned that there is the risk that parasitic exploitation of specific styles through an AI system might amount to the infringement of the general personality right of the author or performing artist.

The preferred scenario identified both by legal scholars and stakeholders is the **reliance on unfair commercial practices**. Hence more than half of the stakeholders (55%; 29) participating in the Delphi survey indicated their agreement with the creation of a claim against unfair commercial practice of parasitism for cases in which the style of an artist is mimicked/copied by an AI system (see Figure 56). This preferred option was followed by the status quo (49%; 26 participants agreeing) and the creation of a remuneration right for cases in which the style of an artist is mimicked/copied by an AI system (45%; 24).

Another legal expert interviewed considered this issue was also relevant to the protection of artists and interpreters, as deepfake practices could lead to legal actions. According to the expert, the artist's personality is also transferred during the process of AI training. This could be addressed by personality rights rather than "protection of style" because by protecting the style, there might be "convolutions of style" between the style of the work and the artist's style.

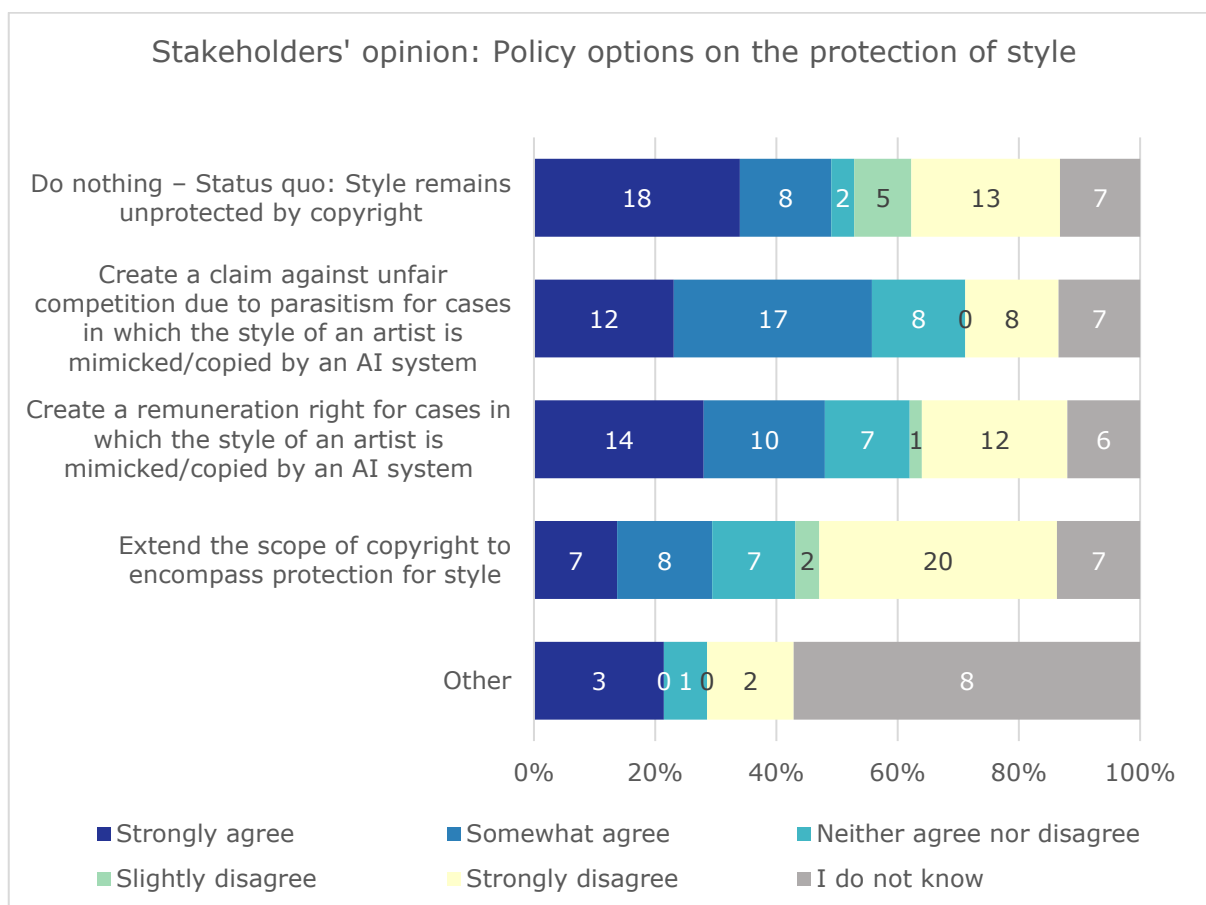


Figure 56: Stakeholders' opinion on the scenarios for the protection of style against AI-generated output and performance rights (N=53)

Source: Technopolis Group Survey

The three most preferred policy scenarios were further assessed by the Delphi participants in the second round of the survey. Their opinions on the impact of each of the scenarios on authors/rightholders, users of AI and producers of AI can be observed in Figure 57. While the status quo was assessed as the most favourable scenario for users and producers of AI in the cultural sector, the other two are almost unanimously considered positive for authors and rightholders.

Stakeholders were asked to assess the two alternative policy scenarios in terms of effectiveness and manageability. Figure 58 shows that in the opinion of the participating stakeholders, the most effective measure is the creation of a claim against unfair commercial practices of parasitism for cases in which the style of an artist is mimicked/copied by an AI. However, this measure is regarded as less manageable than the creation of a remuneration right, which is regarded as efficient and manageable by more than half of the respondents. One stakeholder working for an organisation in the field of text and images explained that, although they believe appropriate for style to remain unprotected by copyright, they do not consider a remuneration right as an advisable policy option. The reason is that, in their opinion, "a remuneration right would encourage users to copy/mimic the work, who would be able to do so unchallenged by the rightholders whose works they are copying". As a result, the rightholder works would be de-valued, and

the AI producer would be able to profit from the rightholders as little reward would be offered to them. Hence this stakeholder, suggests that a claim against unfair commercial practices would be more helpful to strengthen the position of creators/rightholders of human works. However, in the opinion of another stakeholder, unfair commercial practice claims would require a right of action by associations, as it would not be reasonable for an individual artist to be involved on their own in a legal dispute against a technology company.

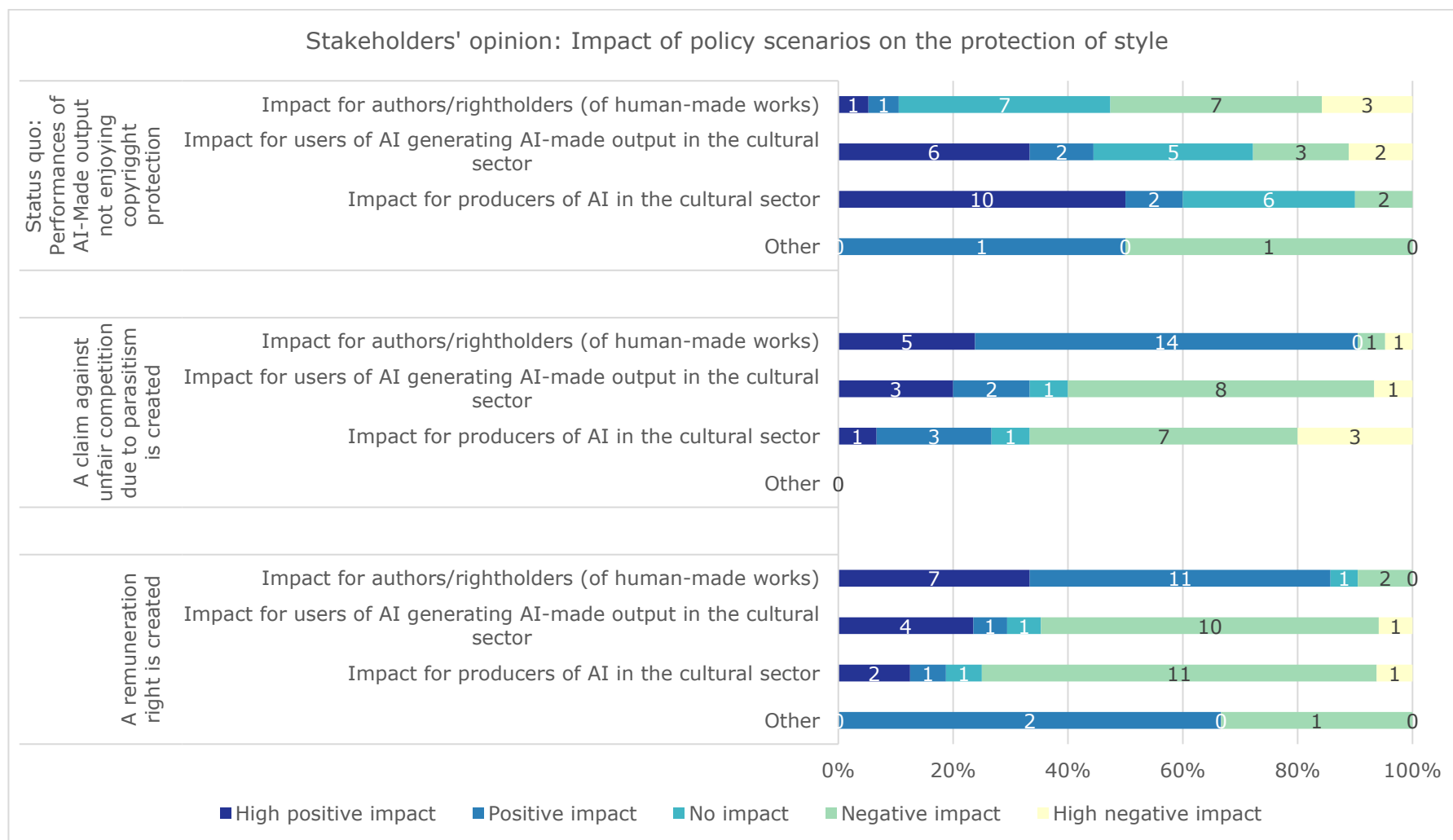


Figure 57: Stakeholders' opinion on the impact of policy scenarios on the protection of style (N=21)



Source: Technopolis Group Survey

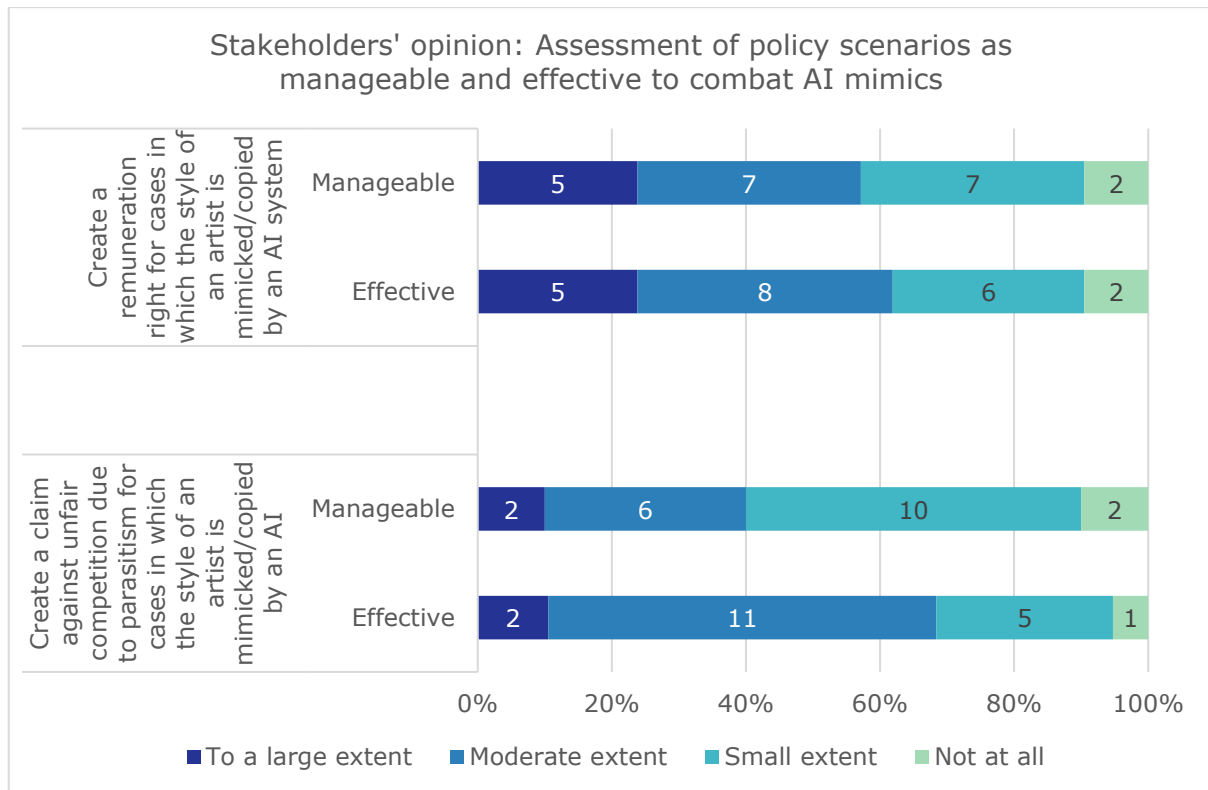


Figure 58: Stakeholders' assessment of the effectiveness and manageability of policy scenarios to combat AI mimics (N=21)

Source: Technopolis Group Survey

3.4.2.2.3. Discussion

Even if the “copy” of the style of an artist, author or composer might arise as one major challenge when it comes to the deployment of AI solutions within the cultural sector, an **extension of copyright law’s scope to protect the ‘style’ of a creator is unwarranted.**

It is a key principle of copyright that only the original expression of a copyright-protected work be protected, not the underlying ideas (such as style) – which must remain free to be used and reused by all (above see 3.3.2.1.1). The protection of style by copyright law would amount to a significant extension of copyright scope and the correlative restriction of artistic expression (including for non-machine creators). Protecting the style would in reality mean abolishing the principle that only original expressions are protected, by offering protection not only to the expressed result of creative choices, but also to the underlying idea. This would result in strongly restricting the public domain and hence impacting the freedom of creation.

At a more practical level, the notion of style is neither fixed nor exclusive: it is merely a mode of classification. Increasing the scope of copyright to encompass artists’ or authors’ styles would therefore automatically raise several practical issues, almost impossible to solve. How can an artist’s style be defined? How can a style be attributed to one author, knowing that human authors as well draw inspiration from other authors’ work and can

collectively contribute to a “style”? In that case, should several authors share rights on a given style?

For the same reasons, the **creation of a remuneration right in case of mimicking also seems unjustified**. Even if the restriction of the public domain would be less intense than in the case of the extension of copyright law, the same unsolvable practical issues would occur as to the definition and attribution of a particular style to a given artist or creator.

If a legal intervention were to be considered, it should happen outside the scope of copyright law. In this regard, **harmonisation at the European level of the claim for parasitism as a form of unfair commercial practice could be considered**. Such an idea has already been considered by the EU Commission, which in 2010 commissioned a study on parasitic copying.⁵⁵⁷ The report pointed “a variety of approaches being used to target parasitic copying”, going from IP to “unfair competition laws or unfair commercial practices or consumer protection laws” between the different Member states”. However, no legislative actions followed. Today, with the UK having left the EU, the prospective of harmonising this body of rules might be reactivated (Indeed, the UK was reluctant to harmonise this area of the law since the notion of unfair commercial practice or of an unfair competition tort was alien to UK law).

3.4.2.3. Should the human performance of an AI-generated creation be protected as a “performance”?

The issue of the protection of human performances of non-human creation is discussed under several policy scenarios, on which the opinions of experts and stakeholders were gathered. Lastly the question is examined in the broader copyright context.

3.4.2.3.1. Issue and policy scenarios under consideration

As explained in Section 3.3.2.2, artistic performances of an AI autonomously generated output cannot enjoy the protection offered by performer rights, because of the required link between the performance and a ‘work’.

During the interviews and the Delphi survey, the following policy scenarios were investigated:

- **Status quo.** Under this baseline scenario, no policy action is taken and the legal framework remains unchanged.
- **Revision of the related right subject matter definitions.** Without modifying the overall structure of the protection based upon a related right, the definitions of ‘performance’ or ‘performer’ are revised to also allow the protectability of performances having as support an AI autonomously generated output. This might be done either by generally deleting the reference to a work or by explicitly extending the definition to AI output (as is the case for instance for “expressions of folklore” under the WPPT).
- **Recognition of independent copyright protection to artists upon their performance.** This solution implies a more in-depth modification of the legal framework by questioning the ratio of the distinction operated between performers and creators within the existing legal framework.

⁵⁵⁷ HOGAN LOVELLS, “Study on Trade Secrets and Parasitic Copying (Look-alikes)”, *Final Report on Parasitic Copying for the European Commission* MARKT/2010/20/D.

3.4.2.3.2. Stakeholders' and scholars' opinions

As a matter of principle, a consensus exists among the stakeholders consulted that performers of AI autonomously generated output should not be deprived of protection. In this regard, the unanimity of the scholars consulted during the ECS experts consultation (see Figure 59) also reflects the outcome of the interviews conducted.

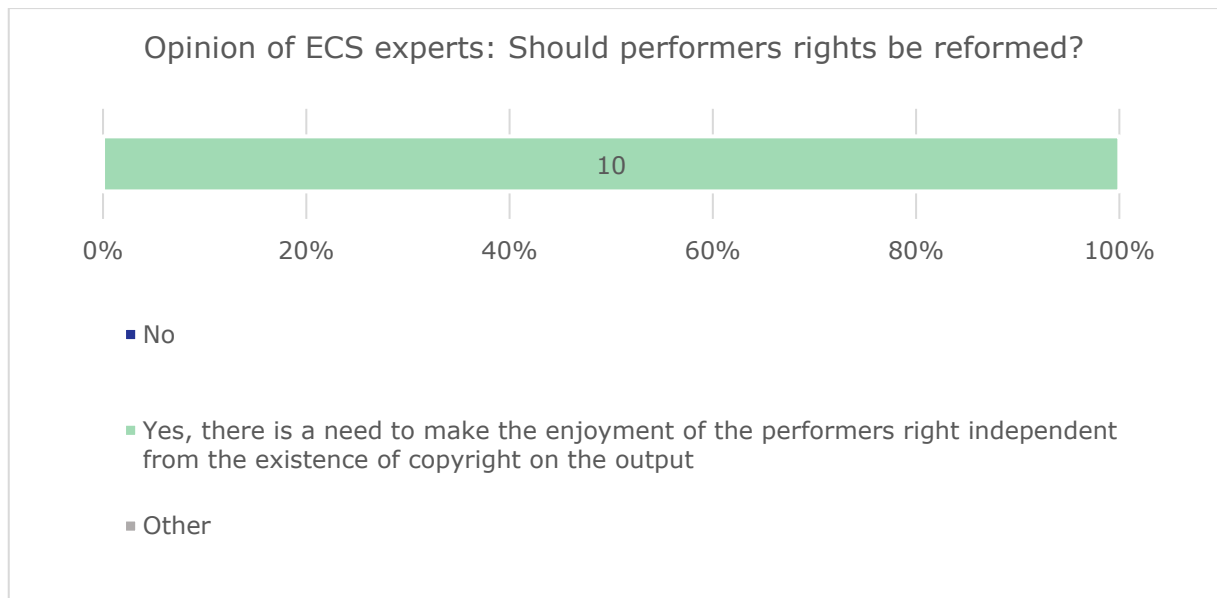


Figure 59: Opinion of ECS experts on a reform of performers' right (Total votes = 10)

Source: Technopolis Group Survey

The stakeholders participating in the Delphi survey also confirm this attitude, with significant rejection towards the status quo (see Figure 60). Among the policy scenarios presented above, their opinion seems to be slightly inclined towards granting copyright protection to artists for their performances (36%; 23 strongly or somewhat agree), although several participants slightly and strongly disagree (22%; 14).

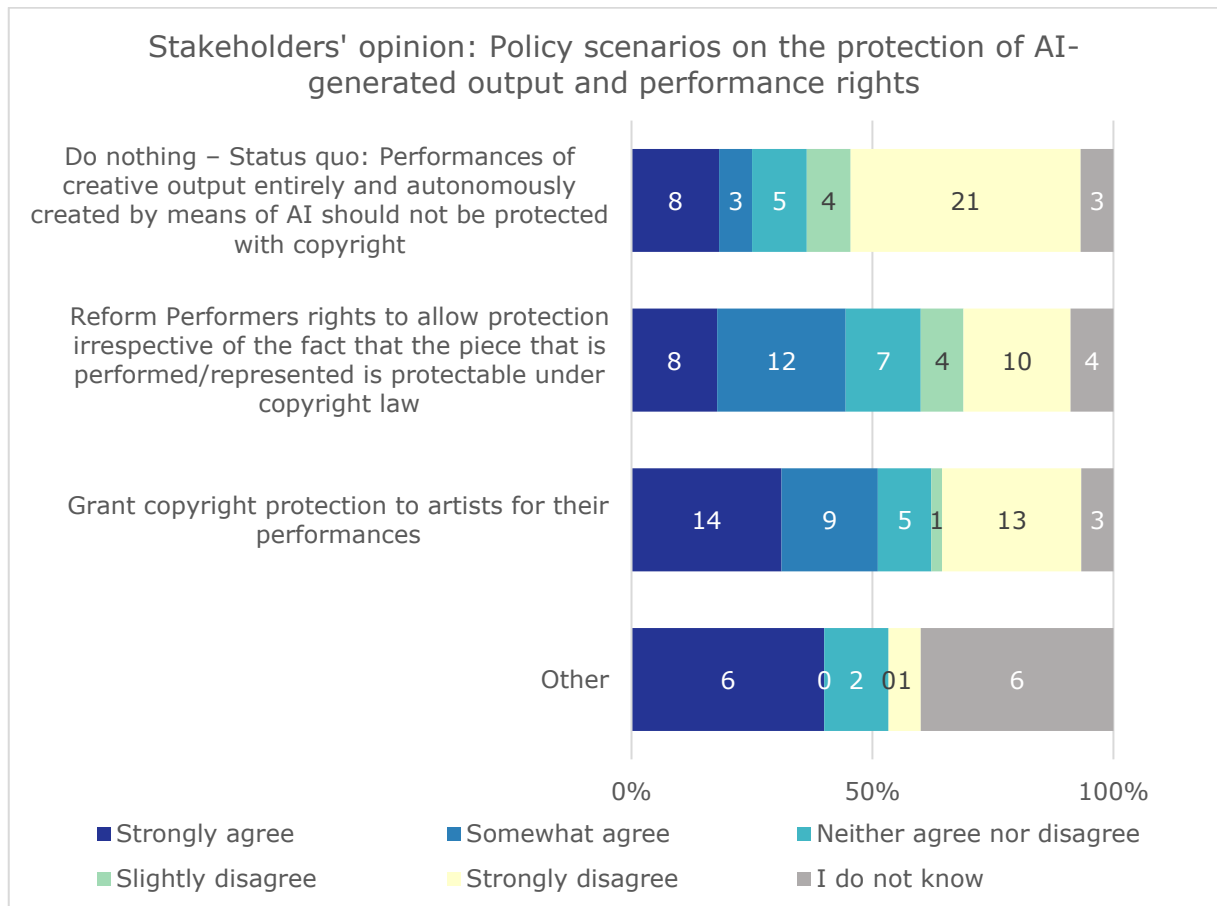


Figure 60: Experts' opinion on the scenarios on the protection of AI-generated output and performance rights (N=45)

Source: Technopolis Group Survey

In the opinion of six Delphi survey participants, performers' right should not depend on whether the performed piece is protectable with copyright but whether the performance brings a creative condition to it. An example given by a participant is that the same AI-generated song may substantially change when performed by a rock band or a ballad singer. Such proposition tends in fact to align the conditions of protectability of performers toward the copyright requirements. A participant lastly argued that the issue of performers protectability should be solved by recognising that given that AI is created by humans, the output of AI could be considered as a human creation. In such a case, the performance of an AI output would then be operated upon a work in the copyright sense.

In the second round of the Delphi survey, participants were asked to assess the impact of two scenarios on the revenue of authors and performing artists using AI-generated cultural output. The two scenarios under assessment and the participants' opinion were:

- The status quo (i.e. performances of AI-generated output do not enjoy copyright protection): Although 53% of the stakeholders believe that the status quo has no impact on the revenue for authors/rightholders, most of the remaining participants believe that it has a better impact on the revenue of authors/rightholders than on that of artists performing AI output.

- An alternative scenario in which performances of AI-generated output enjoy copyright protection: In this case, 63% of the participants believe that it will have a highly positive or positive impact on the revenue of performing artists using AI-generated cultural output. Regarding the impact on authors/rightholders, the opinions vary; while most believe that there will be no impact, 3 participants believe that the impact will be positive and 4 believe it will be negative.

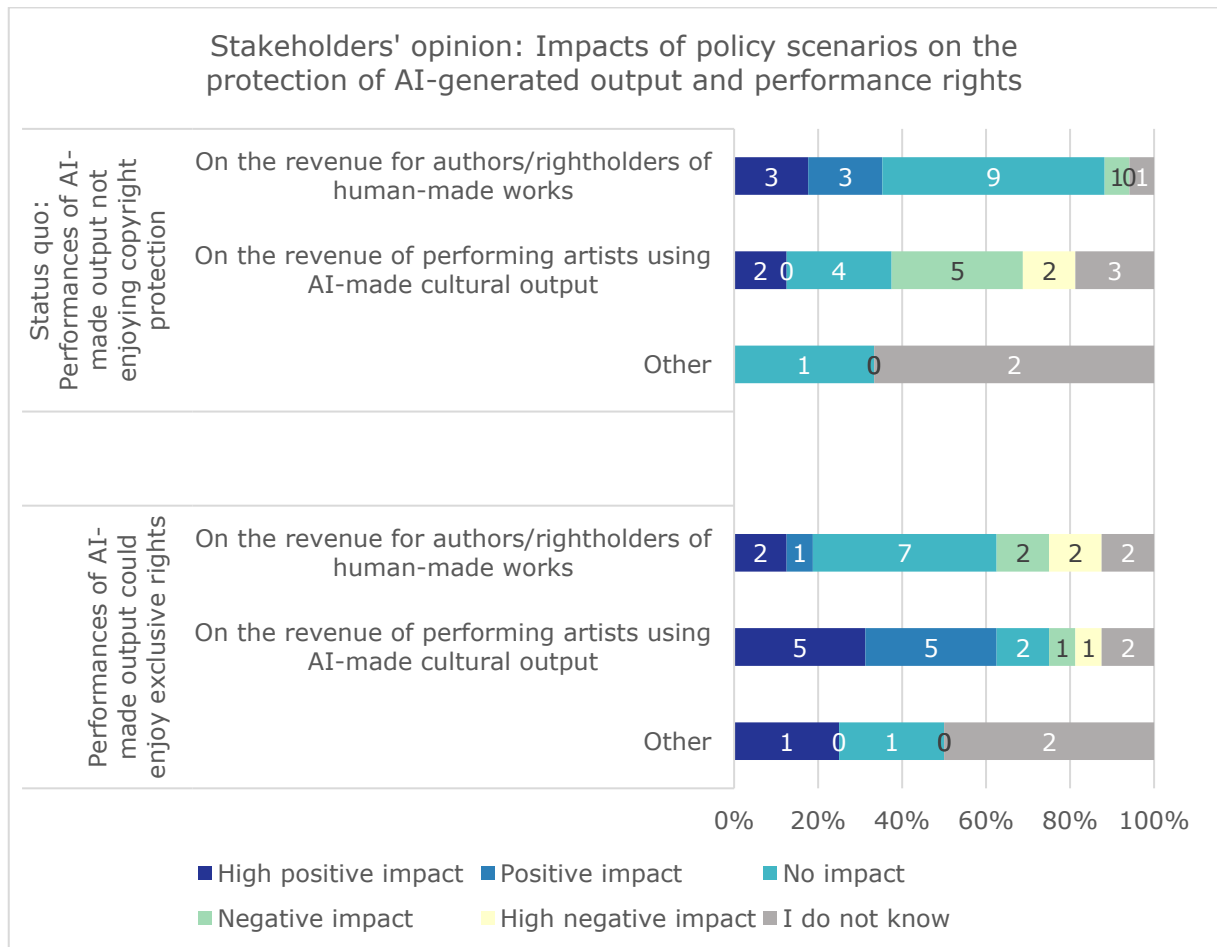


Figure 61: Stakeholders' opinion on the impacts of policy scenarios on the protection of AI-generated output and performance rights (N=17)

Source: Technopolis Group Survey

3.4.2.3.3. Discussion

Overall, it first appears that **the current status quo is not fully satisfactory to address the question of performance of AI-generated creations**. Where the definition of the protected "performance" is unclear (in its relation to copyright protected works), there is a risk that human performances of AI autonomously generated output are not protected because the performed creations are not protected under copyright.

If the absence of protection of human performances of AI-generated productions is the unintended result of the existing definitions, it may be useful to clarify the notion of “performance”, as the protected subject matter. If the purpose of the current definition is to distinguish between the human renditions in the **artistic or cultural sphere** and the human performances in other spheres, such as sports, then this view could be expressed more clearly. Instead of focusing on the performance of “copyright protected works” as such, the issue of protecting performances could be examined under a broader perspective more in line with the *ratio legis*, e.g. by referring to performances ‘in the creative sphere’ or ‘in the cultural sector’. Alternatively, a broader definition of “performances” with an explicit enumeration of the types of performances falling outside its scope could also be considered.

It could however be a conscious choice to restrict the protection of the performances to the renditions of copyright protected creations. The definition may be **meant to exclude** from protection all human renderings of creations that are not protected under copyright, including performances of AI autonomously generated outputs in the cultural sphere. Such choice would result in a discrimination between (identical) human performances of similar human-sourced creations. Such discrimination could be justified by the intention to provide an incentive to the performance of creative works of human origin.

Such justification however seems to disproportionately affect the performers: why should performers, who themselves are artists, be limited in the protection of their artistic expressions for the benefit of other creators? Such an approach would be in sharp contrast with the authors who, in order to enjoy copyright protection, are free to use AI solutions as tools in their creative expression. A human composer can generate a piece with an AI composer solution (ex: AIVA⁵⁵⁸), and by exercising creative choices upon the initial AI output, be granted copyright protection for the final outcome. It seems unfair to deny the human performer protection of their performance, merely because of the non-human origin of the creation they perform.

As a matter of principle, and in accordance with the consensus existing among stakeholders and IP scholars, a clarification of the notion “performance” may be welcome to avoid that, in some jurisdictions, the performers of non-copyright protected creations in the cultural or artistic sphere would not enjoy protection for their performance.

To achieve this goal, several options could be considered by lawmakers.

The first option would be to amend (at the national level) or to have (at the European level) a **harmonised definition of ‘performance’** as the related rights subject matter. This definition could clarify that a rendition can still qualify for protection, even if it is a performance of cultural output not protected by copyright.

This option presents the advantage of being first compatible with the international framework, since Art. 9 Rome Convention foresees that: “any Contracting State may, by its domestic laws and regulations, extend the protection provided for in this Convention to artists who do not perform literary or artistic works.” The legislator could decide to fully disconnect the notions of “performance” and “work”, an option that would have far-reaching consequences beyond AI applications and that would extend the protection to renditions traditionally not protected under the performer’s right (e.g. sports performances). It could also be clarified that performance of protected works or other expressions in the cultural sphere, to which certain examples can be added of renditions that are protected and some that are not protected (e.g. sports performance of athletes). Lastly, it could be

⁵⁵⁸ See case study, section 3.2.2.3.

added that the AI origin of a performed creation does not exclude eligibility for protection under the performers' rights.

A more radical way would be to fully restructure performers' protection and award them **copyright protection** of their performances as soon as they meet the copyright condition of originality. From a theoretical perspective, this option is appealing. Apart from historical reasons, it can be questioned why performers should not enjoy autonomous protection for their creative expression. The commentator H. Cohen Jehoram argued: "if personal expressions are protected by copyright (...) then it is difficult to see why a performance - which is always a personal expression - should be excluded. The mere fact that a performance presupposes a pre-existing musical or other work of a traditional author cannot be an obstacle. (...) From the beginning, translations and other adaptations of pre-existing works have been protected by copyright".⁵⁵⁹

From a pragmatic perspective, this option is however more challenging. Its compatibility with the international framework is first questionable⁵⁶⁰ and certain restrictions of the performers' rights (e.g. certain statutory licences) are not permitted under copyright and would consequently no longer be available for the use of the performances. This tension between the pros and cons of this option might explain the division of stakeholders toward this option, who are either strongly supporting it or strongly rejecting it.

3.4.2.4. Should AI output be protected under other related rights (producers rights)?

AI output in the cultural sector can be very similar to human created cultural content. Whether such AI output should be protected under the related rights is the object of different policy scenarios, discussed with experts and stakeholders and finally framed in the broader copyright legal framework.

3.4.2.4.1. Issue and policy scenarios under consideration

As explained in Sections 3.3.2.3 and 3.3.2.5 AI autonomously generated outputs are not protected under copyright, but may, however, enjoy protection under the rights of phonogram producers, film producers, broadcasters or press publishers. Contrary to the human-centric copyright, these related rights do not require a particular threshold, such as "originality" or a demonstrable investment, for a fixation or a recording to be protected. The production of AI outputs raises the question of whether these "industrial" related rights should be maintained in their actual form.

In this regard, during the interviews and the Delphi study, the following policy scenarios were investigated:

- **Status quo.** The related rights remain unreformed. Even excluded from the scope of copyright law, AI autonomously generated outputs can enjoy protection under these rights.
- **Addition of a condition of investment.** A condition of investment, in the form existing for the *sui-generis* protection of databases, and relating to the cost of generating the subject matter of the related right in question (for instance, the

⁵⁵⁹ H.C. JEORAM, "The nature of neighboring rights of performing artists, phonogram producers and broadcasting organizations", *Colum.-VLA JL & Arts* 2019, p. 15.

⁵⁶⁰ Comp. in this regard but concerning the US system: D. GERVAIS, The protection of performers under US law in comparative perspective", *IP Theory* 2015, vol. 5, iss. 1, 125 f.

fixation of the sounds of performance in the case of phonogram producer right), is added. In such a case, the mere 'saving' on the hardware of an AI-generated output would in itself not be sufficient to enjoy the related right protection.

- **Deletion of the related rights.** This option is the most radical since it invites the questioning of whether the different related rights still make sense today.

At the same time, these related rights do not cover all types of creations in a digital world. While the audio and audiovisual recordings may be protected, written creations do not give rise to any protection under the related rights, except if they qualify as "press publications".

3.4.2.4.2. Stakeholders' and scholars' opinions

Evaluation of related rights (other than performers' rights)

Asked about their opinions on the policy scenarios, the majority of participants would not suggest that related rights should be abolished. However, in the remaining policy scenarios, there was a low level of consensus (see Figure 62).

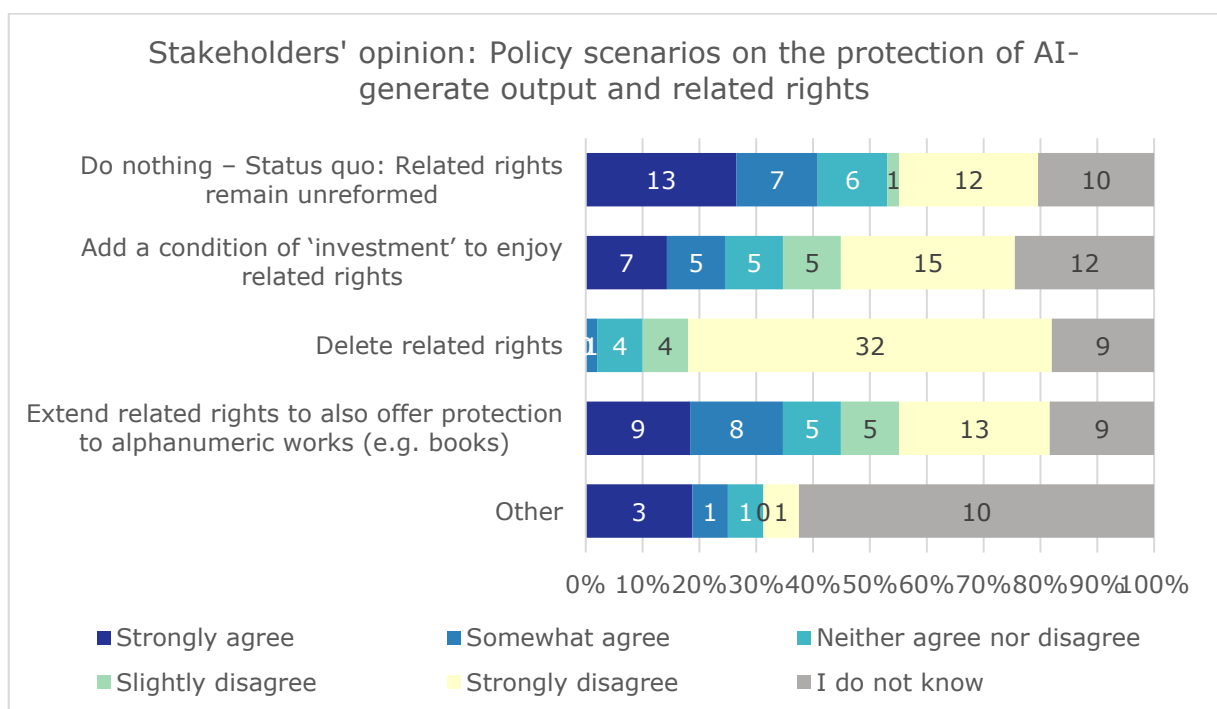


Figure 62: Experts' Stakeholders' opinion on the scenarios on the protection of AI-generated output and related rights (N=50)

Source: Technopolis Group Survey

The respondents in favour of the status quo argue that, currently, data are missing on the widespread use of rights to protect AI-generated outputs, and thus, further evaluations are needed before any policy action is taken. It was also noted that, to date, there was no evidence of market disruption that requires regulatory intervention. In addition, one participant explained that related rights should not be abolished as they constitute important protection in the audiovisual sector, long before the development of AI.

The legitimate use of the neighbouring rights was also discussed by one participant who believed that it would be impossible to assign this right fairly, given that the level of investment would be relative to the creators' resources which would vary widely.

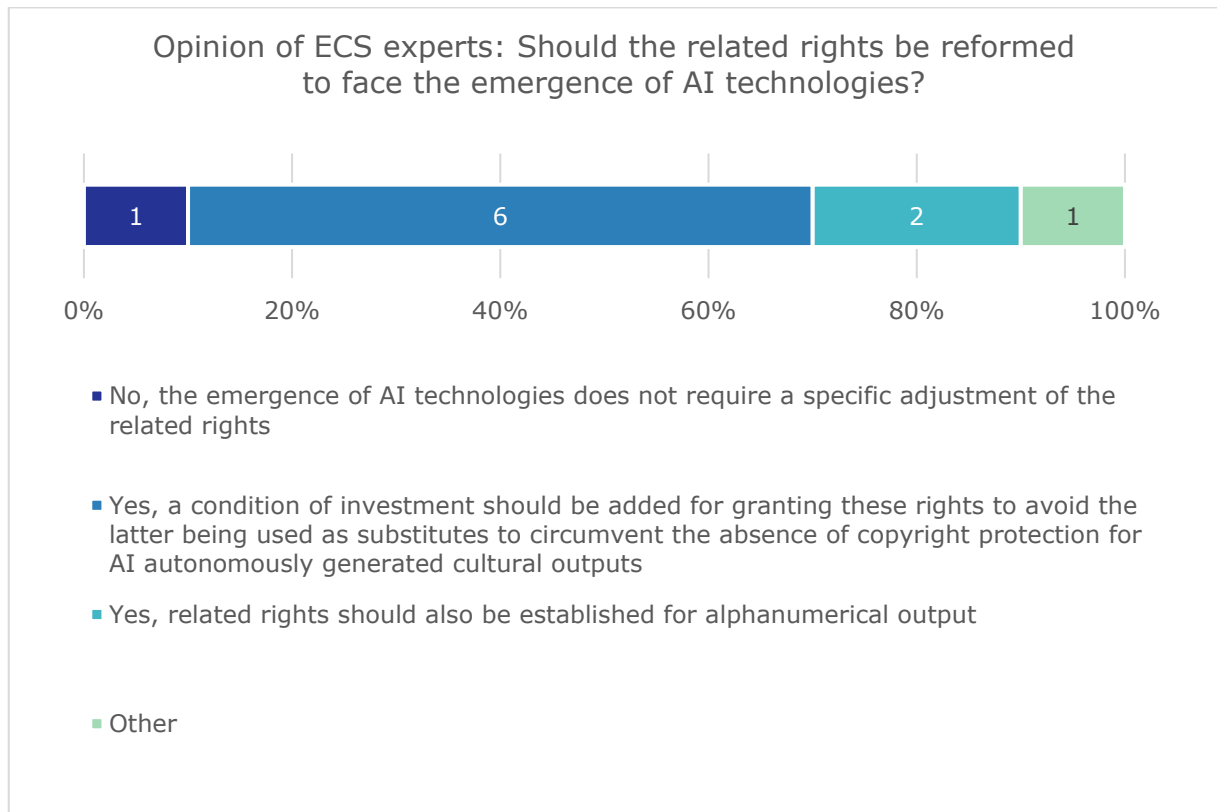


Figure 63: Opinion of ECS experts on reforming of related rights (Total votes = 10)

Source: Technopolis Group Survey

As it resorts from the ECS group, legal scholars appear much more favouring a reformation of the related rights to add at least a condition of investment (see also 3.4.2.4.3). The same conclusion also resorts from the interviews. There, the majority of them point the fact that the requirements of protection are dated and need to be amended to avoid this right being diverted from their purpose. Stakeholders were also asked to provide their assessment on the impact that adding a condition of investment to enjoy related rights of AI-generated output would have on rightholders, consumers, producers and the advancement of AI in the creative sector. There is no consensus on whether the impact of this scenario would be positive or negative, or whether there would be any impact at all (see Figure 64). The aspect where a higher proportion of stakeholders (43%; 9) agreed is that adding a condition of investment would increase the cost of producers to demonstrate whether the cultural product enjoys related rights.

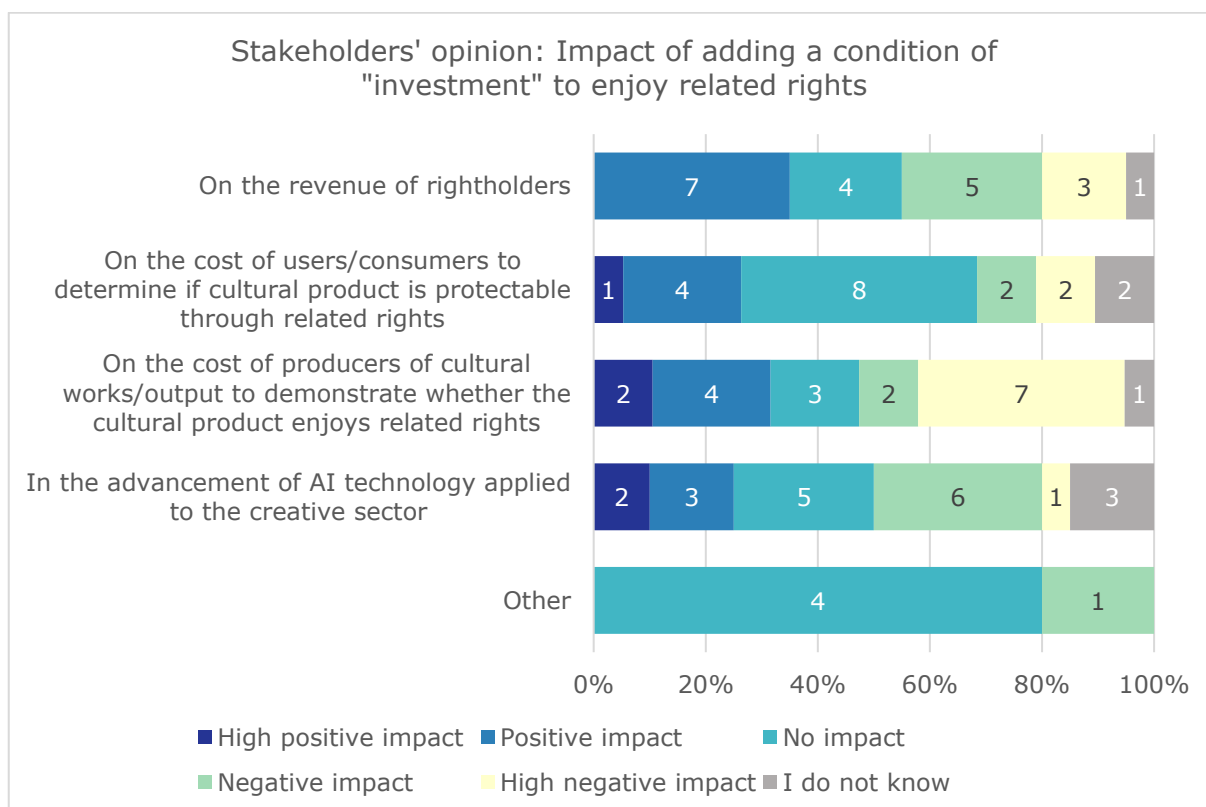


Figure 64: Stakeholders' opinion on the impact of adding a condition of investment to related rights (N = 20)

Source: Technopolis Group Survey

Need of a new related right for all AI-generated output

Not all efforts in relation to the exploitation of copyrighted works are protected under a distinct related right. Unlike productions in the audio and audiovisual sectors, the fixation of an alphanumerical creation as such is not a protected subject-matter, exception made for the press publications with regard to their online use by information society service providers.

The interviews conducted during the research do not support the creation of a new related right for AI autonomously generated output. A legal scholar raised some concern related to the coexistence of copyright with such a related right. According to a stakeholder, autonomously created works should be assessed and protected under copyright.

Legal scholars and a major company within the creative sector recommended caution with regard to the creation of a new right. In their view, one should first enquire whether there is a justification for creating such a right and positive and negative externalities should be weighed against each other (data flow, circulation of works, impact on non-machine authors).

Finally, several legal scholars mentioned the risk related to the creation of such related rights since the recognition of exclusive rights for AI-generated output creates incentives for the production of such output and could intensify their uptake. This might not be desirable as it could further erode the role of human creators, by creating a new source of competition.

Regarding the recognition of a related right on alphanumerical works, the number of stakeholders approving this scenario appears the same as of those opposing it. Ten out of the 17 participants in the Delphi survey who strongly or somewhat agree with this scenario are active in the publishing/book sector. One of them added that such rights should be extended, as this would allow publishers to defend the rights of the authors. However, one stakeholder from the publishing sector who disagreed with the creation of a related right for alphanumerical creations argued that neighbouring rights for the production of books are only relevant for the private copying remuneration, or as a neighbouring right for the narrator of an audiobook. This participant further explained that the production of a book (printing or conversion into digital form) is not a creative process, but a service and these investments get covered by the distribution and the sharing of the resulting revenues.

3.4.2.4.3. Discussion

AI output might be protected under the related rights of phonogram and film producers, broadcasters or press publishers, depending on the kind of output and, at least for the press publishers, the sector of activity. Whether the AI output is also protected under copyright, on the account of an original (human) activity, does not affect this assessment. These related rights might be used as a **substitute** for copyright protection in the case of AI autonomous generated output⁵⁶¹, in the sense that these rightholders may still control the use and exploitation of the AI output. Even in the absence of any copyright protected creation, the use of the AI output would still be subject to exclusive rights.

Concerns

The possibility to rely upon these related rights to protect AI output (even in the absence of copyright protection) raises **at least three concerns**. Firstly, the possibility of relying on these rights could be regarded as a way to **circumvent copyright policy trade-off**, according to which, if no human creative choices are made, no protection should be offered. Indeed, in the absence of human creative choices, no copyright protection applies to an AI-generated output.⁵⁶² The eligible related right has no equivalent conditions of protection.

This anthropocentric conception of copyright is justifiable, even without relying on moral arguments, from a mere economic logic. As trivial as it is, a machine does not need any incentive to generate a given output. The mere "fixation" of such a sound or a moving would be sufficient – even if there is no conscious act or organisation of the "recording" of such sound or image – to create a right coming close to the economic rights under copyright.

Secondly, the protection of AI output under the related rights without any demonstrable investment from the user would show how far these rights have diverted from their initial object. The purpose of these related rights was to recoup the important investments (e.g. in recording studios, technicians and operators, etc.) that the producers or broadcasters had to make. While the development and use of AI systems require important investments (protected under copyright and/or patent rights), the relation between the investment and the AI output is of a different nature than the traditional recordings.

⁵⁶¹ C. HARTMANN et al., "Trends and Developments in Artificial Intelligence. Challenges to the Intellectual Property Rights Framework", *op.cit.*, pp. 89–93.

⁵⁶² See in this regard the IVIR/JIPP study conclusion + J. DREXL et al., "Artificial Intelligence and Intellectual Property Law Position Statement of the Max Planck Institute for Innovation and Competition of 9 April 2021 on the Current Debate", *Max Planck Institute for Innovation and Competition Research Paper*, Nr 21-10, 2021.

Thirdly, AI output may be protected under related rights as it is an audio or audiovisual recording or press publication but will remain unprotected if it does not fall within these specific types of fixations. The **uneven protection** under the related rights, depending on whether the AI output, might be addressed in several ways: by creating a related right for any type of output, by revisiting the criteria for protection or even by abolishing the related rights (other than the performers' rights) altogether.

Ratio legis of related rights (other than performers)

The question of whether AI output should – as a rule – be protected under the related rights (other than performers' rights) raises the more general question of the *raison d'être* of the related rights.

Copyright protection is traditionally justified by two types of arguments, the remuneration of the author and the incentives for creation and dissemination of the work by the author. Similarly, related rights are justified to the extent that they create an incentive for the investment in the production and dissemination of the recording or fixation of various content. While some argue at length that the user of a robot needs the incentive to create AI output⁵⁶³ and consequently might require a distinct related right, the examples of the use of AI in the cultural sectors under examination do not provide any evidence of this hypothesis. As far as AI solutions and AI output are concerned, the business models are more complex and more multi-faceted than the traditional exploitation models on which the related rights were based: the users of AI systems may be open to other incentives for using these technologies and the resulting AI output than financial remunerations or the need to recoup their (limited) investment. Some scholars even contend that these investments, when it comes to AI autonomously generated outputs, are almost inexistent:⁵⁶⁴ as soon as the AI solution has generated the output, the user basically only has to click on "save as" to create a fixation and to be qualified as "producer".

Incentives

Despite the uncertainty about the legal protection of all kinds of AI output, the overview of AI applications in the various sectors shows that the developers and users of AI solutions have not awaited the clarification of this legal situation before developing an offer of AI solutions or adoption the commercially available solutions for generating AI output with limited or no human intervention.

The user's incentive for adopting AI tools may reside in the automation of certain labour-intensive or repetitive tasks (e.g. in the audiovisual or gaming sector). In that case, the AI output may be a part of the complex, human-machine mixed creation that might be protected as an audiovisual work and as a first fixation of film. The incentive for adopting the AI technology resides however in the efficiency gain (in terms of time and effort) and arguably no further incentives in the form of exclusive rights or remuneration rights under the related rights are required.

Other AI tools may autonomously generate AI outputs, with little or no human intervention from the AI user. As the case studies in the visual sector and musical sector have demonstrated (sub 3.2.1.3 and 3.2.2.3), the customers of Generated Photos and Aiva have

⁵⁶³ M. SENFTLEBEN and BUIJTELAAR, note 89. These authors distinguish the incentives required for the robot (which lacks legal personhood), the robot programmer (whose software creations may be protected under copyright and patent rights) and the robot user (whose creation may not be protected under copyright, nor under any of the related rights).

⁵⁶⁴ For BERNT HUGENHOLTZ, these technological investments completely disappeared irrespective of the AI context and these rights are in any way obsolete see: *Hugenholtz*, 2019, 1006.

access to the AI tools “as a service”. Differently priced offers allow the user to generate AI outputs at great ease and to use such AI output with little (contractual) restrictions. While the autonomously generated photos are not protected under any related rights and the autonomously generated music may or may not be protected under the phonogram producer’s right, both services seem to find a public of interested customers with incentives that can only be supposed to exist outside the related rights.

Other examples might be found in applications with more indirect remuneration models. For instance, a social media provider might integrate AI tools in its platform that the social media user can use to generate the accompanying music for their skateboard or dance videos. In that case, the AI user arguably does not need more of an incentive to use the AI tool than its mere availability as a feature of the platform. The social media provider will find an incentive in the possibility to retain its users on the platform and the resulting increased opportunities to monetise the users’ attention.

Where AI solutions are likely to be offered “as a service”, access to AI technology will be more realistic because no major up-front investments are required on the user’s side. More users will consequently produce AI output, which will lead to an increase of **related rights** – even though their beneficiaries might not have an interest in enforcing such rights. This means that exclusive rights are created, which hamper the free use of such output, although the holders of such rights do not see any need for this exclusive control.

These considerations beg the question of which role the existing related rights actually play in the creation of AI-generated content. If the offer is developing for AI-generated content that is protected or not protected alike, then should the contours of the related rights (other than the performers’ rights) not be revisited?

Investment in recording or fixation

The ratio of the “industrial” related rights should be examined to align the conditions for protection and the expression of the rights – keeping in mind that these rights were **not created as a policy answer to the emergence of AI technologies in particular**.

The European legislators have explicitly stated, concerning the related rights other than the performers’ rights, that “the investments required particularly for the production of phonograms and films are especially high and risky. The possibility of (...) recouping that investment can be effectively guaranteed only through adequate legal protection of the rightholders concerned” (rec. 5 Rental Dir). Similarly, in the InfoSoc Directive, it was acknowledged that “If authors or performers are to continue their creative and artistic work, they have to receive an appropriate reward for the use of their work, as must producers in order to be able to finance this work. The investment required to produce products such as phonograms, films or multimedia products, and services such as “on-demand” services, is considerable. Adequate legal protection of intellectual property rights is necessary to guarantee the availability of such a reward and provide the opportunity for satisfactory returns on this investment” (rec. 10 InfoSoc Dir).

While it is undisputed that the protection of the investment in the recording of sounds or films is the underlying reason for granting exclusive rights and remuneration rights, this “investment” was not explicitly required as a condition for protection (unlike the database sui generis right). It might be possible to consider that the necessity of some investments is already implied in the law.⁵⁶⁵ However, such an implied condition could not be regarded as equivalent to an explicit condition of substantial investments as stipulated for the

⁵⁶⁵ See the argumentation developed by the Court of Justice in the Pehlam decision, as presented in 3.3.4.1.2.

database sui generis right. If only implied, the condition indeed mainly allows interpreting the other (explicit) conditions, as it becomes a condition "of its own" that need to be independently fulfilled for the related right to be granted.

Tracing the international origin of the producers' rights back to the Rome Convention (1961), it can be supposed that in the existing technological context any recording required an important investment (in technical equipment and specialised professionals) so it was perhaps superfluous to make this "investment" an explicit requirement for protection. Meanwhile, with the digitisation of the past 25 years, the recording and communication technologies have developed and diversified, up to a point that a simple phone will do for audio and video recording of sufficient quality to be disseminated online. The balance between the object of legal protection (without investment requirement) and the purpose of protection (protection of the investment) was thus upset by the technical innovation of the past decades. This democratisation of recording and communication technologies raises the question of whether there is a need to fine-tune the scope of protection under the related rights and to grant such rights only if the producer indeed has made an actual investment in the creation of the audio or video recording, which they could not recoup without an exclusive right or a remuneration right.

Explicitly requiring that the investment in the fixation or recoding be demonstrated, as a condition for protection under the rights of the phonogram producer, the producer of first fixations of films but also the broadcaster and the publisher of press publications, would allow a realignment of the protected subject matter and the ratio for protection, independently of the technological context. Arguably, producers, publishers and broadcasters make important investments to achieve commercial success. It should however be avoided to reward commercial investments of a general nature with an exclusive right, where other undertakings do not enjoy such right to cover their business risks.

This would mean that the user who has used an AI tool to generate audio- or audiovisual output, without demonstrable investment that needs to be recouped (e.g. monthly subscription for the use of the AI tool), would not be eligible for protection. One could indeed argue that the investments are made in the development of the AI tool and that the AI solution provider requires the means to recoup their investment – not the AI user who has the benefit of using the AI tool for free or for a monthly subscription fee.

Finally, an investment in the fixation or the recording would be demonstrated for the producers to enjoy protection under the related rights. The criteria for protection and for finding infringement should however not be applied symmetrically. In line with some recent decisions of the CJEU, the criterion for finding infringement could take into account the possibility for the rightholder to obtain a proper return on investment (*supra* sub 3.4.1.1). Can the authorship presumption be applied to AI output?

The authorship presumption may raise some issues in relation to AI output, which have been presented to stakeholders and experts in various policy options. The outcome is then placed in a broader copyright discussion.

3.4.2.5. Should the presumption of authorship-rule be adapted to address potential AI-related abuse?

3.4.2.5.1. Issue and policy scenarios under consideration

As explained in section 3.3.3.3, false authorship attribution to a human of an AI autonomously generated output may, in practice, allow the absence of copyright protection for this type of output to be circumvented.

During the interviews and the Delphi survey, the following policy scenarios were investigated:

- **Status quo.** The conditions for authorship attribution (including the authorship presumption) remain unchanged.
- **Modification or deletion of the conditions of the authorship presumption.** Under this option, a reinforcement of the requirements for benefiting from the authorship presumption could be considered.
- **Introduction of an information obligation concerning the reliance on AI solutions to generate the cultural output.** This scenario is discussed in more detail in section 3.4.2.6.
- **Introduction of a sanction mechanism in case of false authorship attribution.** Under this scenario, the adoption of criminal types of norms, sanctioning intentional false authorship attribution, is considered.

3.4.2.5.2. Stakeholders' and scholars' opinions

Concerning the policy scenarios, the majority of the legal experts and scholars interviewed, as well as the broad majority of respondents to the Delphi survey, agreed that when it comes to the **condition of authorship attribution, including the presumption of authorship, the status quo should be maintained** (see Figure 65).

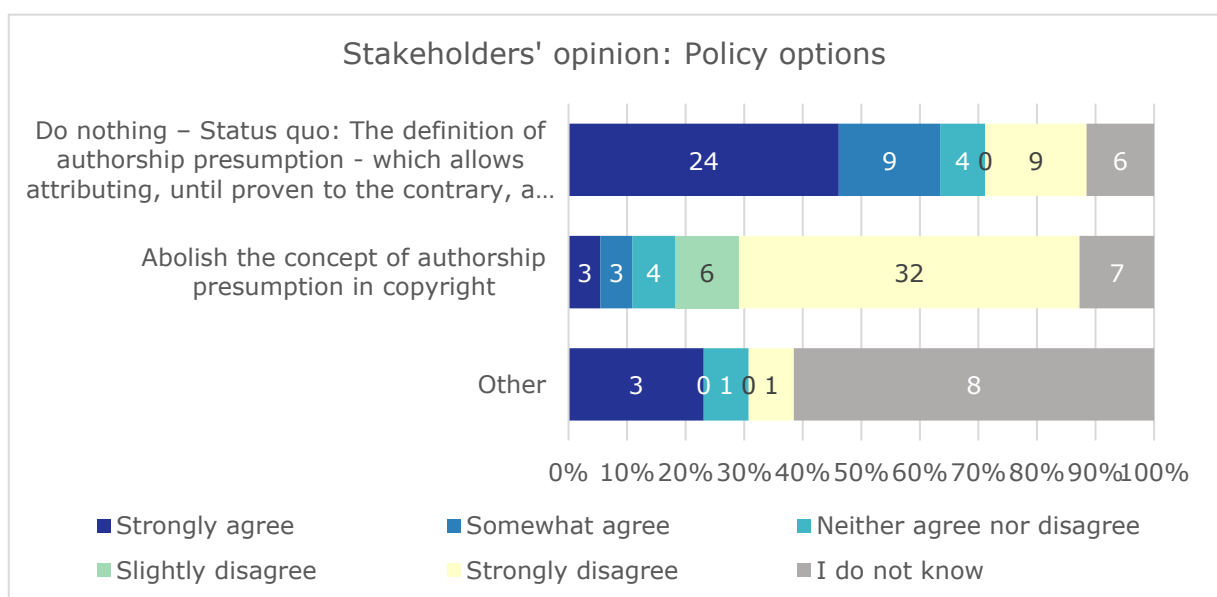


Figure 65: Experts' opinion on scenarios on false authorship and authorship presumption (N=52)

Source: Technopolis Group Survey

Some stakeholders are indeed convinced that the existing AI solutions are not modifying the need for human involvement in the generating of cultural outputs. Hence, two stakeholders (one in the publishing sector and another in the visual sector) in favour of this approach argued that humans still needed to be involved in the correction or improvement of every AI-generated output. Therefore, humans still contribute with artistic choices and there should not be any objection to their authorship.

The importance of the presumption of authorship was mentioned by several survey participants, in relation to the fight against counterfeiting and piracy, and the whole system of intellectual creation in general. A legal scholar specified that in its absence, authors would be deprived of a useful tool while false authorship attribution will remain possible. In addition, one respondent to the Delphi added that the fact that the concept of author presumption might be misused with AI did not mean that it should not be a reason to abandon it.

A stakeholder from the publishing sector and one scholar specialising in IP and the music industry further argued that the question as to who would suffer from a false presumption of authorship in the case of an entirely AI-generated output should be raised. In this participant's view, this will be the competing claimant, as in false authorship cases involving no AI. Therefore, the existing mechanisms for challenging false presumption would be adequate.

The opinion of the ECS experts is along the same line. According to the majority of the scholars in this group, the conditions for the presumption of authorship should not be reformed. However, they emphasise a need for clarification concerning the presumption's purpose. The presumption should be used only to attribute a given cultural output to a person, not as proof that this output is indeed the result of creative choice made by this person (and therefore protectable under copyright).

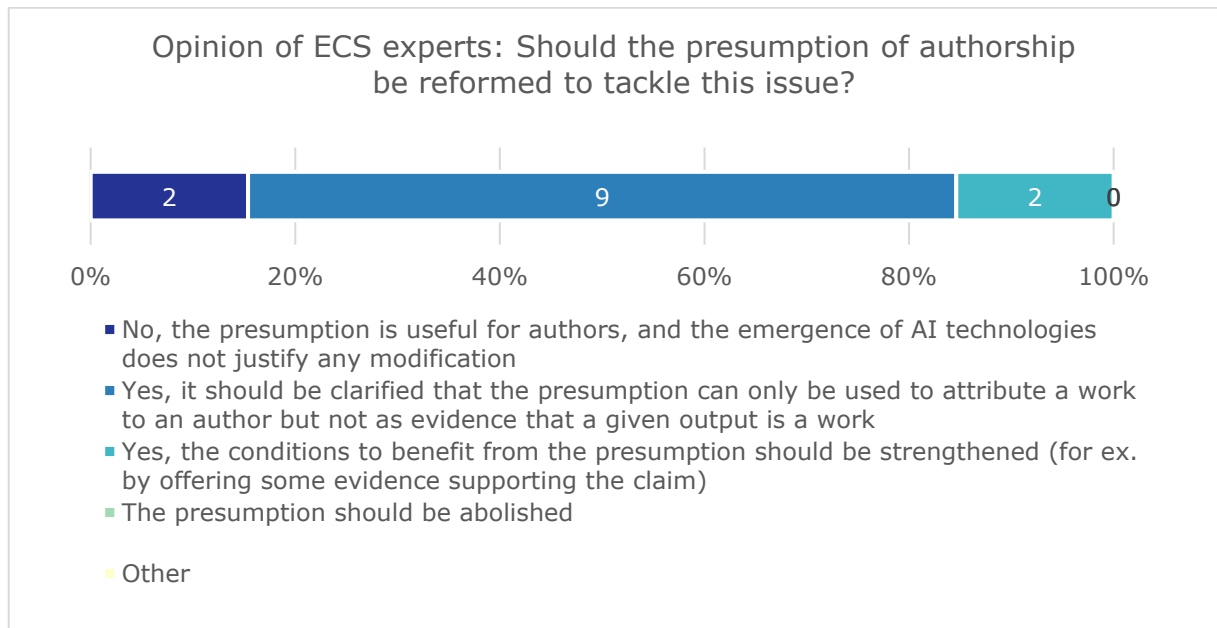


Figure 66: Opinion of ECS experts on a reform of the presumption of authorship (Total votes = 13)

Source: Technopolis Group Survey

In contrast, one respondent argued that the presumption of authorship of humans in cases of AI-generated output could be seen as unfair to algorithms.

One participant in the first round of the Delphi survey mentioned the need to mitigate the phenomenon of false use of the legal presumption, for instance by imposing **sanctions** on those persons who falsely declare they are the author of the output automatically generated by AI. In the second round, one stakeholder (an academic in the field of IP) agreed with this option and argued that sanctioning would be an adequate approach, even in criminal law.

Some participants also discussed the necessity of an information obligation. In the opinion of two of them, there should be bespoke rules for art pieces that have been aided or entirely created by AI, such as a **labelling** obligation for AI-generated output. In the second round, two participants from the publishing sector and one scholar in the field of IP recommended that AI solutions should generate output with a note that allows the identification of AI-generated output, for instance in digital form through a watermark or in the file's metadata.

Lastly, other options were proposed by some stakeholders. One participant raised the question of whether, in the case that copyright protection would be extended to artistic outputs entirely generated by AI, a default position on authorship or first ownership should be introduced into the law. A legal scholar raised the issue of the possibility to create a *sui-generis* right for computer-generated works (as in the UK). A stakeholder from the music sector suggested improved enforcement, accounting and audits of rights, both for rightholders and users.

Two stakeholders (one from the publishing sector and another from the visual sector) in favour of this approach argued that humans still need to be involved in the correction or improvement of every AI-generated output. Therefore, humans still contribute with artistic choices and there should be no objection to their authorship.

3.4.2.5.3. Discussion

Despite not being fully acknowledged by all stakeholders, AI autonomously-generated outputs are today a reality. The present impossibility of determining – from the outputs alone – how creative outputs were developed could lead to false attribution phenomena and have an impact, for instance, on the remuneration of creators through CMOs.

Before considering any legal actions, it should be noted that AI solutions are currently still in their development stage. Even if, according to our knowledge, no technical solutions currently exist to detect AI autonomously generated cultural output, research is, however, conducted in the area of deepfake detection.⁵⁶⁶ Therefore, it is possible that in the same way the issue emerges because of technical innovation, it is also possible **that technological solutions to solve this issue will appear in the future.**

If policy actions were to be considered, there is no reason at this stage to revise **status quo for the authorship attribution, including the authorship presumption.** The effect of the presumption of authorship could be to – de facto – reinforce the risk of false authorship attribution. To remedy this issue, it could be considered modifying or abolishing the presumption of authorship, in order to solve this false authorship attribution. Such intervention would however be excessive, considering the negative impact it would entail on human creators. As explained in Recital 19 of the Enforcement Directive, the decision to rely on such a presumption results from the absence of formal registration. Hence, in its absence, any human creator would also have to establish their authorship upon a given work before being entitled to institute an infringing proceeding.

However, **the purpose of the presumption should be clarified.** It should only be used as a presumption of attribution, i.e., the identified person is presumed to be the one to which copyright should be attributed over the cultural output – if the output qualified as work. On this opposite, the presumption should not be regarded as a presumption of authorship, in that the identified person should be entitled as an 'author' with regard to the piece concerned. Such a qualification would indeed imply that the output qualified as work. The burden to demonstrate that the piece concerned qualifies as work should hence remain on the person claiming copyright protection.

While the introduction of **an information obligation** to inform that a given output has been developed with the assistance of or autonomously by an AI has been suggested by some stakeholders, it will be explained in section 3.4.2.6, why this option **should rather be discarded.**

Lastly, sanction mechanisms could be considered in cases where a person deceptively claim to be the author of an AI autonomously generated output in order to claim subsidies that should accrue to the creators, for instance by registering such output with CMOs. Sanctioning mechanisms should, in principle, remain the *ultima ratio*. However, for as long as technical solutions have not been found to help identify AI autonomously generated outputs and no other legal answer proves to be fully satisfying, the reliance on such type of sanction, with a deterrent effect, could be justified. In this regard, these 'sanction' rules could already be partially implemented by CMOs through the penalty clauses in their statutes.

⁵⁶⁶ See for instance: J. VINCENT, "Facebook develops new method to reverse-engineer deepfakes and track their source", *The Verge*, 2021, <https://www.theverge.com/2021/6/16/22534690/facebook-deepfake-detection-reverse-engineer-ai-model-hyperparameters>, (accessed in July 2021).

3.4.2.6. Should additional transparency or information obligations be imposed?

The development and use of AI solutions raises transparency issues in general but also in the cultural sector. Different policy scenarios can be considered and have been discussed by stakeholders and experts. The outcome is framed in the broader copyright context.

3.4.2.6.1. *Issue and policy scenarios under consideration*

As explained in Section 3.3.3.3 and 3.4.2.5, it might be difficult, if not impossible, to know from the cultural output only if and to what degree AI solutions were involved in its development. The idea of the imposition of an information obligation to solve this issue and empower consumers with information that might impact their experience was therefore investigated.

During the interviews and the Delphi survey, the following policy scenarios were investigated:

- **Status quo.** No information obligation relating to the reliance on an AI solution is added.
- **Imposition of the information obligation.** Under this scenario, AI solution users producing AI cultural outputs must disclose how these outputs were developed. The imposition of such an information obligation requires a precise definition of the cases in which such information is required (only for AI autonomously / also for AI-assisted generated outputs) and of the sanctions applicable in case of violation.
- **Creation of a “human-made”.** This option is comparable to geographical indications. As the opposite of an information obligation, the reliance on such a label, which supposedly offered a competitive advantage, would only be a possibility for those fulfilling its condition of use. Such an option would here again require a careful definition of these conditions.

3.4.2.6.2. *Stakeholders’ and scholars’ opinions*

Overall, the interviewees and the participants in the Delphi survey appear to share a **certain acceptance of the implementation of an obligation of information**, at least in the case of AI autonomously generated outputs.

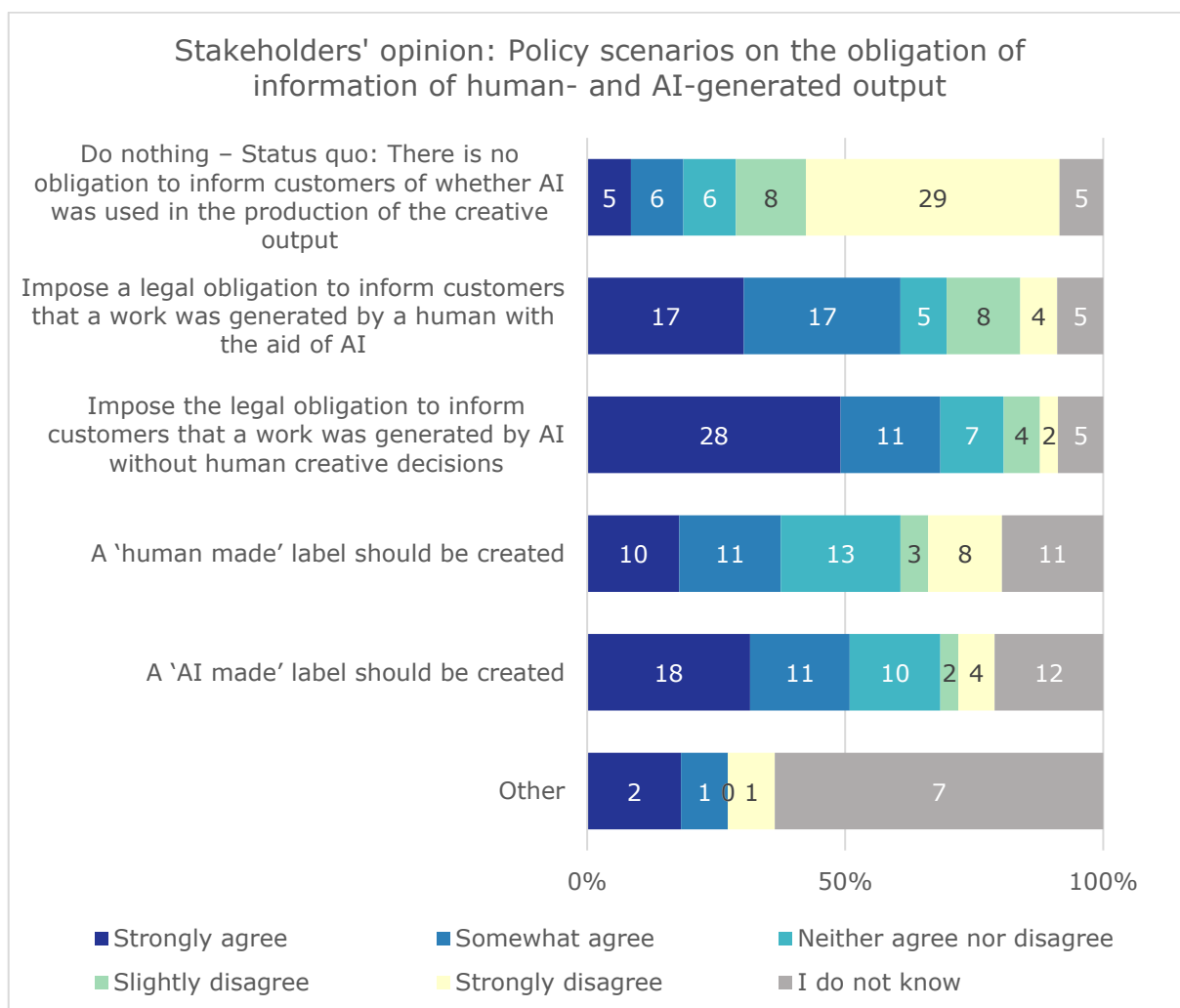


Figure 67: Experts' opinion on the scenarios for the obligation of information on the human/AI generation of creative output (N=59)

Source: Technopolis Group Survey

Hence, in the opinion of 68% (39) of the stakeholders participating in the Delphi survey, the imposition of a legal obligation to inform customers on the use of AI for the generation of the cultural piece would be important if a human does not contribute with creative decisions. In general, the opinion of these participants is that in the creative sector, authorship is a very important concept (e.g. in terms of authenticity), and therefore, transparency is important. In the same way, none of the interviewees was clearly opposed to the implementation of such an obligation of information. Three interviewees (a legal expert, a stakeholder and a major market player) clearly supported the implementation of such an information obligation. However, according to the majority of the interviewees, two questions that should be addressed before implementing such an obligation were raised: (i.) what should be considered as an AI-autonomous work and what should not (the boundary is blurry)?, and (ii.) what is the default assumption of consumers? One legal

expert also expressed concerns concerning Art. 5 of the Berne Convention. When it comes to AI-assisted outputs, the results are more contrasted. In the Delphi results, the proportion of participants (60%; 34) who indicated that they also strongly or somewhat agree to inform customers when the creation of the work is aided by AI, is slightly lower than for AI autonomously generated outputs, but still high. However, it should be noted that four participants differed and argued that the use of AI as aid is very common in the creative sector as many post-processing programs incorporate AI features without users being fully aware of it (e.g. many basic functionalities in Adobe Photoshop run AI in the background). In such cases, they regard AI as a mere tool with an ancillary role, which does not have a significant impact on the work as the expression of the author's personality. This reluctance is also shared by the scholars interviewed, as the opinion of the ECS experts demonstrates.

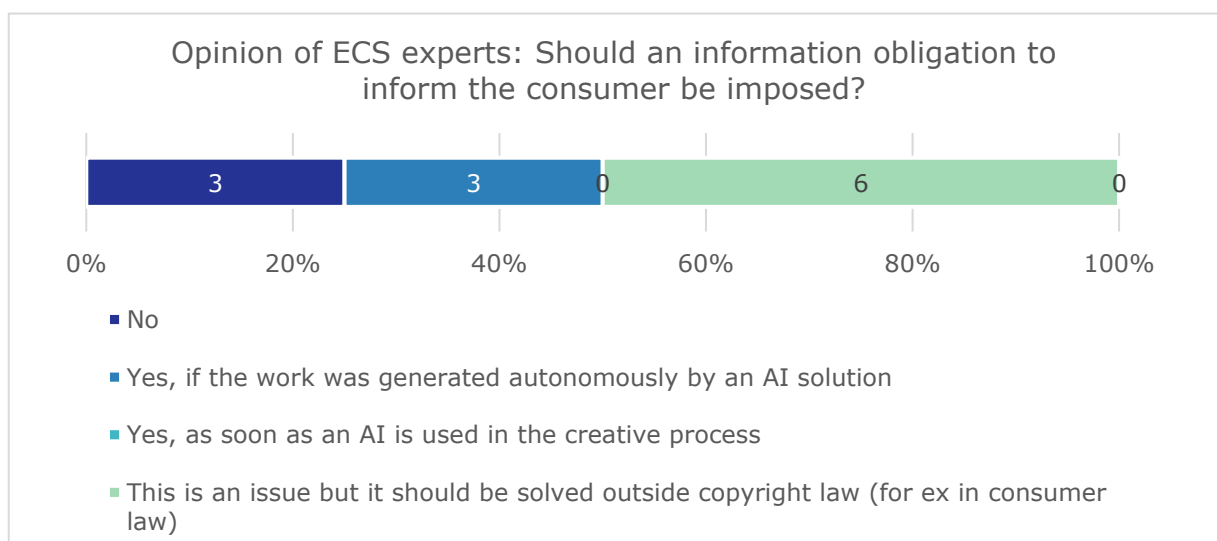


Figure 68: Opinion of ECS experts on the imposition of an information obligation towards consumers (Total votes = 12)

Source: Technopolis Group Survey

The policy scenarios with less acceptance were the **status quo** and the **instauration of labels**

Lastly, it should be noted that the participants often question the necessity to address this issue within copyright law. One of the interviewees, for instance, mentions that the AI regulation proposal, which includes such an obligation, is undoubtedly a better place than copyright law. In the same way, the participants of the ECS survey were mainly of the opinion that this issue should be solved outside copyright law.

In the second round of the Delphi survey, stakeholders were asked to assess the potential impact on costs of the policy scenario that they regarded as most adequate in the first round of the survey: the imposition of a legal obligation to inform customers that a work was generated by a human with the aid of AI (see Figure 69). In the opinion of the stakeholders, the policy scenario would not lead to significant costs, but it would have a high positive or **positive impact on consumer behaviour** (e.g. willingness to pay for AI-made output or in the consumer choice between human-made and AI-made content) and

on the misuse of AI (e.g. individuals claiming authorship of AI-made output; when AI is used to create misinformation, etc).

A participant added that compulsory labelling, although it should be mandatory, would not be easy to implement in every creative sector. This participant gave an example of the music sector, explaining that music advertising jingles are often AI-generated. Informing the radio listener about the artificial nature of the melody used in advertisements would require an interruption, such as "The following music for this product was created by [name of the AI tool]", which is not attractive to companies. However, this participant argued that in the case of AI-generated music offered as complete products to consumers, the labelling informing about the AI nature of the piece should be required.

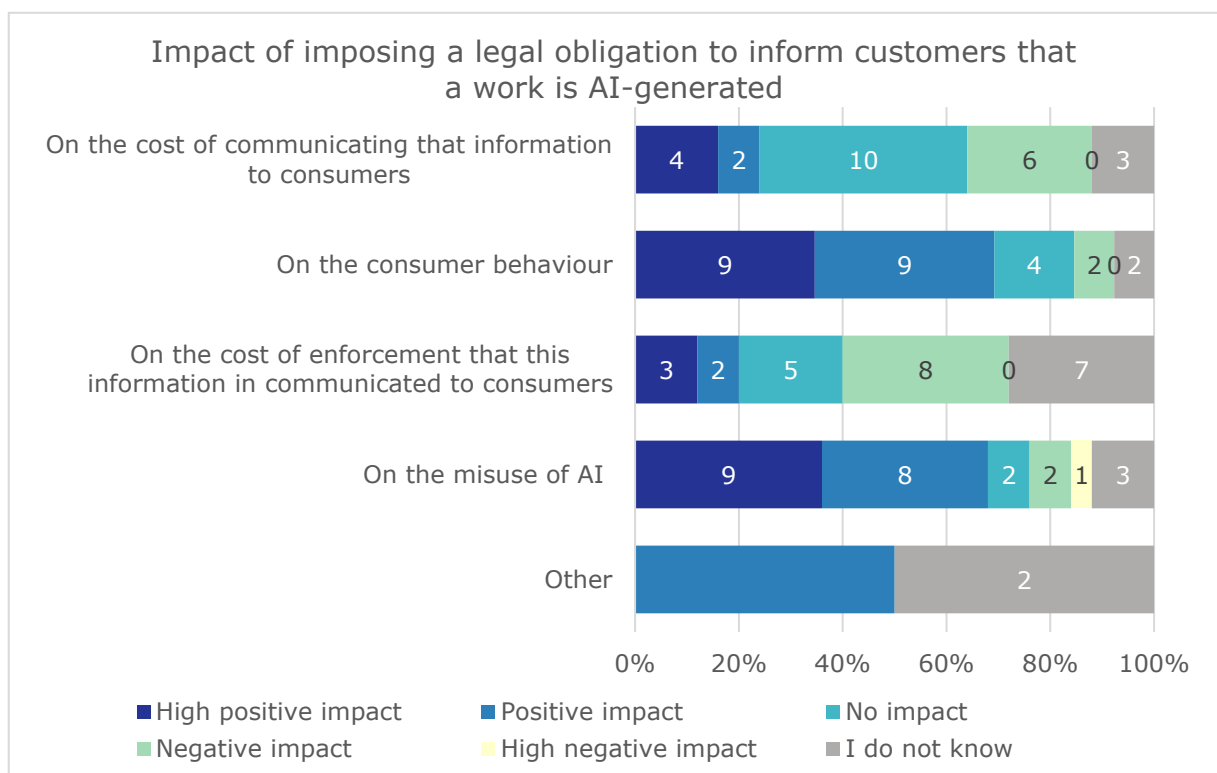


Figure 69: Stakeholders' opinion on the impact of imposing an obligation to inform customers that a work is AI-generated (N=26)

Source: Technopolis Group Survey

3.4.2.6.3. Discussion

Whether an art piece in the cultural or artistic sphere is created by a human or an AI system may affect the perception and the experience of the public. Depending on the type of creation, they may attribute a different meaning to a painting in a museum or a musical performance in a concert hall when they know such creation is not the result of human experience and effort. For other creations (such as stock photos, background music or logos), the human origin may not be a factor of importance. The answer may also depend on personal, subjective factors, which makes it difficult to provide a general answer to the question of whether the AI origin of a creation should be disclosed (ex-ante or ex-post).

However, for (creative or tech-driven) users of AI content, it may be interesting to know that a particular creation is AI-generated and consequently not subject to copyright, so it can be freely used, without the prior consent of the copyright holder.

From a mere copyright perspective, a justification for imposing an information obligation under copyright law could be the difficulty in differentiating, from the output only, between those outputs resulting from human creative choices (and therefore deserving copyright protection) and those autonomously generated by an AI solution. Imposing such an obligation to disclose the AI origin, backed by sanctions in case of deceit, would facilitate the identification of the potentially protected outputs. It could also help address the concerns on the legal presumption of authorship: if the AI origin of a creation must be disclosed, then the presumption of authorship will not apply.

However, an **information obligation** concerning the AI origin of any creation raises several concerns. Firstly, as noted by one of the IP experts, the introduction of such an obligation within copyright law could raise concerns regarding its compatibility with international law. Art. 5(2) Bern Convention provides that “the enjoyment and the exercise of these rights shall not be subject to any formality”. Hence, the information obligation must not be understood as a condition to enjoy copyright protection.

Secondly, and more fundamentally, requiring a creator to disclose their creative process raises certain concerns regarding their **artistic freedom** and **personality rights**. The choice to disclose – or not – how a given work was generated might indeed be considered as part of the creative process. To enjoy copyright protection, creators have never been required to disclose the technique they deploy or the instrument or method they use to achieve a given creative work. The author or artist may enjoy some freedom to construct their own origin story.

Lastly, the **scope** of such an information obligation may not be so clear, especially concerning the importance and degree of autonomy of the AI solution in the creation of the AI output. On the sliding scale of AI autonomy, it may be difficult to determine at which stage an output “autonomously generated by AI” and therefore subject to an information obligation. Alternatively, all use of AI tools should be disclosed. Since AI is and will in the future increasingly be implemented in tools supporting creative processes, a general imposition of such an obligation would lead to flag very high numbers of cultural outputs with a disclaimer. In such a case, the information obligation would hence lose its informative function.

All in all, it seems reasonable to adopt a **wait-and-see attitude** before introducing an obligation to disclose the AI origin of any given creation in copyright. AI solutions are still in development. They raise legitimate concerns, but until this tool has reached a certain maturity, it appears difficult to envisage a well-tailored legislative solution. The definition of the relevant case for an information obligation today could reveal itself as outdated in only some years.

There may however be reasons for introducing such information obligation in other legislation than copyright. While horizontal AI regulations are going through the legislative process, the European legislature could consider adopting copyright-relevant information obligations in broader transparency requirements – depending on the adoption and evolution of AI solution in the cultural sector (such as the one in art. 52 of the proposal for an AI regulation, which concerns certain AI systems interacting with natural persons).

4. Summary and Conclusions

4.1. Conclusions on copyright data management in the European creative industries

In the ambition of the EU to become a leader in the data economy, a **well-functioning IP protection** system is a crucial framework condition. For creative industries, the effective protection and monetisation of their productions are of the utmost importance – not only for commercial reasons but also as an incentive to foster the creative output of individual creators. However, for the different players in the content value network, the **management of copyright-protected work**, including the identification of relevant rightholders, the negotiation of licensing arrangements, and the identification of pirated content, is **complex**. The extent of these complexities depends on the industrial context, the business (and creative) processes within a specific sector, and the legal and institutional framework. It seems, however, fair to say, that all creative industries – to a larger or smaller extent and despite important ongoing technical and standardisation developments which have brought important improvements for example with respect to interoperability of rights metadata systems – could benefit from new **technical advances** (for example enabling the swift **identification of the authorship, right ownership and availability** of content in the context of a digital economy). Metadata understood as “data about data” or “data about content” is instrumental in achieving this goal. This is specifically true in a digital world characterised by a value network in which highly complex use and re-use of digital assets take place and in which any actor can interact with any and all other stakeholders (for example in the context of a platform economy or user-generated content).

In this study, we have investigated the role of identification and rights metadata or “rights management information” as defined in the European Directive 2001/29 on the harmonisation of copyright in the information society as follows:

“Rights Management Information means any information provided by rightsholders which identifies the work or other subject matter [], the author or any other rightsholder, or information about the terms and conditions of use of the work or other subject-matter, and any numbers or codes that represent such information. [This] shall apply when any of these items of information is associated with a copy of, or appears in connection with the communication to the public of, a work or other subject matter []”⁵⁶⁷.

It has analysed the extent to which **different challenges related to rights metadata**, which have been discussed by various stakeholders – for example in the context of the Finnish Presidency of the Council of the European Union in 2019 – can be empirically substantiated. In doing so, we have taken stock of and acknowledged the many different initiatives that are already addressing these issues – be it standardisation processes or technical developments.

A main conclusion of the study is that **analyses of the status of rights metadata**, as well as their potential impacts, must take place at a granular level of analysis and must be context sensitive. Therefore, the study team adopted a sectoral approach to analysing metadata challenges. The team took a differentiated perspective on the music, film and

⁵⁶⁷ European Community (2001). “Directive 2001/29/EC of the European Parliament and of the Council on the harmonisation of certain aspects of copyright and related rights in the information society”. Article 7 §2.

publishing industries. Moreover, even within industries such as the music or publishing industry, differences with respect to the relevance of metadata problems became very clear. Metadata challenges are more prevalent in some areas, such as the digital news publishing sector, than in others, for example book publishing. In the former, issues such as a lack of rights metadata in photographs used on news websites lead to missing attribution of copyright ownership and therefore also challenges regarding the remuneration of creators. In the book publishing sector, and especially in the traditional brick-and-mortar distribution context, which was *not* the focus of this study, this seems not to be as problematic: authorship is usually clear and licensing negotiations are relatively straightforward. In addition, the question of the authoritativeness – i.e., the question of whether stakeholders can trust in the correctness of metadata – seems not to be a problem in this sector.

Similarly, in the music industry, the study team gathered evidence from different primary and secondary sources that rights metadata management and data quality has improved thanks to many ongoing initiatives, but that it still is an important industry issue with seems have significant commercial implications (for example, costs associated with retrospectively correcting erroneous metadata in older works). At the same time, stakeholders from the recorded music sector signalled that rights metadata issues are for them less important than the inaccuracy or lack of usage metadata provided by online platform services.

The sector and subsector-specific assessment of needs to improve the rights data framework is related to an impressive number of **ongoing industry initiatives**. This report lists highly important industry initiatives that improve the exchange and interoperability of metadata within an industry. The DDEX initiative and the new RDx, Cube and URights projects to interconnect CMO databases are examples in the music sector of an industry-driven approach to improving data interoperability. ONIX and DOI are other prominent examples in the publishing industry.

Overall, we discussed and highlighted four different “**pain points**” related to rights metadata. These are the costs of rights management, inefficient processes with respect to rights licensing, challenges concerning payments processes, and misappropriation or piracy of copyright-protected content.

In general, the study points out that, first, a much higher awareness of the importance of rights metadata and the potential benefits of an improved metadata situation is needed – especially among creators themselves. Although many initiatives are ongoing and significant efforts have been invested by, for example, CMOs or organisations such as WIPO on this topic, there remains much to do.

Improving the general **awareness** of the benefits rights metadata (as well as copyright rules in general) would be an important condition to making the rights metadata system work better (metadata education). Additionally, we conclude that the **potential of new technologies for metadata management** (for example Artificial Intelligence and Blockchain) has not yet been fully exploited. Finally, integrating the existing rights data frameworks in the sense of a **cross-sector rights data network could bridge gaps** between standard content identifiers such as ISRC, ISWC, ISBN or ISAN and digital manifestations of the content they denote. This would increase interoperability also between different media or content sectors. The ultimate objective of this endeavour would be to break the silos of different creative industries and improve the efficient rights data management and licensing across sectors. It could help to release even more of the digital potential of Europe’s creative sectors.

4.2. Conclusions on copyright and artificial intelligence in the field of creative industries

The development of AI solutions will likely be a game-changer in the cultural industry, raising several challenges for copyright and related rights. Currently, the deployment of these technologies in the four sectors investigated for this study (namely visual arts, music, audio-visual & film, and videogames) is still in its infancy, but new AI solutions are emerging at a swift pace, varying significantly from sector to sector. Mostly used for improving the production and marketing of cultural artefacts or for enhancing the consumer experience, some available AI tools assist or perform the process of creation itself, commonly considered as within the sole realm of humans, and not only repetitive or mechanical duties.

The reliance on AI technologies for or during the creative process might challenge copyright and/or related rights. The study has distinguished the upstream or input issues, from downstream or output issues related to the use of AI. On the input side, AI applications might require to be trained with large datasets of creative works enjoying copyright or related rights protection, and the question arises whether an authorisation is needed. On the output side, the applications might permit the generation of cultural artefacts whose protection under copyright or a related right raises new issues, among others because those artefacts are difficult to distinguish from human-created cultural content.

The copyright and related rights system should aim at ensuring a “fair balance” between various fundamental rights and between various interests. With the growing use of AI tools, the balance should be struck between the interests of the creators, the other holders of copyright and related rights, the developers of the new AI applications, the providers and users of innovative AI-based services, the end-users of cultural content and the general interest. This should be reflected in the way each issue of the copyright and related rights system is addressed: the definition of protected subject-matter and the interpretation of the conditions of protection (e.g. with regard to the exclusion of style or the investment-based related rights), the protected acts under the economic rights (in particular the reproduction and extraction rights) and the moral rights (with regard to the integrity issue) and the exceptions (in particular for TDM and research). Various challenges regarding issues of the copyright and related rights system have been identified and discussed with experts and stakeholders in the course of the research.

Input stage

Tensions are likely to grow between the copyright holders of works used as AI input and the developers and users of such AI solutions. This situation might require some adjustments to strike the right balance. Over-protection of the input could lead to a decrease in the development, deployment and use of AI solutions, and thus in the creation of new creative artefacts. Under-protection could prejudice the rightsholders, who may miss opportunities to monetise their rights and do not have the power to oppose the use of their protected subject matter.

The examination of AI solutions in the selected cultural sectors has revealed several challenges for copyright and the related rights.

The processing of datasets containing cultural content for training an AI solution is bound to involve protected **works** and, depending on the AI application, also protected **performances and sound and/or audio-visual recordings**. Each type of

protected object has its own conditions for protection but the need for massive volumes of data to train the AI algorithms entails that the AI-developer will usually not be able to assess whether the training dataset contains protected elements for which an authorisation should be sought, in addition to the **database itself** which may be protected in its own right.

The users would often be facing complex situations due to the possible cumulation of exclusive rights and a multitude of rightsholders. In the cases where the AI developer or user must clear the exclusive rights, the sheer **number of rightsholders** and **the fragmentation of exploitation** (according to the exclusive rights and/or territories involved) may complicate the acquisition of rights.

The challenges of massive processing of protected subject matter could be addressed by an adequate delineation of the reproduction and extraction rights, and by the adequate application of the exceptions for text and data mining.

The wide interpretation of the **reproduction right**, which has often been applied in a rather mechanical way, in combination with the volume of rights and the number of different rightsholders concerned, could create a considerable challenge for the developers or users of AI solutions in the cultural sector. The scope of the reproduction right may be evolving in the light of the recent decisions of the CJEU (in particular in *Pelham* and *CV-Online*) concerning the related rights of the phonogram producer and the sui generis database rights. This evolution to a more functional and teleological interpretation of what constitutes a reproduction (or an extraction/reutilisation) would be useful for the processing of cultural content in view of training AI solutions, where the purpose is not the true exploitation of the work or protected subject matter itself.

Although one should still wait for the complete transposition into national laws and for the application of the **TDM exceptions**, certain concerns have been expressed on (i) the scope of the TDM exceptions and (ii) the efficient exercise of the opt-out. It is recommended to safeguard the uniform application of the exceptions in the EU to reduce the differences between national laws and practices to a minimum. Some close monitoring seems important here.

In the same way, the **moral rights** in the national copyright laws might be invoked to prevent the use of protected works, even where such use is permitted under statutory exceptions to the economic rights. Depending on the future developments, it might be advisable to ensure some level of harmonisation regarding the exercise of the moral rights in view of the applicable exceptions, in particular the TDM exceptions. A more ambitious approach would be to consider (partial) harmonisation of moral rights.

Output stage

On the **output side**, the research, interviews and surveys conducted for the completion of the study indicate that the current policy trade-off within copyright law, according to which **no copyright exists in the absence of human creative choices**, should be maintained. In most cases, the AI output generated through the use of an AI solution does not take over the protected elements of works, performances or recordings, meaning that it does not constitute an infringement of copyright or the related rights under the current rules. An extension of copyright protection to the elements of an artistic style does not seem desirable at this stage.

Under the current copyright rules, in the absence of human creative choices, **autonomously generated AI output would not be protected under copyright**. Furthermore, no convincing reasons currently exist for extending copyright protection to such AI-generated output: creation, research and innovation into AI solutions and the resulting AI output, as well as the business model behind it, are not hampered by the absence of protection under copyright.

By contrast, the fact that an AI output is autonomously generated by an AI solution may, in certain jurisdictions, have as a consequence that the (human) **performance** of such AI output is not protected under the performers' rights. On the other hand, the fixation of such AI output and its performance could be eligible for protection under the **producers' rights**, since no other condition than fixation is explicitly posed for this related right. To avoid that some of those related rights are used to circumvent the current copyright policy trade-off, it might be advisable to revisit the conditions for protection under the related rights by (i) clarifying the relation between the protected "performance" and the protected "work" and (ii) revising the conditions for the producers' rights by making it express that some investment is required.

The closer AI solutions will come to producing cultural content that cannot be distinguished from human-created cultural content, the more the issue of **false authorship**, used to circumvent the absence of copyright protection for this type of output, will become a cause for concern. However, the study concludes that a restriction or abolishment of the **presumption of authorship**, in order to render false authorship claims more difficult, would be excessive and could have negative effects for the human creators on whom the burden of establishing authorship would lie. The study concludes that no **information obligation** concerning the use of an AI solution for the development of the work should be added within the copyright framework either. The scope of such legal obligation would be difficult to determine and its impact on the creators' artistic freedom and their personality rights might be disproportionate.

5. Annexes

5.1. Annex 1: Exemplary technical challenges related to the *acquis communautaire* on rights metadata – textual references

This annex contains the texts analysed in section 2.2.3 gathered here for the reader's convenience.

5.1.1. Copyright in the information society (EC/2001/29) – Article 7 §2⁵⁶⁸

Article 7 §2 of EC/2001/29 refers to obligations **concerning rights-management information** in the following form:

Rights Management Information means any information provided by rightsholders which identifies the work or other subject matter [], the author or any other rightsholder, or information about the terms and conditions of use of the work or other subject-matter, and any numbers or codes that represent such information. [This] shall apply when any of these items of information is associated with a copy of, or appears in connection with the communication to the public of, a work or other subject matter [].

In the official French translation of EC/2001/29 accessible on the EUR-Lex Portal, this paragraph is formulated as follows:

On entend par «information sur le régime des droits» toute information fournie par des titulaires de droits qui permet d'identifier l'œuvre ou autre objet protégé [], l'auteur ou tout autre titulaire de droits. Cette expression désigne aussi les informations sur les conditions et modalités d'utilisation de l'œuvre ou autre objet protégé ainsi que tout numéro ou code représentant ces informations. [Ceci] s'applique lorsque l'un quelconque de ces éléments d'information est joint à la copie ou apparaît en relation avec la communication au public d'une œuvre ou d'un objet protégé [].

And, the German version reads as:

[] der Ausdruck „Informationen für die Rechtewahrnehmung“ [bezeichnet] die von Rechteinhabern stammenden Informationen, die [] geschützten Werke oder Schutzgegenstände, den Urheber oder jeden anderen Rechteinhaber identifizieren, oder Informationen über die Modalitäten und Bedingungen für die Nutzung der Werke oder Schutzgegenstände sowie die Zahlen oder Codes, durch die derartige Informationen ausgedrückt werden. [Dies] gilt, wenn irgendeine der betreffenden Informationen an einem Vervielfältigungsstück eines Werks oder eines sonstigen Schutzgegenstands [] angebracht wird oder im Zusammenhang mit der öffentlichen Wiedergabe eines solchen Werks oder Schutzgegenstands erscheint.

Transposition into the French Copyright Act – Code de la propriété intellectuelle – Article L331-11⁵⁶⁹

⁵⁶⁸ Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, O.J. L 167, 22 June 2001, p. 10–19

⁵⁶⁹ See https://www.legifrance.gouv.fr/codes/texte_lc/LEGITEXT000006069414

In the French Copyright act (Loi n° 2006-961 du 1er août 2006 (transposition of the Directive EC/2001/29) modified by Loi n° 2019-775 du 25 juillet 2019 (transposition of the Directive EU/2019/790) the topic is addressed as follows:

*On entend par information sous forme électronique toute information fournie par un titulaire de droits qui permet d'identifier une œuvre, **une interprétation, un phonogramme, un vidéogramme, un programme, une publication de presse ou un titulaire de droit**, toute information sur les conditions et modalités d'utilisation d'une œuvre, d'une interprétation, d'un phonogramme, d'un vidéogramme, d'un programme ou d'une publication de presse, ainsi que tout numéro ou code représentant tout ou partie de ces informations.*

*Conformément à l'article 14 de la loi n° 2019-775 du 24 juillet 2019 tendant à créer un droit voisin au profit des agences de presse et des éditeurs de presse, cette disposition **ne s'applique pas** aux publications de presse publiées pour la première fois avant la date d'entrée en vigueur de la directive du Parlement européen et du Conseil sur le droit d'auteur dans le marché unique numérique.*

Transposition into the Urheberrechtsgesetz (UrhG) – German Copyright Act, §95c Schutz der zur Rechtswahrnehmung erforderlichen Informationen⁵⁷⁰

The German text in the UrhG reads as follows:

[Von Rechtsinhabern stammende] Informationen für die Rechtswahrnehmung im Sinne dieses Gesetzes sind elektronische Informationen, die Werke oder andere Schutzgegenstände, den Urheber oder jeden anderen Rechtsinhaber identifizieren, Informationen über die Modalitäten und Bedingungen für die Nutzung der Werke oder Schutzgegenstände sowie die Zahlen und Codes, durch die derartige Informationen ausgedrückt werden.

5.1.2. Collective management of copyright (EU/2014/26) – Article 24⁵⁷¹

The English version of Article 24 of EU/2014/26 reads as follows:

Capacity to process multi-territorial licences

1. Member States shall ensure that a collective management organisation which grants multi-territorial licences for online rights in musical works has sufficient capacity to process electronically, in an efficient and transparent manner, data needed for the administration of such licences, including for the purposes of identifying the repertoire and monitoring its use, invoicing users, collecting rights revenue and distributing amounts due to rightsholders.

2. For the purposes of paragraph 1, a collective management organisation shall comply, at least, with the following conditions:

(a) to have the ability to identify accurately the musical works, wholly or in part, which the collective management organisation is authorised to represent;

⁵⁷⁰ See <http://www.gesetze-im-internet.de/urhlg/>

⁵⁷¹ Directive 2014/26/EU of the European Parliament and of the Council of 26 February 2014 on collective management of copyright and related rights and multi-territorial licensing of rights in musical works for online use in the internal market, O.J. L 84, 20 March 2014, p. 72–98

(b) to have the ability to identify accurately, wholly or in part, with respect to each relevant territory, the rights and their corresponding rightsholders for each musical work or share therein which the collective management organisation is authorised to represent;

(c) to make use of unique identifiers in order to identify rightsholders and musical works, taking into account, as far as possible, voluntary industry standards and practices developed at international or Union level;

(d) to make use of adequate means in order to identify and resolve in a timely and effective manner inconsistencies in data held by other collective management organisations granting multi-territorial licences for online rights in musical works

The French translation of the Directive is formulated as follows:

1. Les États membres veillent à ce qu'un organisme de gestion collective qui octroie des licences multiterritoriales de droits en ligne sur des œuvres musicales soit doté d'une capacité suffisante pour traiter par voie électronique, de manière transparente et efficace, les données requises pour la gestion desdites licences, y compris aux fins de l'identification du répertoire et du contrôle de l'utilisation de ce dernier, pour la facturation aux utilisateurs, pour la perception des revenus provenant des droits et pour la distribution des sommes dues aux titulaires de droits.

2. Aux fins du paragraphe 1, un organisme de gestion collective remplit au minimum les conditions suivantes:

(a) avoir la capacité d'identifier avec précision les œuvres musicales, en tout ou en partie, que l'organisme de gestion collective est autorisé à représenter;

(b) avoir la capacité d'identifier avec précision, en tout ou en partie, sur chaque territoire concerné, les droits et les titulaires de droits correspondants pour chaque œuvre musicale ou partie d'œuvre musicale que l'organisme de gestion collective est autorisé à représenter;

(c) faire usage d'identifiants uniques pour identifier les titulaires de droits et les œuvres musicales, en tenant compte, dans la mesure du possible, des normes et pratiques sectorielles volontaires élaborées à l'échelle internationale ou au niveau de l'Union;

(d) recourir à des moyens adéquats pour déceler et lever, avec rapidité et efficacité, les incohérences dans les données détenues par d'autres organismes de gestion collective qui octroient des licences multiterritoriales de droits en ligne sur des œuvres musicales.

And thirdly, the German translation:

1. Die Mitgliedstaaten stellen sicher, dass die Organisationen für die kollektive Rechtewahrnehmung, die Mehrgebietslizenzen für Online-Rechte an Musikwerken vergeben, über ausreichende Kapazitäten zur effizienten und transparenten elektronischen Verarbeitung der für die Verwaltung dieser Lizenzen erforderlichen Daten verfügen, darunter zur Bestimmung des Repertoires und Überwachung von

dessen Nutzung, zur Ausstellung von Rechnungen, zur Einziehung von Einnahmen aus der Rechtenutzung und zur Verteilung der den Rechtsinhabern zustehenden Beträge.

2. Für die Zwecke des Absatzes 1 müssen die Organisationen für die kollektive Rechtewahrnehmung mindestens folgende Voraussetzungen erfüllen:

(a) sie müssen über die Fähigkeit zur korrekten Bestimmung der einzelnen Musikwerke – vollständig oder teilweise –, die die Organisationen für die kollektive Rechtewahrnehmung repräsentieren dürfen, verfügen;

(b) sie müssen hinsichtlich eines jeden Musikwerks oder Teils eines Musikwerks, das die Organisation für die kollektive Rechtewahrnehmung repräsentieren darf, über die Fähigkeit verfügen, die Rechte – vollständig oder teilweise und in Bezug auf jedes Gebiet – sowie den zugehörigen Rechtsinhaber zu bestimmen;

(c) sie müssen eindeutige Kennungen verwenden, um Rechtsinhaber und Musikwerke zu bestimmen, unter möglichst weitgehender Berücksichtigung freiwilliger branchenüblicher Standards und Praktiken, die auf internationaler oder Unionsebene entwickelt wurden;

(d) sie müssen geeignete Mittel verwenden, um Unstimmigkeiten bei den Daten im Besitz anderer Organisationen für die kollektive Rechtewahrnehmung, die Mehrgebietslizenzen für Online-Rechte an Musikwerken vergeben, **rasch** und wirksam zu erkennen und zu **beheben**.

Transposition into the French Copyright Act – Code de la propriété intellectuelle – Article L325-2

Ordonnance n°2016-1823 du 22 décembre 2016 stipulates:

I.- Les organismes de gestion collective peuvent, dans des conditions fixées par décret en Conseil d'Etat, octroyer des autorisations d'exploitation multiterritoriales de droits en ligne sur des œuvres musicales sous réserve qu'ils disposent des moyens leur permettant de traiter par voie électronique les données nécessaires à la gestion de ces autorisations.

In turn, the **Verwertungsgesellschaftengesetz (VGG) – German Collective Management Act, §61 Besondere Anforderungen an Verwertungsgesellschaften**⁵⁷² formulates:

(1) Die Verwertungsgesellschaft muss über ausreichende Kapazitäten verfügen, um die Daten, die für die Verwaltung von gebietsübergreifend vergebenen Online-Rechten an Musikwerken erforderlich sind, effizient und transparent elektronisch verarbeiten zu können.

(2) Die Verwertungsgesellschaft muss insbesondere

⁵⁷² See <https://www.gesetze-im-internet.de/vgg/>

1. jedes Musikwerk, an dem sie Online-Rechte wahrnimmt, korrekt bestimmen können;
2. für jedes Musikwerk und jeden Teil eines Musikwerks, an dem sie Online-Rechte wahrnimmt, die Online-Rechte, und zwar vollständig oder teilweise und in Bezug auf jedes umfasste Gebiet, sowie den zugehörigen Rechtsinhaber bestimmen können;
3. eindeutige Kennungen verwenden, um Rechtsinhaber und Musikwerke zu bestimmen, unter möglichst weitgehender Berücksichtigung der freiwilligen branchenüblichen Standards und Praktiken, die auf internationaler Ebene entwickelt wurden;
4. geeignete Mittel verwenden, um Unstimmigkeiten in den Daten anderer Verwertungsgesellschaften, die gebietsübergreifend Online-Rechte an Musikwerken vergeben, **unverzüglich** und wirksam erkennen und **klären** zu können.

5.2. Annex 2: List of sources for the literature review

- Bodó, Balázs and Gervais, Daniel and Quintais, João Pedro: Blockchain and smart contracts: the missing link in copyright licensing? (September 25, 2018). *International Journal of Law and Information Technology*, Volume 26, Issue 4, Pages 311–336.
- Butler, Susan P.: Collective Rights Management Practices Around the World - a survey of CMO practices to reduce the occurrence of unclaimed royalties in musical works (April 2020). Commissioned by the U.S. Copyright Office.
- Council of the European Union: Developing the Copyright Infrastructure-Stocktaking of work and progress under the Finnish Presidency (December 20, 2019). Note from the presidency to the working party on intellectual property (Copyright), 15016/19.
- Europe Economics and Guibault, Lucie and Salamanca, Olivia and van Gompel, Stef: Remuneration of authors and performers for the use of their works and the fixations of their performances (June 2015). Final Report for the European Commission.
- European Commission: Access to rights data via identification technologies optimisation - Periodic Reporting for period 1 (30 June 2018).
- European Commission: Making the most of the EU's innovative potential - An intellectual property action plan to support the EU's recovery and resilience (November 25, 2020). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Region, COM(2020) 760 final.
- European Commission: Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act) (November 25, 2020). COM(2020) 767 final.
- French Ministry of Culture: Mission Report - towards more effectiveness of copyright law on online content sharing platforms: overview of content recognition tools and possible ways forward (January 29, 2020). Higher Council on Literary and Artistic Property, National Centre for Cinema and the moving Image, High Authority for the Dissemination of Works and The Protection of Rights on the Internet.
- German Federal Ministry of Justice and Consumer Protection: Erklärung der Bundesrepublik Deutschland zur Richtlinie über das Urheberrecht und verwandte Schutzrechte im Digitalen Binnenmarkt; insbesondere zu Artikel 17 der Richtlinie (April 15, 2019).
- Gervais, Daniel J.: Exploring the Interfaces Between Big Data and Intellectual Property Law (March 26, 2019). 10 (2019) *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* (JIPITEC) 22, Vanderbilt Law Research Paper No. 19-36.
- Hardjono, Thomas and Howard, George and Scace, Eric and Chowdury, Mizan and Novak, Lucas and Gaudet, Meghan and Anderson, Justin and d'Avis, Nocolle and Kulis, Christopher and Sweeney, Edward and Vaughan, Chander: Towards an Open and Scalable Music Metadata Layer (November 20, 2019). Draft.
- Jessop, Paul: Functional Requirements for the International Music Registry Analysis Report for WIPO. Analysis Report on Functional Requirements for IMR.
- Lyons, Paul and Hyojung, Sun and Collopy, Dennis and Curran, Keven: Music 2025 – The Music Data Dilemma: issues facing the music industry improving data management (June 2019). The Intellectual Property Office.
- Megías, David and Kuribayashi, Minoru and Qureshi, Amna: Survey on Decentralized Fingerprinting Solutions: Copyright Protection through Piracy Tracing (April 3, 2020). *Computers* 9, no. 2: 26.

Motion Picture Laboratories: The Evolution of Media Creation - A 10-Year Vision for the Future of Media Production, Post and Creative Technologies (July 2019).

Nordemann, Jan Bernd: Art. 17 DSMCD a Class of Its Own? How to Implement Art. 17 Into the Existing National Copyright Acts – Also a Comment on the Recent German Discussion Draft (July 12, 2020). German version published in GRUR (Gewerblicher Rechtsschutz und Urheberrecht) 2020, 569-576.

Paskin, Norman: The Digital Object Identifier: From Ad Hoc to National to International (July 2015). The Critical Component: Standards in the Information Exchange Environment (2015), edited by Todd Carpenter, published by ALCTS (American Library Association Publishing).

Rendas, Tito: Copyright, Technology and the CJEU: An Empirical Study (October 12, 2017). 49(2) IIC – International Review of Intellectual Property and Competition Law 153 (2018).

Senftleben, Martin and Margoni, Thomas and Antal, Daniel and Bodó, Balázs and Gompel, Stef van and Handke, Christian and Kretschmer, Martin and Poort, Joost and Quintais, João and Schwemer, Sebastian Felix: Ensuring the Visibility and Accessibility of European Creative Content on the World Market - The Need for Copyright Data Improvement in the Light of New Technologies and the Opportunity Arising from Article 17 of the CDSM Directive (February 12, 2021).

Xalabarder Raquel: International legal study on implementing an unwaivable right of audiovisual authors to obtain equitable remuneration for the exploitation of their works (May 2018). CISAC – International Confederation of Societies of Authors and Composers.

5.3. Annex 3: List of current and ongoing initiatives in data interoperability within the content rights infrastructure

Table 2: Overview on current and ongoing initiatives in data interoperability within the content rights infrastructure

Ref	Initiative	Summary (usually based on website information)			
A	Working groups and reports	<i>Groups, consultations and reports seeking solutions for rights infrastructure</i>	<i>Primary Content/Media</i>	<i>Primary Sector</i>	<i>Data scope</i>
A1	Developing the Copyright Infrastructure Finnish Ministry of Culture and Education	<p>The Finnish Ministry of Culture and Education has launched the project "Developing the Copyright Infrastructure 2020-2022". Subgroups for the copyright infrastructure meet in different configurations to discuss topical issues related to practices around identifiers in the creative industries, metadata, technology and infrastructure development, and legal issues linked to data and intellectual property. The progress is reported at the meetings of the Finnish Copyright Delegation to discuss next steps. The discussions support work at EU and global level.</p> <p>The following aspects are interesting to mention:</p> <ul style="list-style-type: none"> • The EU Communication IP Action Plan mentions the need to develop a functioning copyright infrastructure. The Copyright Infrastructure is defined in the Action Plan as a set of rules, technologies and institutions that provide a framework for information management practices in the creative sectors. • During its Presidency of the Council of the EU in 2019, Finland emphasized the need to develop the copyright infrastructure, i.e., the practices concerning the submission of data on works and authors, incl. identifiers, tags, standards and data sharing formats. Finland also aims to 	Any	Any	Parties Content Rights Awareness and Understanding

		support the development of a human centred, balanced and thriving European data economy. As part of this work, the growth and competitiveness of the creative industries were also taken into account, e.g., at the Data Economy Conference in Helsinki on 25-26 November 2019. The data principles promoting the sharing and use of data were developed together with the assistance of the industries and Sitra.			
A2	Music 2025 UK Intellectual Property Office	<p>The British Intellectual Property Office commissioned a report on the music data dilemma. The findings of this report reflect the views and opinions gathered from over 50 interviews from high profile music industry representatives.</p> <p>The report recommended:</p> <ul style="list-style-type: none"> • Education and awareness - improving the standards of data input at all points to ensure correct attributions in the digital space, • Collaboration – seize opportunities to improve communication across a diverse and fractured sector, • Interoperability – discourage all players to develop their own datasets and databases in isolation, • Governance – seize opportunities for the music industry to learn from the banking sector in terms of governance, to ensure adherence to data standards. 	Music	Any	Any Awareness and Understanding
A3	Music Explained Copyright	Music Copyright Explained is a free user-friendly guide to how music copyright works in the UK. It explains how copyright gives music-makers control over the songs and recordings they create.	Music	Any	Any

	UK Intellectual Property Office	It talks through how music-makers and the music industry generate income out of their music rights. It outlines all the key things music-makers and other creators need to know about music copyright and licensing. Among other things, it presents the top five music copyright facts, the top five tips for music makers, and the top five tips for users of music – all explained in detail.			Awareness and Understanding
A4	Reversion Rights in the European Member States Dr Ula Fergal, Postdoctoral Researcher at CREATE, University of Glasgow	This working paper maps provisions allowing authors and performers to reclaim their rights (reversion rights) which are currently or were historically a part of the national laws of the EU Member States. The impulse came from the introduction of the right of revocation in Article 22 of the Directive EU/2019/790 on copyright in the Digital Single Market, a reversion right following a use-it-or-lose-it logic. General provisions, applicable to all types of works and agreements are not a rule. Reversion rights often tackle narrow sector-specific issues, and general provisions tend to factor in specificities of different types of works, such as differences in their commercial lifespan. The provisions do not always lead to the termination of agreements. Since most of the rights are not brought to effect automatically but require creator's action to make any changes to the contractual relationship, there is a space for renegotiation of existing contracts but also potentially for blacklisting. The procedure and formalities which authors and performers need to observe are rarely addressed, leaving creators without a guidance on how to exercise their rights. Termination is only one of the options offered by the existing reversion provisions. Others include the change of exclusive into non-exclusive assignments and an authorisation to perform acts otherwise	Any	Legal	Rights

		reserved to other parties. Digital uses are not reflected in the existing provisions which sometimes date back nearly a century. There is no guidance yet about what may constitute a lack of exploitation in the digital context, but this confirms the importance of exhaustive (identifying both authors and rightsholders) and up-to-date metadata mentioned earlier.			
A5	Working Party on Intellectual Property Council of the European Union	<p>On 20 December 2019, under the Finnish presidency, the Council issued an important stocktaking document on developing the Copyright Infrastructure. It states a clear objective, unleashing the digital potential of all European creative sectors through effective metadata, improved licensing efficiency, and automated revenue distribution. Moreover, the recommended approach includes interoperable identifiers of works and rightsholders, attribution of authorship, dispute resolution, standardised rights management information, declaration of rights, governance, Collective Management Organisations, data exchange standards, connection with digital copies of works, and usage data.</p> <p>It is worth noting that the document refers to EU/2019/790, EC/2001/29 Art 7.2, Rec 55, EU/2014/26, EC/2004/48 Art 5.2, and mentions the standard stakeholder identifiers Interested Parties Information (IPI), Instant Payment Notification (IPN), International Standard Name Identifier (ISNI), and Open Researcher and Contributor ID (ORCID), the standard content identifiers Digital Object Identifier (DOI), International Standard Book Number (ISBN), International Standard Serial Number (ISSN), International Standard Musical Work Code (ISWC), International Standard Recording Code (ISRC), International Standard Audiovisual Number (ISAN), International</p>	Any	Any	Parties Content Rights Awareness and Under- standing

		Standard Text Code (ISTC), and the initiatives Access to Rights Data via Identification Technology Optimisation (ARDITO), Digital Data Exchange (DDEX), Repertoire Data Exchange (RDx), and Cube , most of them presented hereafter.			
B	Data frameworks	<i>Generic frameworks and schemas for interoperability for content and rights data</i>	<i>Primary Content/Media</i>	<i>Primary Sector</i>	<i>Data scope</i>
B1	<indecs> Interoperability of Data in E- commerce Systems (with the support of the European Commission)	<p>The Interoperability of Data in E-commerce Systems <indecs> project was an international initiative of owners of intellectual property rights, that developed a framework of metadata standards to support networked commerce based on intellectual property. It was supported by the European Commission's Info 2000 programme and produced –</p> <ul style="list-style-type: none"> • A complete generic data model for intellectual property trading in a networked environment, • Mapping of other metadata initiatives to this common model, • A specification for the development of a <i>metadata registry</i> which would make it possible for applications to use this mapping to make different metadata schemes interoperable, • Specification for the linking of <i>person identifiers</i>, an essential part of the infrastructure, • A Resource Description Framework (RDF) model of the generic data model, • Implementation guides (managerial and technical) for those who need to work with the model, 	Any	Any	Parties Content Rights

		<ul style="list-style-type: none"> Proposals to appropriate standards bodies for formal standardisation. <p>The <indecs> project was very successful and paved the way to major hereunder-described initiatives which are still ongoing, such as ONIX, DDEX, DOI or MDDF described hereafter. The sum of these initiatives represents an intellectual and network capital that must be considered for the options to move forward and solve copyright metadata issues.</p>			
B2	Distributed Trust Rights Framework Digiciti Networks OÜ ⁵⁷³	<p>Digiciti Networks assembles new, breakthrough technologies that focus on the fundamentals of rights management, namely content identifiers, stakeholder identifiers, metadata associations, authoritative assertions, and the use of trusted, multi-party, distributed, dynamic data management systems to create and share Rights Management Information. The Distributed Trusted Rights Framework, provides trustworthy Rights Information, provenance, authenticity, and compliance with agreements, and enables the automated distribution of content and associated rights compensation. The approach is minimally prescriptive but maximally supportive and inclusive. It allows many solutions to be used while enabling numerous ways in which individuals and organisations can cooperate in originating, enriching, governing, and distributing trusted Information, helping streamline current processes and trigger innovative businesses.</p>	Any	Any	Parties Content Rights Authority

⁵⁷³ Disclosure: Philippe Rixhon is the Research & Innovation Director at Digiciti Networks OÜ.

B3	European Blockchain Services Infrastructure (EBSI) European Blockchain Partnership	From its start, the European Blockchain Services Infrastructure (EBSI) has been busy with four use cases relevant to rights management information: notarisation (of digital works or assets), certification, self-sovereign identity (of authors, rightsholders or other stakeholders) and trusted data exchange. The European Blockchain Partnership has received the proposals of consortia asked to build a performing EBSI. The consortia were asked to test the developed distributed ledger technologies and evaluate the level of performance and the improvements achieved by the new capacities at hand of one or two use cases. The “EU wide management of IP rights (like patents, trademarks, copyrights), including also the management of copyrights that can be directly associated to digital content in near real-time” was one of them (see the EU Blockchain procurement documents (p. 10) available at https://etendering.ted.europa.eu/cft/cft-document.html?docId=81917)	Any	Any	Rights Authority
B4	Experiments towards a copyright infrastructure Estonian government	The Estonian Government – through its platform AccelerateEstonia – has launched an initiative to perform practical experiments leading to the implementation of a viable copyright infrastructure, and to identify and launch a sustainable operating model for that infrastructure and its core services in Estonia. A prototype will be piloted with the Estonian music ecosystem in autumn 2021. The roles of the public sector in defining and assuring the governance will be investigated, as well as the benefits and incentives of all stakeholders.	Any Music prototype	Any	Parties Content Rights Awareness and Under- standing Authority
B5	Functional Requirements for	FRBR (for which the link here is to the final report as there is no ongoing website) introduced the critical Item-Manifestation-Expression-Work distinctions for content into the global	Any	Biblio- graphic	Parties Content

	Bibliographic Records (FRBR) International Federation of Library Associations	bibliographic community, paralleling some of the work of the <indecs> project. As indecs has influenced much standards work in the commercial sector, FRBR has since done the same in the bibliographic world, including on the Resource Description and Analysis (RDA) initiative on library cataloguing rules. The analysis is critical for rights data interoperability between the bibliographic and commercial sector, as the two domains have different criteria for identifying distinct creations, and therefore for the application of rights data.			Rights
B6	Linked Content Coalition (LCC)	The Linked Content Coalition (LCC) is a not-for-profit global consortium of standards bodies and registries. The purpose of the LCC is to facilitate and expand the legitimate use of content in the digital network through the effective use of interoperable identifiers and metadata. LCC members create and manage data standards associated with content of one or more types, particularly for identifiers, metadata and messaging. The LCC supports interoperability between the computer systems of any and all legitimate participants in the digital network, including creators, rightsholders, publishers, aggregators, rights and content exchanges, retailers, consumers, cultural institutions (including libraries, museums and archives) and their agents and associations. Participation may be on any scale, from that of private individuals to multinational organisations.	Any	Any	Parties Content Rights
B7	Movielabs Digital Distribution Framework (MDDF) Motion Picture Laboratories Inc.	For automation of digital workflows and supply chain efficiency, MovieLabs recommends adoption of a suite of compatible standards and specifications. They cover core aspects of online distribution, including identification, metadata, avails, asset delivery, and reporting. Developed and delivered through	Audio-visual	Comm.	Content Rights

		industry collaboration, these standards and technologies enable automation, cost reduction, and improved consumer experiences across the industry. Collectively, they call these the MovieLabs Digital Distribution Framework (MDDF). MDDF offers Unique Identification, Scalable and Localizable Metadata, Precise Licensing (Avails and Title Lists), Efficient Delivery, Delivery Process Support, Engaging Digital Extras and Reporting. It is based on a cohesive set of specifications that support business models and promote efficiency across interrelated entertainment workflows.			
C	Identifier standards	<i>Standards for identifiers of parties or content</i>	<i>Primary Content/ Media</i>	<i>Primary Sector</i>	<i>Data scope</i>
C1	Universal Resource Identifier (URI)	A Universal Resource Identifier (URI) is a member of this universal set of names in registered name spaces and addresses referring to registered protocols or name spaces. A Uniform Resource Locator (URL), defined elsewhere, is a form of URI which expresses an address which maps onto an access algorithm using network protocols. Existing URI schemes which correspond to the (still mutating) concept of IETF URLs are listed here. The Uniform Resource Name (URN) debate attempts to define a name space (and presumably resolution protocols) for persistent object names. This area is not addressed by this document, which is written in order to document existing practice and provide a reference point for URL and URN discussions.	Any	Any	Any

C2	International Standard Audio-visual Number (ISAN) ISAN Agency	<p>The ISAN (International Standard Audiovisual Number) is a voluntary numbering system for the identification of audiovisual works. It provides a unique, internationally recognized and permanent reference number for each audiovisual work registered in the ISAN system.</p> <p>An ISAN uniquely distinguishes one audiovisual work from all other audiovisual works. Other methods of identifying audiovisual works, such as by title, can result in confusion about the specific work being referenced. For example, one title can be very similar to another. Titles also change when a work is distributed beyond its country (or countries) of origin and the title is translated into other languages.</p> <p>Because each ISAN is a unique number that is permanently assigned to an audiovisual work, it can identify that work across national boundaries and language barriers.</p> <p>As a unique identifier, the ISAN is useful in a wide range of computerized applications, particularly those which involve databases or the exchange of information about audiovisual works. Some of its possible applications are:</p> <ul style="list-style-type: none"> audiovisual cataloguing; to assist collecting societies in the management of rights; to reduce unauthorized use of protected content. <p>The issuance of an ISAN is in no way related to any process of copyright registration, nor does the issuance of an ISAN provide evidence of the ownership of rights in an audiovisual work.</p>	Audio-visual	Any	Content
----	---	---	--------------	-----	---------

C3	International Standard Book Number (ISBN) ISBN Agency	<p>An ISBN is an International Standard Book Number. ISBNs were 10 digits in length up to the end of December 2006, but since 1 January 2007 they now always consist of 13 digits. ISBNs are calculated using a specific mathematical formula and include a check digit to validate the number.</p> <p>An ISBN is essentially a product identifier used by publishers, booksellers, libraries, internet retailers and other supply chain participants for ordering, listing, sales records and stock control purposes. The ISBN identifies the registrant as well as the specific title, edition and format.</p> <p>ISBNs are assigned to text-based monographic publications (i.e. one-off publications rather than journals, newspapers, or other types of serials). Any book made publicly available, whether for sale or on a gratis basis, can be identified by ISBN. In addition, individual sections (such as chapters) of books or issues or articles from journals, periodicals or serials that are made available separately may also use the ISBN as an identifier.</p>	Books	Any	Content
C4	International Standard Content Code (ISCC) The ISCC Foundation ⁵⁷⁴ <i>Update: ISCC is a "working draft" at ISO and not yet approved as</i>	<p>The ISCC is a universal identifier for multiple generic media-types (text, image, audio, video). It is a lightweight and similarity-preserving fingerprint designed for digital content to identify content in decentralized and networked environments across the creative industries (journalism, books, music, film, etc.). It is free, open-source and transparent. ISCC identifiers are generated algorithmically from the content itself. Content</p>	Any	Any	Content Authority

⁵⁷⁴ Disclosure: Philippe Rixhon is a member of the Advisory Board of the ISCC Foundation.

	<p>an international standard.</p>	<p>files are processed to build the identifier. The ISCC does not have to be manually assigned, neither does it have to be carried around or embedded within the content. The content itself is the source and authority of the ISCC. The ISCC is a unique, hierarchically structured, composite identifier. It is built from a generic and balanced mix of content-derived, locality-sensitive, and similarity-preserving hashes generated from metadata and content.</p> <p>The ISCC is used in the following projects –</p> <p>Content Blockchain (<i>The ISCC Foundation and partners</i>), a project run as an open, non-profit initiative. The Content Blockchain Network is open for anyone to create and develop new and innovative applications and tools that can connect to the distributed ledger. The initiative provides the foundational tools and the infrastructure for a new content and media ecosystem based on blockchain technology. These include chain-agnostic standards for content identification (ISCC), smart licences and transaction models. Content Blockchain supports data streams, native tokens and on-chain governance.</p> <p>Open Content Certification Protocol (<i>Posth Werk BV</i>), the OCCP suggests a process to create and verify content certificates by using open, content-derived, decentralised content identifier technology on public blockchain networks. It is based on the International Standard Content Code (ISCC) and the decentralised cross-chain registry protocol. The goal of the certification process is to create trust in assertions, claims and the authenticity of the original content, and to ensure</p>			
--	-----------------------------------	--	--	--	--

		accountability of entities, even if they must or prefer to remain pseudonymous.			
C5	International Standard Recording Code (ISRC) ISRC Agency	<p>The International Standard Recording Code (ISRC) enables sound recordings and music videos to be uniquely and permanently identified. ISRC helps to avoid ambiguity among recordings and simplifies the management of rights when recordings are used across different formats, distribution channels or products. The ISRC for a recording remains a fixed point of reference when the recording is used across different services, across borders, or under different licensing deals.</p> <p>ISRC identifies sound recordings and music videos. ISRC is not used to identify compositions/musical works (typically identified by ISWC), music products or performers.</p>	Sound recordings	Any	Content
C6	International Standard Serial Number (ISSN) ISSN Agency	<p>An ISSN is an 8-digit code used to identify newspapers, journals, magazines and periodicals of all kinds and on all media—print and electronic.</p> <p>An ISSN (International Standard Serial Number) identifies all continuing resources, irrespective of their medium (print or electronic), including newspapers, annual publications (reports, directories, lists, etc.), journals, magazines, collections, websites, databases, blogs, etc.</p> <p>In many countries, an ISSN is mandatory for all publications subject to the legal deposit.</p> <p>The ISSN takes the form of the acronym ISSN followed by two groups of four digits, separated by a hyphen. The eighth digit is a check digit calculated according to a modulus 11 algorithm on the</p>	Serial publications	Any	Content

		<p>basis of the 7 preceding digits; this eighth control digit may be an "X" if the result of the computing is equal to "10", in order to avoid any ambiguity (eg: ISSN 0317-8471, ISSN 1050-124X).</p> <p>The ISSN role is to identify a publication. It is a digital code without any intrinsic meaning.</p>			
C7	International Standard Textual Work Code (ISTC)	<p>The International Standard Text Code (ISTC) is a numbering system for the unique identification of text-based works; the term "work" can refer to any content appearing in conventional printed books, audiobooks, static e-books or enhanced digital books, as well as content which might appear in a newspaper or journal. The ISTC provides sales analysis systems, retail websites, library catalogues and other bibliographic systems with a method of automatically linking together publications of the "same content" and/or "related content", thus improving discoverability of products and efficiencies. An ISTC number is the link between a user's search for a piece of content and the ultimate sale or loan of a publication.</p> <p>2017, the International ISTC Agency ceased operation. After becoming aware of this, a public call for a new host to assume the Registration Authority responsibilities was circulated, which unfortunately received no responses. As a result, the ISO Technical Subcommittee (ISO TC 46/SC 9) responsible for the International Standard Text Code (ISTC) - ISO 21047:2009 withdrew the standard in September 2021.</p>	Textual works	Any	Content
D	Identifier standards with metadata	<i>Standards for identifiers of parties or content for which metadata registration is required</i>	Primary Content/ Media	Primary Sector	Data scope

D1	Digital Identifier (DOI) International Foundation	Object DOI	<p>The DOI system (ISO 26324) provides an infrastructure for persistent unique identification of objects of any type over the Internet. The governance body of the DOI system is the DOI Foundation, that is also the ISO Registration Authority for the standard. The DOI system is implemented through a federation of registration agencies coordinated by the DOI Foundation.</p> <p>The DOI system comprises the unique persistent identifier – the DOI name – assigned to an object (for example a piece of content) and the resolution system that enables the creation of a persistent network link to current information about the identified object, including where the object, or information about it, can be found on the Internet. This is achieved by binding the DOI name to metadata about the identified object and to a URL, indicating where the DOI should be resolved to on the Internet. A DOI name can resolve to some form of access to information (usually in form of a metadata record) or service (including licensing services) related to which the DOI refers or to the object itself. While information about an object can change and be updated over time, its DOI name will not change, thus links and data retrieval methods based on DOI won't change. Referring to an online document by its DOI provides a more stable link than simply using its URL alone, as also in cases where the URL associated to a DOI name no longer works, the broken link can be curated by simply associating a new URL to the DOI name. Ensuring the persistence of the DOI resolution is one of the core missions of the all the stakeholders of the DOI system, users, publishers, registrants, registration agencies and the DOI Foundation alike.</p>	Any	Any	Content
----	---	----------------------	---	-----	-----	---------

		Traditionally DOIs are in wide use to identify academic publications, such as journal articles, research reports and official publications, however over time DOI usage has extended to other content types such as datasets, government information, movies (EIDR) and other information resources. Applications of the DOI system are also emerging in other sectors, not related to publications. As a result of the ARDITO project (see below), the ARDI is a DOI assigned to a “rights declaration” which resolves to a right metadata record and/or to licensing services. The DOI system enables the construction of automated services, transactions and optimisations of operations along supply and value chains.			
D2	Entertainment ID Registry (EIDR) , using DOI US not-for-profit membership corporation	<p>The emergence of digital technologies has transformed every aspect of the professional audio-visual supply chain from content creation and post-production to distribution and consumption and created new opportunities for all stakeholders. With these opportunities also come challenges, such as more complex value chain interactions and an explosion in the number of assets relevant to commerce.</p> <p>Effective monetisation of these assets through an increasing number of distribution channels requires a widely adopted industry standard universal identifier registry that supports the full range of asset types and relationships between assets. The Entertainment Identifier Registry Association (EIDR) is established to provide this foundational service to all industry participants at low cost. It is a not-for-profit industry association that was founded by MovieLabs, CableLabs, Comcast and TiVo to meet a crucial need across the entertainment supply chain for universal identifiers for a broad array of audio-visual objects.</p>	Audio-visual	Comm.	Content

		These founding members are also members of the Board of Directors which governs the registry.			
D3	Entertainment ID Registry (EIDR) , using DOI US not-for-profit membership corporation and International Standard Audio-visual Number (ISAN) ISAN Agency	<p>The Entertainment ID Registry Association (EIDR) and the International Standard Audiovisual Number International Agency (ISAN-IA) launched a dual registration service, a result of the European Commission's Audiovisual Standard ID policy aimed at boosting opportunities for audiovisual arts, entertainment, information, and archival management.</p> <p>Audiovisual standard IDs will highly enable the production and distribution of such content at the best cost, facilitate its exchange, and enhance its discoverability by professional and fans alike. EIDR and ISAN have been identified by the European Commission as appropriate candidates for the systematic registration of EU's Creative Europe MEDIA program funded projects. Under the monitoring of the commission, all parties have worked to implement a fully interoperable system that allows content creators, owners and distributors to obtain identifiers from both organisations with a single application. Both ISAN-IA and EIDR operate as non-profit, cost-recovery organisations to the benefit of the industry. The registered IDs will be exchanged between the organisations, allowing ID users willing to do so to work with either organisation while accessing many of the benefits of both registration systems.</p>	Audio-visual	Comm.	Content
D4	Interested Parties Information System (IPI System) SUISA, the Cooperative Society of Music Authors	<p>The purpose of the IPI system is the global unique identification of rightsholders acting across multiple creation classes (musical work, literary work, work of art etc.), assuming different roles (musical creator, film director, author of fine art etc.), and owning all rights (performing right, reproduction right, radio</p>	Any	CMOs	Parties

	and Publishers in Switzerland, used by all BIEM/CISAC societies	<p>broadcast right etc.), determined by each creation class they deal with. This is related to the data exchange not only within the copyright societies, but also in the world-wide transaction processing with third parties, such as user organisations like radio and TV stations, sound carrier producers etc.</p> <p>The IPI system contains the real and artistic names of the rightsholders, as well as their nationalities and place of birth (natural persons) or foundation (legal entities), including their territory, time and share dimensioned agreements with the corresponding copyright societies. Around 4.5 million rightsholders are today included in the IPI system. The system is designed, developed, and operated by SUIISA to support all types of intellectual property rights protection, not limited to copyright. It directs the international economic flow, between creation users and creation rightsholders, through administrative organisations. The IPI system is in fact a backbone administration tool for collective, intellectual property protection organisations.</p>			
D5	<p>International Standard Name Identifier (ISNI)</p> <p>ISNI International Agency Limited</p>	<p>ISNI is the ISO certified global standard number for identifying the millions of contributors to creative works and those active in their distribution, including researchers, inventors, writers, artists, visual creators, performers, producers, publishers, aggregators, and more. As ISO 27729, it is part of a family of international standard identifiers that includes identifiers of works, recordings, products and right holders in all repertoires, e.g. DOI, ISAN, ISBN, ISRC, ISSN, and ISWC.</p> <p>The mission of the ISNI International Agency (ISNI-IA) is to assign to the public name(s) of a researcher, inventor, writer,</p>	Any	Any	Parties Content

		<p>artist, performer, publisher, etc. a persistent unique identifying number in order to resolve the problem of name ambiguity in search and discovery; and diffuse each assigned ISNI across all repertoires in the global supply chain so that every published work can be unambiguously attributed to its creator wherever that work is described.</p> <p>By achieving these goals, the ISNI will act as a bridge identifier across multiple domains and become a critical component in Linked Data and Semantic Web applications.</p>			
D6	International Standard Musical Work Code (ISWC) International Confederation of Societies of Authors and Composers (CISAC)	<p>The International Standard Musical Work Code (ISWC) identifies a musical work as a unique intangible creation. It relates to the result of an intangible creation of one or more people, regardless of copyright status, distributions or agreements that cover this creation.</p> <p>The International Standard Musical Work Code (ISWC) for a musical work is usually stored in a database on a computer system. It is divided into three elements, consisting of the letter T (the "prefix element"), followed by nine digits (the "work identifier"), and a numeric check digit (example: ISWC T-034.524.680-1).</p> <p>The ISWC cannot appear on the musical work (the intangible creation) itself. It should however be printed on all correspondence pertaining to the work to which it has been allocated. The ISWC should always be printed in type large enough to be easily legible. The ISWC should also be printed along with the copyright notice.</p>	Musical works	Comm.	Content

		<p>Elements of such descriptive data shall include, at a minimum, the following:</p> <ol style="list-style-type: none"> 1. At least one original title for the work, together with the appropriate title type code. 2. All creators of the work (composers, authors, arrangers, translators etc.) identified by their IP-numbers and role codes. Performing artists should also be provided where known. 3. An indication of whether or not the work is derived from an existing work and, if so, the type of derivation indicated by a derivation code. 4. One value of category must be attributed to the work when an ISWC is allocated (composite type, version type or excerpt type category). <p>The ISWC is managed and maintained by the ISWC System, with a separate database and mechanisms to allocate the ISWC centrally. ISWCs are also made available on CIS-Net by societies, but it is not the place where they are maintained.</p>			
E	Metadata standards (content & rights)	<i>Metadata schemas for both content and rights data.</i>	<i>Primary Content/Media</i>	<i>Primary Sector</i>	<i>Data scope</i>
E1	CISAC Cue Sheet standards and rules	The Cue Sheet Standards & Rules project harmonises music cue sheets. It will bring significant benefits to everyone involved in	Music	Audio-visual	Content Rights

	CISAC, International Confederation of Societies of Authors and Composers	<u>music productions for audio-visual works. Implemented in a collaboration between CISAC, representing authors societies worldwide, and music publishers and producers through the Society Publisher Forum, the Cue Sheet Standards & Rules simplify the rules governing the identification of musical works used in audio-visual productions. The harmonisation improves the administration of music rights, brings a new consistency to the use of cue sheets, and will lead to increased efficiencies and potentially reduced costs for rightsholders and users.</u>			
E2	Digital Data Exchange (DDEX) Digital Data Exchange LLC ⁵⁷⁵	<p>The purpose of <u>DDEX</u> as set out in its Operating Agreement is “to develop standards relating to metadata creation and management, identification of entities and the communication of such information in relation to media rights, agreements and content to enable a highly automated, timely and cost-effective transaction processing environment providing the highest possible level of operational efficiency amongst participants in the physical and digital media value chain...and to promote, through the creation of publicity material, the holding of training seminars and the like, global awareness and compliant implementation of those standards”.</p> <p>DDEX is playing a critical role in the development of the digital music value chain as more and more companies recognise the vital importance of standardisation. From just looking to develop standard format messages, DDEX is looking now at all aspects of</p>	Music	Comm.	Parties Content Rights

⁵⁷⁵ A standards-setting organisation focused on the creation of digital value chain standards to make the exchange of data and information across the music industry more efficient

			<p>standardisation within the digital music value chain. DDEX also carries out regular discussions with organisations that have or are developing standards of relevance to the digital music value chain, <i>as well as those developing standards for other media industries</i>. All these discussions are about identifying how the community as a whole can work to improve the operation and interoperability of these standards with the goal of automating as much of the value chain as possible. Also, in this context, DDEX has been making tentative strides towards finding ways of improving the overall quality of the metadata itself as it moves around the whole value chain. This activity is very much in its early stages but is seen as critical in the longer term.</p>			
E3	EBU Core European Union	Broadcast	<p>The EBUCore is an initiative of the European Broadcast Union. It defines a set of concepts, relationships and properties that apply to media. This is a part of metadata that can be used to describe any multimedia content. It is based on a model of metadata, Dublin Core. It was first published in 2000 as a set of definitions for the audio archives. At that time, XML was in its infancy and its use has increased dramatically since, requiring a more structured approach to describe audio-visual content information. In addition, other more semantic languages have greatly influenced the way of modelling audio-visual object. The EBUCore followed this evolution into what it is today: a Dublin Core Media⁵⁷⁶. In 2015, the version 1.6 of this standard has considered the latest developments in the Semantic Web and Linked Open Data communities. EBUCore 1.6 is available as an</p>	Broadcast	Any	Content

⁵⁷⁶ The Dublin Core, also known as the Dublin Core Metadata Element Set, is a set of fifteen *core* elements for describing resources. This fifteen-element Dublin Core has been formally standardized as ISO 15836, ANSI/NISO Z39.85, and IETF RFC 5013.

		RDF ontology entirely compatible with the W3C Media Annotation Working Group ontology, which model is common. The definitions in EBUCore 1.6 have been refined and the schema structure has been reinforced, aiming at enhancing business cooperation between European and US producers and being in phase with developments undertaken at the International Telecommunication Union level (ITU: United Nations specialised agency for Information and Communication Technologies).			
E4	IPTC photo metadata	<p>The IPTC Photo Metadata Standard is the most widely used standard to describe photos, because of its universal acceptance among news agencies, photographers, photo agencies, libraries, museums, and other related industries. It structures and defines metadata properties that allow users to add precise and reliable data about images.</p> <p>IPTC Photo Metadata Standard consists of two schemas – IPTC Core and IPTC Extension – developed by the IPTC for professional use with a focus on news and stock photos. IPTC worked alongside Adobe on the technical implementation of the metadata that employs Adobe’s XMP technology (now an ISO Standard) as an enriched alternative to the IIM format.</p> <p>IPTC Core and IPTC Extension define metadata properties with comprehensive sets of fields that allow users to add precise and reliable data about people, locations, and products shown in an image. It also supports dates, names and identifiers regarding the creation of the photo, and a flexible way to express rights information.</p>	Photos	Any	Content Rights

E5	Online Information Exchange (ONIX) EDItEUR Ltd ⁵⁷⁷	<p>The ONIX family includes standards for Books, Serials and Licensing Terms & Rights Information (including RROs). All ONIX standards are designed to support computer-to-computer communication between parties involved in creating, distributing, licensing or otherwise making available intellectual property in published form, whether physical or digital. All are expressed in XML.</p> <p>ONIX for Books was the first, and is the most widely adopted, member of EDItEUR's ONIX family of standards. It was initially developed by EDItEUR jointly with Book Industry Communication (UK) and the Book Industry Study Group (US) and is now maintained under the guidance of an International Steering Committee including not only BIC and BISG but also national user groups in Australia, Belgium, Canada, China, Egypt, Finland, France, Germany, Italy, Japan, Korea, The Netherlands, Norway, Russia, Spain, and Sweden. The ONIX for Books Product Information Message is the international standard for representing and communicating book industry product information in electronic form.</p> <p>Other ONIX standards include ONIX for Serials, and ONIX for Publications Licences aimed at communication of rights and repertoire data between RROs (Reproduction Rights Organizations), as well as more specialised formats for metadata associated with the registration of identifiers (DOIs, ISTCs, etc) and Thema, a subject category scheme.</p>	Books, Serials	Any	Content Rights
----	---	---	----------------	-----	----------------

⁵⁷⁷ An international group coordinating development of the standards infrastructure for electronic commerce in the book, e-book and serials sectors.

F	Metadata standards (rights)	<i>Metadata schemas specifically for rights data.</i>	<i>Primary Content/Media</i>	<i>Primary Sector</i>	<i>Data scope</i>
F1	Creative Commons	<p>Creative Commons (CC) is a not-for-profit organisation that helps overcome legal obstacles to the sharing of knowledge and creativity to address the world's pressing challenges.</p> <p>CC provides Creative Commons licenses and public domain tools that give every person and organization in the world a free, simple, and standardized way to grant copyright permissions for creative and academic works; ensure proper attribution; and allow others to copy, distribute, and make use of those works</p> <p>CC works closely with major institutions and governments to create, adopt and implement open licensing and ensure the correct use of CC licenses and CC-licensed content</p> <p>CC supports the CC Global Network, a community initiative working to increase the volume, breadth, and quality of openly available knowledge worldwide</p> <p>CC develop technology like CC Search that makes openly licensed material easier to discover and use</p>	Any	Any	Rights
F2	Europeana Rights Statements Europeana and partners	Twelve different rights statements can be used by cultural heritage institutions to communicate the copyright and re-use status of digital objects to the public. The rights statements have been designed with both human users and machine users (such as search engines) in mind and are made available as linked data. Each rights statement is located at a unique Uniform Resource Identifier (URI). The rights statements have been	Any	Cultural heritage	Rights

		specifically developed for the needs of cultural heritage institutions and online cultural heritage aggregation platforms and are not intended to be used by individuals to license their own creations. The rights statements fall in three categories: statements for works that are in copyright, statements for works that are not in copyright, and statements for works where the copyright status is unclear. The statements provide end users with easy-to-understand high level information about the copyright and re-use status of digital objects. Except for the statements for objects with an unclear copyright status, these statements should only be applied after the copyright status of a work has been established.			
F3	IPTC Web Statement of Rights International Press Telecommunications Council	In Autumn 2018, Google Images introduced new features to their image search results. Next to a selected photo, the image's creator, credit line, and a copyright notice are shown instantly. It works by reading the corresponding embedded IPTC photo metadata fields from the image file. This feature is now extended to also display across an image a licensable badge and a link to its licence information – using the Web Statement of Rights field – under defined conditions in the results of all image searches. Google may also show a text linked to a web page where a licence to re-use the image can be obtained – using the Licensor URL field. These fields are defined by the IPTC Photo Metadata Standard, which defines the best way to fill metadata fields such as Creator, Credit Line, Copyright Notice, Web Statement of Rights, and Licensor URL.			

F4	Open Digital Rights Language (ODRL), RightsML World Wide Web Consortium, W3C	The Open Digital Rights Language (ODRL) is a policy expression language that provides a flexible and interoperable information model, vocabulary, and encoding mechanisms for representing statements about the usage of content and services. Communities adopting ODRL can use standardised actions for permissions, prohibitions, and duties that are expressed in policy statements. RightsML is the International Press Telecommunications Council's Rights Expression Language ⁵⁷⁸ for news media and is based on ODRL.	Any	Comm.	Rights
F5	Text and Data Mining Reservation Protocol (TDM Protocol) World Wide Web Consortium, W3C	The goal of the W3C Community Group " Text and Data Mining (TDM) Reservation Protocol ", created with the support of the Federation of European Publishers (FEP) and the W3C member European Digital Reading Lab (EDRLab), is to specify a simple and practical machine-readable solution, capable of expressing the reservation of TDM rights following the rules set by the Article 4 of the Directive EU/2019/790 and, optionally, the availability of machine-readable licences for TDM actors. There are three alternative technical solutions for expressing the reservation of TDM rights, all based on well-known web standards and protocols: one based on http headers, another based on a file hosted on the origin server, and a third based on html metatags. The first two can be applied to any digital media type (text, images, video, etc.) while the latter is HTML specific. Overall, these three solutions correspond to different situations and technical skills. The priority in which these techniques	Any	Any	Rights

⁵⁷⁸ A Rights Expression Language (REL) is a machine-processable language used to express intellectual property rights (such as copyright) and other terms and conditions for use over content. RELs can be used as standalone expressions (i.e., metadata usable for search, compatibility tracking) or within a Digital Rights Management (DRM) system.

		should be processed by TDM agents will be provided in the technical specifications that will be finalised after summer 2021. A machine-readable TDM policy has been defined to detail how a rightsholder can be contacted and conditions in which a TDM licence can be acquired. TDM policies will be defined as a profile of ODRL 2.2 to allow users to detail, optionally, how a rightsholder can be contacted and conditions in which a TDM licence can be acquired.				
G	Data exchange	access/	<i>Systems and schemas providing access to content and rights data from multiple sources</i>	<i>Primary Content/Media</i>	<i>Primary Sector</i>	<i>Data scope</i>
G1	CIS-Net CISAC, International Confederation of Societies of Authors and Composers		CIS-Net is a network of databases built upon the Common Information Standard (CIS) standards. Each database constitutes a node within the overall network. There are three types of nodes: (a) local nodes, maintained by individual member societies, (b) regional nodes, developed by regional groups of member societies, and (c) the Works Information Database (WID), the CISAC database of musical works used by many societies. The network can be accessed from a web-based search engine. CIS-Net provides major benefits to all member societies regardless of their technical development or resources by facilitating more extensive and effective licensing for exploitation of the works they manage as well as faster and more efficient distribution of revenues. For these reasons, a commitment to participate in CIS-Net is a condition of CISAC membership for the music societies. It is also open to Independent Management Entities; a new approach was implemented two years ago.	Music	CMOs	Content Rights

G2	Cube ICE Services – International Copyright Enterprise Services	Cube , the new copyright platform of ICE Services, the joint venture between the Performing Rights Organisations PRS (United Kingdom), STIM (Sweden), and GEMA (Germany) will harness cloud computing and machine learning technologies to deliver a highly automated copyright system, which will increase the speed and accuracy with which ICE <i>consolidates multi-territorial copyright data</i> . The system has transparency in its DNA and will 'open the box' on how data is processed according to common data authority rules and will seamlessly integrate rightsholders into the resolution of data conflicts. Cube will also dramatically increase both the speed and capacity of its data processing capabilities and enable the fast onboarding of new customers through the implementation of highly automated and scalable data policy rules.	Music	Comm.	Content Rights
G3	Handle system Corporation for National Research Initiatives (CNRI)	<p>The Handle system is the Corporation for National Research Initiatives's proprietary registry assigning persistent digital identifiers or handles to information resources, and for resolving "those handles into the information necessary to locate, access, and otherwise make use of the resources".</p> <p>The Handle system provides the underlying technology for the DOI and EIDR identification systems (although it is independent of these) as well as others, enabling (1) digital objects to retain their identifiers when their location URL changes and (2) a single handle to direct users to multiple objects (for example, content or rights metadata).</p> <p>The Handle System is a rapid-resolution, globally distributed system run by multiple groups that the public can use for resolving identifiers (handles). The Handle System is an</p>	Any	Any	Any

		implementation of the Identifier and Resolution component of the Digital Object Architecture . The Global handle Registry (GHR) is a distributed registry whose operation is managed collaboratively by the DONA Foundation and multiple organizations that are credentialed and authorized by DONA.			
G4	Linked Open Data Europeana	Linked Open Data is a way of publishing structured data that allows metadata to be connected and enriched, so that different representations of the same content can be found, and links made between related resources. All Europeana datasets can be explored and queried through the SPARQL API. The metadata for all the objects in the Europeana portal is open, in that it is all licensed under the CC0 Public Domain Dedication under the terms of the Data Exchange Agreement (DEA), and can be freely downloaded via the API.	Any	Cultural heritage	Any
G5	Repertoire Data Exchange ⁵⁷⁹ (RDx)	RDx stands for Repertoire Data Exchange. It is the industry data portal for the supply and exchange of performance rights repertoire data between record companies and Music Licensing Companies (MLCs), i.e., CMOs. RDx is a data exchange hub, which puts the DDEX "Recording Data and Rights" (RDR) standard into practice as part of its core functionality. It allows multiple record companies to deliver data about their recordings and rights to multiple MLCs. RDx also returns to the record companies details of the registration status of their recordings and any claims conflicts that arise. RDx is a mechanism for improving the delivery of data but does not affect other aspects	Music	Comm.	Content Rights

⁵⁷⁹ A joint venture of WIN (Worldwide Independent Network), and IFPI (International Federation of the Phonographic Industry), powered by PPL (Phonographic Performance Ltd in the UK).

		of the way performance rights work is managed. Record companies still need to maintain membership and mandates with MLCs. RDx is a joint venture between IFPI and WIN. It was launched in 2020 with 4 leading MLCs and 4 independent and major record companies using the service, and with a growing number of MLCs and rightsholders preparing to join RDx through 2021.			
G6	URights SACEM, Société des auteurs, compositeurs et éditeurs de musique in France	URights is a platform developed by SACEM and IBM. It is open by design to allow other partners to integrate, such as other Collective Management Organisations (CMOs) across the world, allowing them to avoid cost duplications. It will provide customised services tailored to the specific nature of their local markets. URights will allow CMOs to address royalties for creators and publishers in areas such as music and audio-visual content. The open architecture means that other partners can use URights' processing technology while maintaining their own rights databases, with the highest standards of security and data confidentiality. URights will also offer creators and publishers new services to help them access and analyse market analytics, as well as to provide a better understanding of cultural goods consumption.	Music	Comm.	Content Rights
G7	European Strategy Data European Commission	As part of the European Strategy for Data , "the Commission intends to fund the establishment of EU-wide common, interoperable data spaces in strategic sectors. Such spaces aim at overcoming legal and technical barriers to data sharing across organisations, by combining the necessary tools and infrastructures and addressing issues of trust, for example by way of common rules developed for the space. The spaces will include: (i) the deployment of data-sharing tools and platforms;	Any	Any	Any

		(ii) the creation of data governance frameworks; (iii) improving the availability, quality, and interoperability of data – both in domain-specific settings and across sectors. Funding will also support authorities in the Member States in making high value data sets available for re-use in the different common data spaces. The support for data spaces will also cover data processing and computing capacities that comply with essential requirements in terms of environmental performance, security, data protection, interoperability and scalability”			
G8	European Blockchain Services Infrastructure (EBSI) European Commission/ European Blockchain Partnership	<p>The EBSI is an initiative which “supports the creation of cross-border services e.g., for citizens to manage their own identity, [] credentials and register documents”.</p> <p>The EBSI has focused on four use cases relevant to rights management information: notarisation (of digital works or assets), credentials, self-sovereign identity (of authors, rightsholders or other stakeholders) and trusted data sharing. The European Blockchain Partnership has received the proposals of consortia asked to build a performing EBSI. The consortia were asked to test the developed distributed ledger technologies and evaluate the level of performance and the improvements achieved by the new capacities at hand of one or two use cases. The “EU wide management of IP rights (like patents, trademarks, copyrights), including also the management of copyrights that can be directly associated to digital content in near real-time” was one of them.</p>	Any	Any	Any

H	Datasets	<i>Non-proprietary databases and datasets of global value</i>	<i>Primary Content/ Media</i>	<i>Primary Sector</i>	<i>Data scope</i>
H1	Lumière VoD European Audio-visual Observatory	Lumière VoD is a directory of European works (film and TV content) available on on-demand services in Europe. It helps find the services and countries where a film or a TV content is released on pay-video on-demand services, transactional and subscription Video on Demand (VoD), and combines search criteria to create lists of available films by director, country or year of production and available TV content by country of production. Lumière VoD is primarily designed for audio-visual industry professionals: authors, producers, distributors, film funds and regulators in order to help them track the exploitation of works on VoD and to assess the composition of the VoD catalogues. It is not intended to facilitate the rental or purchase of works, nor the subscription to a service.	AV	Comm.	Content Rights
H2	Orphan Works Database EU Intellectual Property Office (EUIPO)	The Orphan Works Database provides information related to orphan works contained in the collections of publicly accessible libraries, educational establishments, and museums, as well as archives, film or audio heritage institutions and public-service broadcasting organisations established in the Member States. Orphan works are works that are still protected by copyright, but whose authors or other rightsholders are not known or cannot be located. Music, books, newspaper and magazine articles and films can be orphan. The Directive EU/2012/28 on certain permitted uses of orphan works sets out common rules to make digitisation and online display of orphan works legally possible. Under Article 3(6) of this Directive, EUIPO is responsible for the	Any		Content Rights Authority

		establishment and management of a single publicly accessible online database on orphan works.			
J	Rights Platforms	<i>Shared solutions for managing rights</i>	<i>Primary Content/ Media</i>	<i>Primary Sector</i>	<i>Data scope</i>
J1	Access to Rights Data via Identification Technology Optimisation (ARDITO) A European Horizon 2020 research and innovation project	The project aimed at filling the gap in the digital content value network and connecting online contents to rights information, by building a complementary digital rights data network. It developed tools and market-driven services to support creators and small and medium-sized businesses in the creative content sector (images, publishing, eBooks, videos, and multimedia) to find new business ideas through monetising the re-use of their content. The project implemented existing components of a rights data network by optimising a range of content identification technologies (including DOI, watermarking and fingerprinting) to provide seamless access from the identifiers to the rights and licensing information and services and integrating them into the Copyright Hub ecosystem.	Text, Images	Comm.	Content Rights
J2	The Copyright Hub	The Copyright Hub is a UK-based non-profit organisation attempting to lower the transaction costs of licensing copyrighted items. It was proposed in the Hargreaves Review which recommended the creation of an "industry-led" body, dubbed the <i>Digital Copyright Exchange</i> , to make licensing more convenient. The Copyright Hub took on responsibility for developing the open-source technology which would deliver the Hub's vision. The idea was to make it easy and free for anyone to attach an identifier to any piece of content, and then to create	Any	Any	Parties Content Rights Awareness and Understanding

		<p>a record with the Copyright Hub which would link that identifier to an authoritative record of information about the piece of content and its owner. The technology would also allow automated licensing processes to take place, making authorisation and creation of records of permission a very low-cost process. The Hub planned to create a browser extension and an online copyright ownership database, the plugin talking to the database to allow Internet users to conveniently license copyrighted works such as images.</p>			Authority
J3	<p>RAIDAR Berklee College of Music</p>	<p>Berklee College of Music has long advocated for the creation of a more open and transparent music business, one that gives artists more control over their music, their data, and ultimately their careers. When the college formed the Open Music Initiative (OMI) in 2016 with Netflix, YouTube, and dozens of other industry players, it set out to do just that. OMI's priority was to fix the transfer of music metadata – the full information about an artist's work, including the musicians, songwriters, producers, labels, and publishers involved in its creation – to ensure that when a song is played, the right people get paid and credited. The industry's lack of standardised protocols for copyright attribution and royalty allocation have led to a tangled mess of licensing deals, rules, and intermediaries, resulting in error-filled royalty statements and millions in lost or misallocated royalties, leaving artists unpaid for their work. The solution was to develop an industry-wide framework linking artists and rights-holders to their works through a network of ledgers – an open-source model that has proven successful in the book publishing, auto parts, and library systems industries.</p>	Music	Any	<p>Parties Content Rights</p> <p>Authority</p>

J4	WIPO Connect World Intellectual Property Office (WIPO)	<p>WIPO Connect is an IT solution, developed and offered by WIPO, aiming at facilitating the collective management of copyright and related rights. It works on two levels:</p> <p><i>WIPO Connect Local</i> is a web application used for day-to-day operations, either installed on a local server or hosted in the cloud (Registration of rightsholders; Management of documentation: works, performances, and sound recordings; Management of licensing agreements; Identification and matching of works which have been used, providing usage reports or data capture; Distribution reports providing the amount of royalties to be distributed to rightsholders based on usage, documentation, and local parameters).</p> <p><i>WIPO Connect Shared</i> is a fully cloud-based solution, synchronizing WIPO Connect Local implementations and exchanging data with industry data sources (Synchronization in the cloud of WIPO Connect Local implementations; Submission of documentation to industry databases and retrieving information for synchronization; Facilitating and automatizing the assignment of industry identifiers; Dissemination of local repertoires to foreign CMOs).</p>	Any	CMOs	Parties Content Rights
K	Authentication	<i>Shared solutions for authenticating content, data or parties</i>	<i>Primary Content/ Media</i>	<i>Primary Sector</i>	<i>Data scope</i>
K1	Content Authenticity Initiative (CAI) Adobe and others	The Content Authenticity Initiative is building systems to provide provenance for digital media, giving creators tools to express objective reality and empowering consumers to evaluate whether what they are seeing is trustworthy. The initiative is	Any	Any	Parties Content

		designing components and drafting standards specifications for a simple, extensible, and distributed media provenance solution. It focuses its efforts on (a) the detection of deliberately deceptive media, (b) education, and (c) content attribution.			Awareness and Under- standing Authority
K2	WIPO Proof World Intellectual Property Office (WIPO) <i>Update: WIPO PROOF will be discontinued in February 2022.</i>	WIPO PROOF is a new digital business service that provides a date- and time-stamped digital fingerprint of any file, proving its existence at a specific point in time. This new service complements WIPO's existing intellectual property systems. It is specifically designed for the increasingly digital world where innovation and creativity are enabled by technology, big data, and global collaboration.	Any	Any	Content Authority
L	Policing	<i>Systems for tracking or acting on breaches</i>	<i>Primary Content/Media</i>	<i>Primary Sector</i>	<i>Data scope</i>
L1	Copyright Clearing House on the Internet (CUII) Selbstregulierung Informationswirtschaft e.V., a German association of rightsholders and Internet service providers	The Copyright Clearing House on the Internet (CUII) is an independent body. It was founded by internet access providers and rightsholders to use objective criteria to check if blocking the access to a structurally copyright-infringing website is lawful. A review committee checks at the request of the rightsholder and, if the conditions are met, recommends a DNS block of this structurally copyright-infringing website. The recommendation of the examination committee is made unanimously and only in the case of clear copyright infringements. The recommendation is sent to the Federal Network Agency for Electricity, Gas, Telecommunications, Post and Railways (BNetzA). If the	Any	Any	Authority

		<p>examination by the Federal Network Agency shows that a DNS block is harmless under the provisions of the Net Neutrality Ordinance (Regulation (EU/2015/2120), the CUII informs the Internet access providers and the applicants. Otherwise, the Internet access providers involved in the CUII block the corresponding domains of the structurally copyright-infringing website.</p>			
--	--	---	--	--	--

5.4. Annex 4: Impact Model

To structure and prepare the analysis for this study, the study team used an **impact modelling approach**. An impact model can be interpreted as a simplified, structured illustration of the causal chains linking rights metadata challenges discussed in this study with potential (socio)-economic effects caused by these challenges.

It is important to keep in mind that the model per se is not making statements per se on whether (and to what extent) the postulated impact paths or causal chains are empirically observable. Some of the relations shown in the model might in fact not be applicable to some industries. It is the objective of the empirical work based on the model to **validate or reject the hypotheses made in the model**. This exercise is carried out in the following sector-specific sections.

The model for the music industry, for example, is depicted in Figure 70 hereunder. The chart can be read starting from the upper left corner. It departs from challenges ("problems") related to rights metadata which are debated in different creative industries to a larger or smaller degree (such as its absence or a lack of interoperability of systems to process rights metadata).

In a next step, **direct or primary impacts** are shown. They represent effects which might be – from a theoretical point of view as well as based on the experience of the study team – caused by the discussed "problems". The primary impacts are still on a rather abstract level and relate to aspects such as accuracy of rights metadata, simplicity to process/manage metadata, as well as the time and cost factor related to the metadata management (speed and affordability).

Ultimately however, this study attempts to approximate the effects of rights metadata issues on real, "hands-on" economic factors in creative industries in Europe. These factors are shown on the right side of the impact model and relate to the **pain points** described in section 2.5, including costs for rights management, inefficient market processes (especially related to rights licensing) and ultimately affecting revenues of different industry participants (see upper right corner of the impact model).

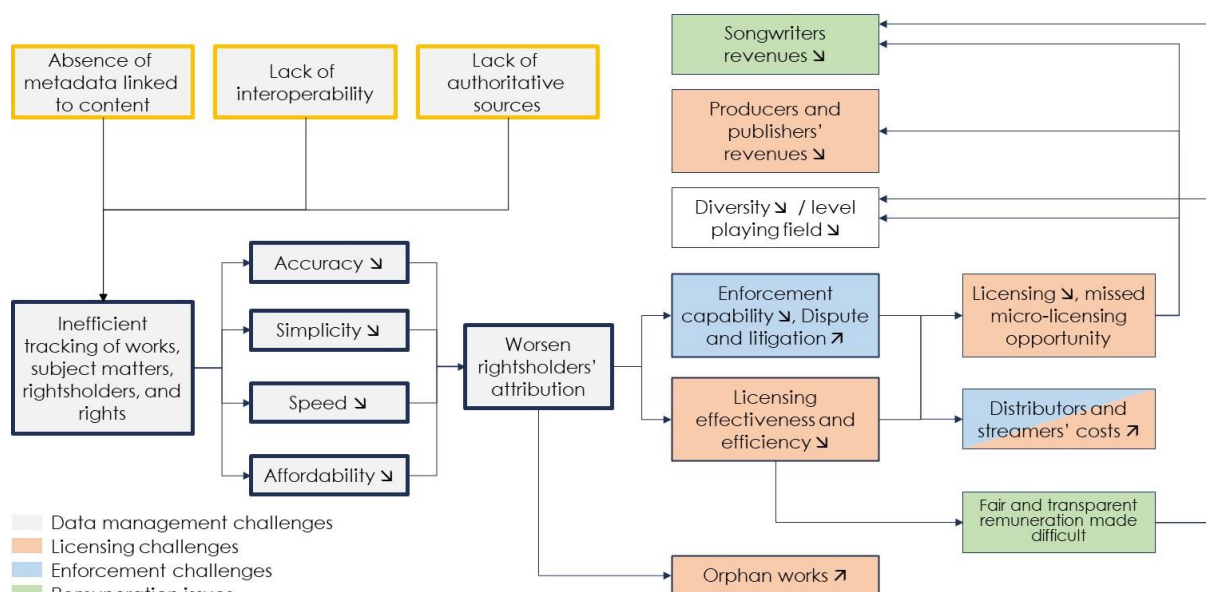


Figure 70: Exemplary impact model for the music industry - potential metadata challenges and impacts⁵⁸⁰

Source: FH Krems/study team

The objective of using impact models such as the one described is to **disassemble the impact pathways** in the way presented above. This gives a more granular view on how impacts might occur and are linked to each other as well as into potential impacts per se. In this sense the impact models have provided the study team with a conceptual basis for the empirical work⁵⁸¹ as well as the presentation of the empirical findings.

5.5. Annex 5: Some characteristics of rights data management

5.5.1. Prevalence of the content value network

Copyright data management must be effective and efficient in a precise context, i.e., on the **content value network**. Already in 2003, Andrew Leyshon⁵⁸² argued that **a set of networks** were emerging on the Internet and would undermine the ability of large media companies to control copyright in the way they have in the past. Following on, Hyojung Sun⁵⁸³ developed a "framework of **Digital Music Value Networks**, which provides a useful tool to capture the complexity of the interactions amongst diverse forces and their interactions in the four major networks of creativity, reproduction, promotion/distribution, and consumption"⁵⁸⁴.

⁵⁸⁰ For online music services, the three problems at the start of this model are mainly found with songs not recordings. With UGC (User Generated Content) and broadcast/public performance, it could affect recordings too.

⁵⁸¹ See also section 2.1.3 on further methodological remarks.

⁵⁸² A. Leyshon, *Scary Monsters? Software Formats, Peer-to-Peer Networks, and the Spectre of the Gift*, Environment and Planning D: Society and Space, 21(5), pp.533-558, October 2003, <https://doi.org/10.1068/d48j>

⁵⁸³ H. Sun, *The Digital Revolution: Tamed: The Case of the Recording Industry*, Palgrave MacMillan, 2019.

⁵⁸⁴ See *Music 2025: the music data dilemma. Issues facing the music industry in improving data management*, page 24 at <https://www.gov.uk/government/publications/music-2025-the-music-data-dilemma>, already mentioned in sections 2.3.1.2 and 2.6.5.8

In parallel, research was launched to enable value to be exchanged as quickly as information, with each flow of digital work or related subject matter being immediately compensated with a counterflow of remuneration. It aims at leveraging the **Internet of Value** to realise the potential of the creative sectors⁵⁸⁵.

On the content value network, **any actor can interact bidirectionally with any other stakeholder** (e.g., CMOs to streaming platforms and vice-versa), i.e., **exchange of value and data** can occur bidirectionally between any node of the network.

The characteristics of the content value network are as follows.

- **Complex.** This relates to consideration of moral and commercial rights, flows and counterflows of rights data; natural and legal persons and different ways to identify them or protect their privacy; collaboration between diverse entities, e.g. not-for-profit and profit-making, and data treatments of the various exceptions and limitations that apply to them or not; groups and their members, for example a school and its students; the fact that a stakeholder can have several copyright-relevant roles at the same time, and the fact that works and rightholders are regulated by different national legislations at the same time.
- **Dynamic.** As a stakeholder's role or status can change over time (e.g., an online platform becoming a record producer or vice-versa, or a journalist being a reader and a reader being a commentator); as the usage of a work can change over time, for example, from not-for-profit to profit-making; or as the rightholders in a work can change over time because rights can be bought or inherited.

And as Martin Schaefer wrote in his article "Why metadata matters for the future of copyright"⁵⁸⁶:

- **Different by type of content**, in practice, clearing and administering rights in, e.g., photographs differ thoroughly from doing so for films, scientific articles, trade publishing books or sound recordings,
- **Multi-layered**, there is almost no piece of content where the rights situation concerns a single category or right holders only. This is true even for a photograph, where there is only one person clicking the button. Apart from the question whether the photo is qualified for copyright protection, there might be copyrights in the objects depicted, or rights of publicity concerning the persons or objects depicted. For a single pop song, you will need data about the composer, the corresponding music publisher and often a CMO (collective management organisation). The same applies for the lyricist. In addition, there is the performer and the label, and probably yet other CMOs representing them for certain types of use but never for all. Some are always administered individually, and it is likely that the different right holders will not exercise their rights uniformly,
- **Fragmented**, to make things yet more complicated, even for a short pop song there will frequently be not one but numerous co-composers and co-lyricists, each of whom might be represented by different music publishers with a different percentage - and this (e.g., in the case of a CD or an album offered for streaming) differs from track to track,

⁵⁸⁵ See P. Rixhon, *New Media Business Models to emerge from the Internet of Value*, in *Enabling the Internet of Value*, Springer, 2021

⁵⁸⁶ Retrieved at <http://copyrightblog.kluweriplaw.com/2020/11/27/why-metadata-matter-for-the-future-of-copyright/> in December 2020

- **International**, often rights ownership of some (if not all) who are holding rights in a production will differ from country to country,
- **Volatile**, as if that was not enough, metadata is not static over time, as rights in content are a tradeable asset. Accordingly, ownership in content changes frequently (e.g., if publisher A sells part of its catalogue to publisher B), just as changing user demand is likely to require new sets of data. Also, while the new owner of a right is usually keen to inform CMOs and other operators of central databases as quickly as possible after rights acquisition (to get paid his share), the former owner does not necessarily feel the same pressure to act - with the result that two different owners might show up for the same content,
- Last but not least, licenses are often granted for a **limited period of time** only so that a rights holder in January 2021 may no longer be entitled to exercise the right in February 2021. Also, rights are being revoked or else terminated from time to time - leaving sub-licenses in force however under certain circumstances.

Managing rights metadata on the content value network differs from managing rights metadata on a limited value chain: **access to data repositories and exchanges will have to be designed and developed accordingly**⁵⁸⁷, a fortiori when the goal is to **synchronise** the consumption or experience of content with the remuneration of authors and rightholders.

5.5.2. Purposes of rights management information

Rights management information accompany digital assets all along the digital supply/value chain⁵⁸⁸. Digital asset management is a series of operations on digital assets (works or other subject matters) that require the use of computer applications to ensure that the owner, and possibly their delegates, can perform operations on the data files without any risk of losing the data.

⁵⁸⁷ I.e., to facilitate multi-purpose, multi-channel, and multi-stakeholder access in real-time.

⁵⁸⁸ "The management of rights requires industry stakeholders to perform a series of data operations on digital content assets (e.g., images, videos, sound files or other types of works or subject matter), the purpose thereof being to distribute, import for distribution, broadcast, communicate to the public or make available to the public works or other protected subject-matter", Recital 56 of Directive EC/2001/29.

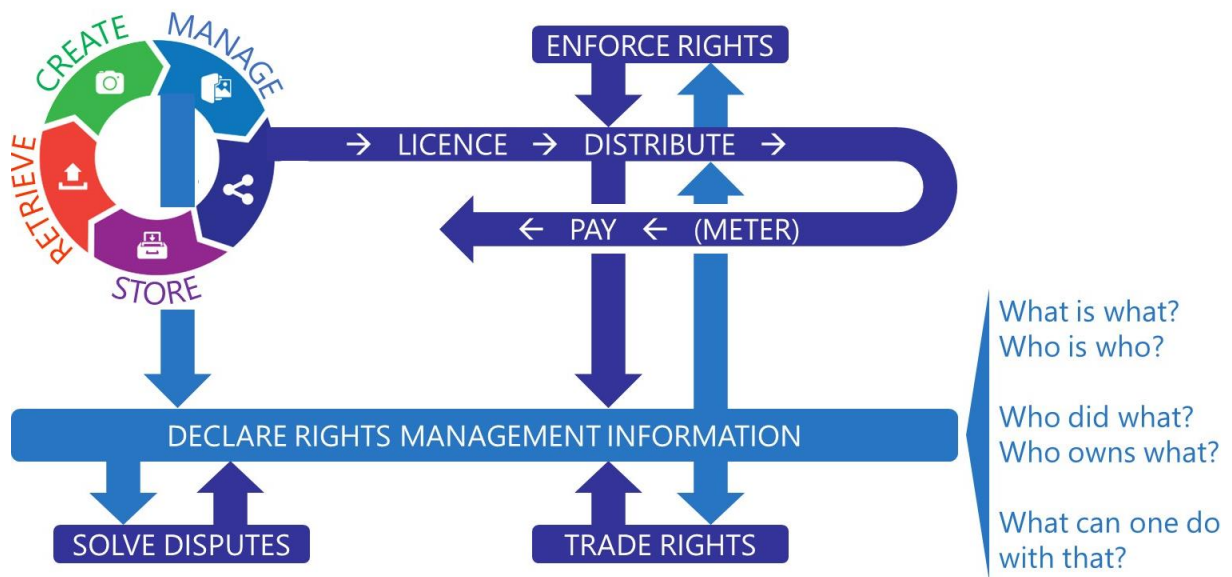


Figure 71: Purposes of RMI

Source: Philippe Rixhon Associates

Rights management information is the data one needs to support the distribution flow – license, distribute, (meter), and pay. Following our modelling of the purposes of rights management information, related among others to Recital 56 of the directive EC/2001/29⁵⁸⁹, the creative industries perform a series of digital operations summarised as follows:

- **Create.** This refers to applications importing digital assets (works or related subject matter) from the analogue and/or digital world (by encoding, scanning, optical character recognition, etc.) or authoring it as a new object. **Identification metadata** for work, related subject matter, authors and rightholders should already be recorded at this point.
- **Manage.** Operations to make digital assets easily available to their users by providing a searchable index that supports retrieval of assets by their content and/or metadata, including RMI. The cataloguing function is usually part of the ingestion process for new assets. This is primarily the domain of **descriptive metadata**.
- **Distribute.** This relates to a series of operations across multiple actors including licensing, distribution, metering of consumption (depending on sectors and business models), and royalty payment. Each of these operations require **rights metadata**. Metering produces **usage metadata**.
- **Store.** Digital assets, representing works and related subject matter in a digital format, have a lifecycle. Such a lifecycle is represented by **identification metadata** which documents various states such as creation, approval, live, archived and deleted. Often, earlier versions of a digital asset will be stored to allow

⁵⁸⁹ "There is, however, the danger that illegal activities might be carried out in order to remove or alter the electronic copyright-management information attached to it, or otherwise to distribute, import for distribution, broadcast, communicate to the public or make available to the public works or other protected subject-matter from which such information has been removed without authority. In order to avoid fragmented legal approaches that could potentially hinder the functioning of the internal market, there is a need to provide for harmonised legal protection against any of these activities".

them to be downloaded or reinstated. Therefore, a storage system can operate as a type of version control system.

- **Retrieve.** This refers to security control operations ensuring that relevant people have access to assets. This involves **administrative metadata** and often integration with existing directory services through technology such as single sign-on.

Complementary digital operations include:

- **Solve disputes** – Ownership and availability of rights can be disputed, and **identification, rights, and administrative metadata** will be used to help resolve the argument.
- **Enforce rights** – Digital assets are distributed and protected, typically against illegitimate exploitation; **identification metadata** will be used to enforce this protection.
- **Trade rights** – Finally, rights may be inherited or bought; the transfer of rights trade is informed by **identification and rights metadata**.

5.5.3. Works and metadata flows

The **synchronisation of content consumption and rightholder remuneration** is an objective of the management of data related to copyright-protected content in the digital era. On one hand, content consumption necessitates a **flow of data** (content) and metadata (information about content) from the content creation to the content consumption. On the other hand, rightholder remuneration requires a **counterflow of data** (remuneration) and metadata (information about remuneration linked to information about content) from the consumer to the creator.

The synchronisation of flows and counterflows requires:

- reliable, exhaustive, and up-to-date **necessary** RMI, whereby what is **necessary** for one use case may be **unnecessary** for another use case;
- a counterflow of remuneration data that is as **unconstrained** as the flow of content data it relates to, *unconstrained* means fast, simple and efficient; these are relative notions, they depend on use cases and rely on procedures that can be anything from fully manual to fully automated.

It is clear, however (see the characteristics of the content value network hereabove) that these procedures can be complex.

An example of flow and counterflow from the music industry

Figure 14 below illustrates the **flows** of works and other subject matter such as recordings, and **counterflow** of remuneration in the case of music streaming.

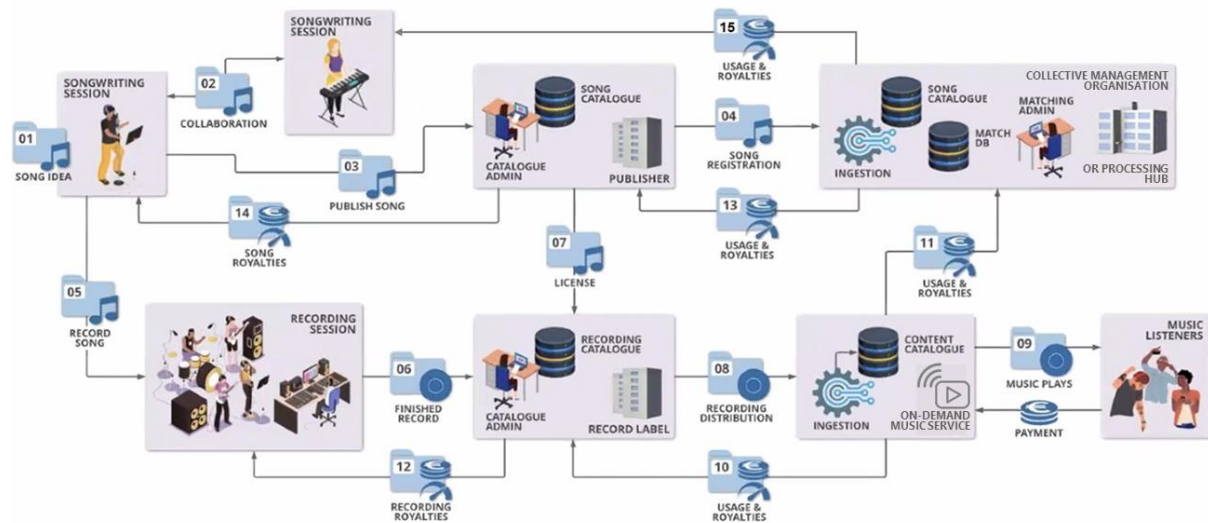


Figure 72: Flow of data pertaining to work and related subject matter, and counterflow of data pertaining to remuneration – example 1: music industry

Source: Stage Enterprises

These flows vary even within one sector, here the music industry. For example, the flows are not the same for music broadcast on the radio or in a shopping centre, and not the same for music embedded in a subsequent work such as a film or user-generated content.

A distinction: preventive vs. corrective rights data management

In the example of Figure 13, an issue of rights data management is the general absence of reliable identification of songwriters and rights splits⁵⁹⁰ at step 08. Consequently, an online music service such as Spotify is not equipped with data to swiftly remunerate the composers. They return that task to record labels, Collective Management Organisations, or Independent Management Entities, which must then tackle the challenge of matching identifications of recording (see ISRC, [International Standard Recording Code](#)) with identification of work (see ISWC, [International Standard Musical Work Code](#)), and deduplicate RMI concerning the same digital asset. This is a **corrective** measure.

One cause of this situation can be the disconnect between the wish to quickly push a new song on the market and the time it takes to issue an ISWC. Both CISAC, the issuer of ISWCs, and technology providers such as Session⁵⁹¹, remedy the situation respectively by reengineering their procedures or deploying applications capturing songwriters' identities as close as possible to the moment of creation. These are preventive measures.

An example of flow and counterflow from the publishing industry

Figure 73 below illustrates the **flow** of works and other subject matter such as performances, and **counterflow** of remuneration in the case of a theatre play⁵⁹². Notwithstanding additional complexities arising when the play is an adaptation or a

⁵⁹⁰ Most songs nowadays are written by several songwriters or composers who split their remuneration rights.

⁵⁹¹ See <https://session.id/>

⁵⁹² A playwright can be represented by a literary agent or not, accordingly the playwright may have a direct or indirect relationship with a Collective Management Organisation.

translation of another original work, it is apparent that flow and counterflow are simpler than in the above music streaming example.

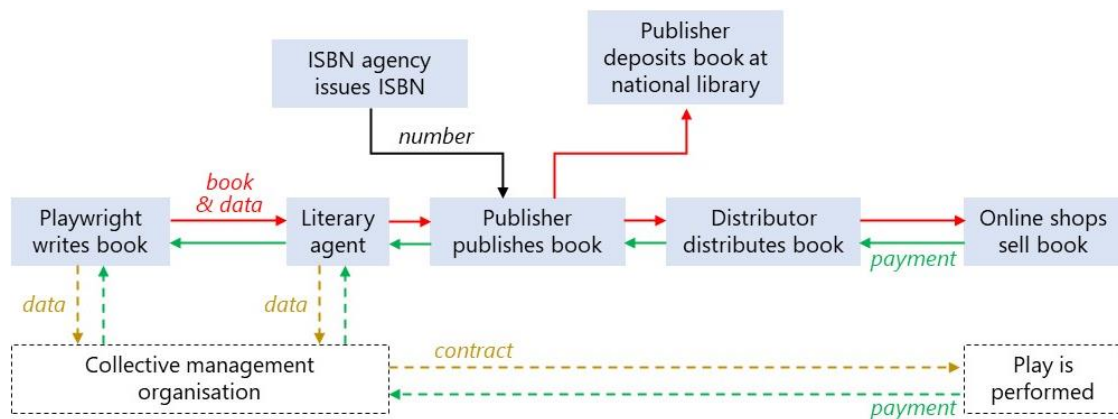


Figure 73: Flow of data pertaining to work and related subject matter, and counterflow of data pertaining to remuneration – example 2: theatre playwriting

Source: Philippe Rixhon Associates

Content identification metadata are issued by a (national) ISBN agency. This example showcases a book that may require in fact several ISBNs, one for the printed version and one per format for the eBook (.epub, .mobi⁵⁹³, .pdf, etc.), since the ISBN is a product identifier and not a work identifier.

Stakeholders on the distribution chain, from playwrights to online shops, may use their own numbers to identify the other stakeholders in the chain.

Descriptive metadata originates from the playwright, first created and registered in the ISBN system by the publisher and then usually quality-checked and enriched by a bibliographic agency (which may or may not coincide with the ISBN agency). Trade metadata (e.g., price and availability) are produced by the publisher and then updated by distributors. Rights metadata are managed in less standardised way involving literary agents, publishers, and – for some secondary rights – CMOs.

A specificity: aggregation of rights

Flows of works, other subject matters, data, and remuneration vary greatly among creative industries. Data flows differ greatly between the music industry, the publishing sector (books, images, newspapers and magazines, and journals) and the film and TV producers and distributors. Generally, the publishing sector and the film industry tend to use rights buyouts and aggregation to facilitate the commercialisation of works and other subject matters.

The principles of fair, appropriate, proportional, and transparent remuneration of the contributors to works or their publications are anchored in the *acquis communautaire*. An accurate application of these principles would require the metadata practitioners to find ways to **disaggregate attribution and usage data**, i.e., individual contributors and contributions would have to be recorded and bound as **rights metadata**.

⁵⁹³ It is now not common that an eBook is published in multiple formats. .mobi, used only in the Amazon shop, is now a transformation of the .epub provided by the publisher and de facto is not assigned with another ISBN.

5.5.4. Drowning in data

What applies to information in general⁵⁹⁴ applies to rights management information. The volume of rights metadata grows exponentially with the volume of content. According to the recent Synchtank report on *Drowning in data*⁵⁹⁵:

- ever-rising data volumes, increasingly fragmented rights, and an historic number of income sources are creating unprecedented challenges for music publishing royalty and finance teams.
- increasing revenue is dependent on processing exponential growth in data, which could be in the hundreds of trillions within five years for CMOs and hundreds of billions for publishers.

New data management approaches may be necessary to control this situation and assure the trustworthiness of information pertaining to rights ownership and rights availability. Metadata practitioners trying to solve copyright data problems will be inspired by Charu Aggarwal's research and developments⁵⁹⁶.

5.5.5. Outline of a generic functional architecture

The study team defined a minimum set of use cases as a basis to outline a generic⁵⁹⁷ functional architecture. Such an architecture must indeed address **needs** of the creative industries, in specific in the areas of:

- **Rights management.** Protect moral and economic rights, essentially the exclusive rights of reproduction and making available defined in the Infosoc directive, and also the rights related to user generated content (Directive (EU) 2019/790 Article 17 of, etc.).
- **Rights licensing.** Operate electronic markets for media assets, in particular to support the licensing for use of illustration in teaching (EU/2019/790 Art. 5 §2) or the micro-licensing of digital assets (Article 15 of Directive (EU) 2019/790, etc.)
- **Rights remuneration.** Assure contributors and rightholders' fair, appropriate, proportionate and transparent remuneration (Articles 16, 18 and 19 of Directive (EU) 2019/790 etc.) and the remuneration of out-of-commerce digitisation (Article 8-11, of (EU) 2019/790)
- **Rights enforcement.** Prevent misappropriation of digital assets (Article 17 §4 (EU) 2019/790), and allow rightholders to reserve their rights in cases of text and data mining (Article 4 of (EU) 2019/790.)

⁵⁹⁴ Claude Shannon defined the entropic uncertainty in information as part of his theory of communication. Entropy in information theory is directly analogous to the entropy in statistical thermodynamics. Accordingly, a data communication system is composed of three elements: a source of data, a communication channel, and a receiver. In Shannon's theory, the fundamental problem of communication is for the receiver to be able to identify what data was generated by the source, based on the signal it receives through the channel. Shannon considered various ways to encode, compress, and transmit messages from a data source, and proved in his famous source coding theorem that the entropy represents an absolute mathematical limit on how well data from the source can be losslessly compressed onto a perfectly noiseless channel. See Claude Elwood Shannon and Warren Weaver, *A Mathematical Theory of Communication*. University of Illinois Press, 1949

⁵⁹⁵ Griffiths, E., *Drowning in Data: Royalty accounting and systems in the digital age*, Synchtank, London, 2021 (<https://www.synchtank.com/blog/musicroyaltyaccountingreport/>).

⁵⁹⁶ Charu Aggarwal, C. (ed.), *Managing and Mining Uncertain Data*, Springer US, New York, 2009.

⁵⁹⁷ Non-sector-specific, but specifiable for each sector.

After analysis, and based on the interviews, the use cases allowed us to distinguish among basic **clusters of functionalities** (see also figure 15):

- **Voluntary registries⁵⁹⁸ of RMI** allowing users to declare rights individually or in bulk, attributing these rights, attesting these attributions, flagging conflicts and providing certifications, with the following core capabilities:
 - **Attribute** works or related subject matters to rightholders by linking granular content-based identification of works or related subject matters with RMI; for example, Maria wrote that song, and John took that picture.
 - **Verify** trust about these attributions, e.g., 'Maria wrote that song' is a reliable information, you can trust me, 'John took that picture' is a reliable information, you can trust me,

This requires immutable identifier/identity bindings pointing to dynamic metadata.

- **Intelligent metadata ingestion** allowing users to declare rights, individually or in bulk, even if the metadata are inaccurate; this will include professional or amateur⁵⁹⁹ declarations of rights pertaining to formats or ready-made creations, catalogues, heritage, or orphan works and related subject matters, with the following core capabilities.
 - **Ingest metadata at creation** which requires intelligent user experience and interfaces (UX/UI) and non-proprietary digital fingerprinting, as creators must be able to protect and exploit their rights without any knowledge of intellectual property or information technologies.
 - **Rely on content-based identification** of works or related subject matters and rightholders' digital ID wallet or self-sovereign identity.
 - **Ingest catalogue and heritage metadata** that may require pattern recognition and rule-based algorithms - namely machine learning aka 'artificial intelligence', to help humans curate metadata that are often inaccurate - missing or erroneous.
- **Electronic markets for media assets** covering business-to-business and business-to-consumer transactions, with the following core capabilities.
 - **Manage and store digital assets**, works and related subject matters.
 - **License**, eventually through rule-based smart licensing, and its corollaries of standardisation, simplification, and automation.
 - **Distribute** including identity and access management.
 - **Track and trace** digital assets.
 - **Analyse and meter** event streams.
 - **Pay and channel** royalties.
 - **Report** analytics.
- Dispute resolution.
- **Rights trading**, in the sense of transferring (selling and buying) rights as opposed to exploiting rights

⁵⁹⁸ This includes services like the ARDI (see section 2.6.6.2), where the "register" is only about the identifier of a right declaration, the natural or legal person who did that declaration, and a link to a web resource containing the declaration.

⁵⁹⁹ For example, in the case of user-generated-content (UGC).

They covered most of the issues related to copyright data management and facilitated the further conduct of the study.

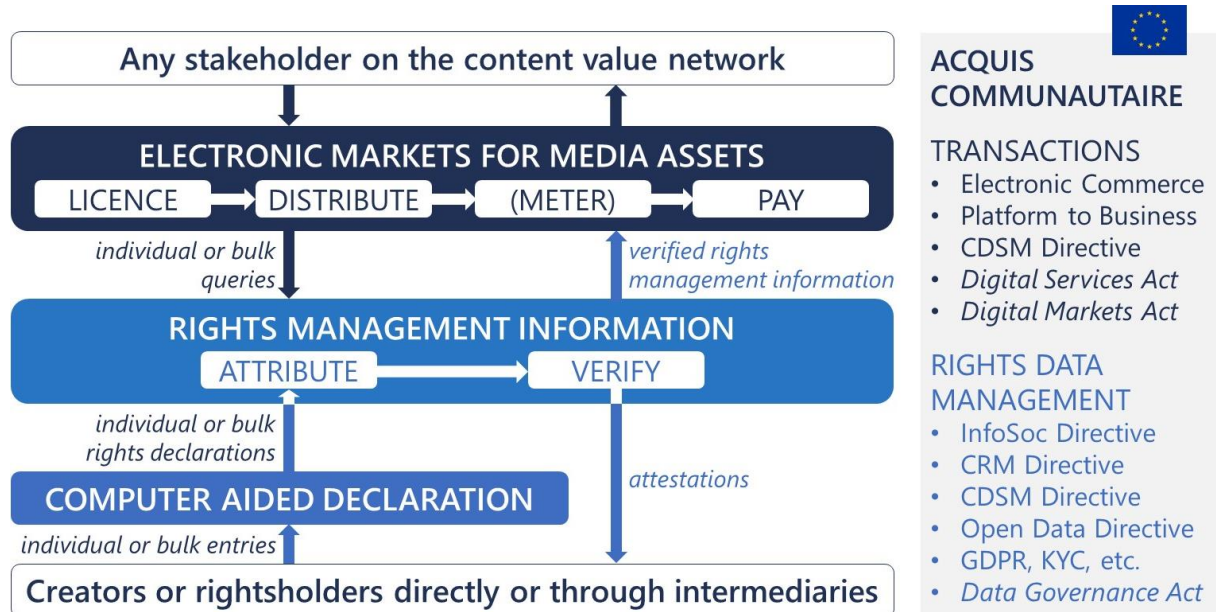


Figure 74: Functionality and regulation

Source: Philippe Rixhon Associates

These clusters of functionalities, or modules, must be **interoperable**. Keeping them **distinct** will allow to accommodate:

- **Distinct regulations.** For example licensing and distribution are governed by directives and legislations related to copyright, e-commerce and platform to business and soon the digital services act and the digital markets act, while the data systems and registries of RMI must comply with, among others with the rules of multi-territorial licensing of rights in musical works and the upcoming data governance act, and the same systems and registries must comply with the General Data Protection Regulation (GDPR), or know your customer/anti-money-laundering (KYC/AML).
- **Distinct business models.** For example, intelligent metadata ingestion and registries of RMI could be the object of a public-private partnership, whereas licencing, asset management and storage, and distribution will surely remain the remit of a multitude of sectoral or generic, national or international, commercial or not-for-profit operators.
- **Distinct technologies.** An example of this is assertion-oriented distributed ledgers for registering RMI versus transaction-oriented distributed ledgers for licensing and distribution.

5.5.6. The need for a holistic integration

A holistic integration fulfils the purposes of rights management information through processes (re)engineering and change management (i.e., organisation including cooperation, training including awareness and understanding, and communication). In turn, technical systems support people and processes. The integration of purposes, processes, organisations, and systems is called *holistic* by the study team.

Approaching copyright data management in the digital era would require an iterative holistic integration at three levels: industry practices, data technology, and policy & law. These levels are interdependent. A first iteration would sketch a solution, a second draft roadmaps, and a third design blueprints.

An interview and subsequent discussion with the Estonian Ministry of Economic Affairs and Communications outlined three potential levels of possible interactions between the public sector and creative industries, data scientists, and system engineers:

- The public sector, as an *honest broker*, could help –
 - Prototype, pilot, scale up, and co-finance a registry of rights management information and an intelligent metadata ingestion,
 - Embed the resulting copyright infrastructure in the national digital infrastructure and leverage the national ID system (fostering data authentication),
 - Mediate (including campaigning for cooperation), incentivize, educate (including campaigning for awareness and understanding), and communicate with the national creative industries,
 - Liaise with the European institutions and WIPO,
- The public sector, as a *change agent*, could have to adapt its procedures to define and implement the governance of the national copyright infrastructure (fostering data authority),
- The public sector, as *policymaker*, could have to consider the necessity to adapt national regulations concerning – for example – the copyright title, standards, or the role of Collective Management Organisations.

Creative industries, technologists, and the public sector will continue to benefit from the **joint development and deployment** of data solutions.

Moreover, **cross-sectoral collaborations** will not only allow to address data issues related to multimedia, mash-up content, or hybrid user generated content, but will uncover new business opportunities.

Similarly, **cross-national data solutions** will match the nature of the Internet – a **global** system of interconnected computer networks, and the objectives of the **Digital Single Market**.

5.5.7. The lowest common denominator to frame rights metadata

In view of the complexity and sectoral intricacies of copyright data management, it appears reasonable to consider the **lowest common denominator** which should benefit all stakeholders on the content value network.

This common denominator would be an open⁶⁰⁰, distributed, and transparent rights data framework. Its remit would be limited to:

- **Accepting** rights declarations,
- **Attributing** works and other subject matters to creators and rightholders,

⁶⁰⁰ "Open" does not mean "free of charge" but offered at fair, reasonable, and non-discriminatory terms (FRAND conditions).

- **Pointing** to the rights management information necessary to licence, distribute and remunerate works and other subject matters,
- **Verifying**, i.e., attesting attributions,
- **Answering** queries, i.e., providing verified rights management information.

This framework could be materialised through:

- **Non-proprietary standards** of identifiers, metadata formats and exchanges to ensure normalisation, interoperability, and transparency,
- **Inclusive architecture**⁶⁰¹ to cater for any rightholder, stakeholder, incumbent or new intermediary,
- **Proven techniques** to address issues and integrate solutions compliant with industry and jurisdiction practices.

5.5.8. The potential role of Artificial Intelligence

Artificial intelligence could be particularly relevant to rights metadata:

- **Rights management:** (a) one needs to distinguish between humans creating copyrightable work and other subject matters and machines that do not produce copyrightable content, and (b) one may be able to apply pattern recognition, rule-based algorithms, and machine learning to create, curate and clean rights management information.
- **Rights licensing:** (a) expert systems can support humans to automate copyright-data-based licenses and micro-licences, and (b) AI can be used to inform rightholders about the qualitative and quantitative use and monetisation of their works, whether licensed or not, and put them in a better position to negotiate and enforce agreements.
- **Rights enforcement:** the EUIPO launched a study on the impact of AI on infringement and enforcement of copyright and design. It will suggest a mechanism for collecting and presenting new AI applications relevant to copyright infringement and enforcement to maintain the information gathered during the research and produce training and capacity building materials in relation to AI and copyright.

Algorithm transparency: as mentioned in the introduction of the proposed Digital Services Act, the use of metadata by online platforms to feed AI algorithms and promote content still lacks transparency, e.g., for the establishment of playlists or display of search results. Authors and performers are requesting more transparency on the exploitation of their works and more information about monetisation to balance the bargaining powers between them and the platforms, increase fairness and secure cultural diversity⁶⁰².

⁶⁰¹ It could – for example – adopt the OpenAPI specification for machine-readable interface files for describing, producing, consuming, and visualising web services. See <https://www.openapis.org/>.

⁶⁰² G. Mazziotti, A Data-Driven Approach to Copyright in the Age of Online Platforms, EUI Department of Law Research Paper No. 2020/07, 2020 available at SSRN: <https://ssrn.com/abstract=3655027> or <http://dx.doi.org/10.2139/ssrn.3655027> and D. Antal, A. Fletcher, P. Ormosi, Music Streaming: Is It a Level Playing Field? 2021 available at <https://www.competitionpolicyinternational.com/music-streaming-is-it-a-level-playing-field/>

5.6. Annex 6: Outline of an Open Rights Data Framework

In this outline, the study team develops the concept of open rights data framework mentioned in 2.7.4.

The study on Copyright Data and New Technologies asked the question: “**could digital technologies help solving some copyright data issues?**”. The answer is positive. Eight months of primary and secondary research including tens of interviews with experts from all sides of the creative ecosystem have led to the results given in other sections of this report.

They have also led to the formulation of avenues for the semantic and technical development of an **open rights data framework** presented first during the workshop that took place in June 2021 and developed afterwards with rights metadata experts.

5.6.1. Functionality

An open rights data framework should be:

- **Available to anyone who wants to use it:** any EU citizen, resident, and business in the Union who would like to make use of the framework would be able to do so.
- **Widely useable:** the framework should be useable widely as a way to build applications answering the questions who did what, who owns or controls what, what can one do with it, and what was done with it for the purpose of creating, accessing and remunerating digital content.
- **At rightholders’ discretion:** the framework should enable creators and rightholders to choose which rights information should be shared with third parties, and to keep track of such sharing.

While it may be defined initially in relation to the EU, such a framework must be designed to extend to support content and rights for any territory or jurisdiction.

A minimal application of the framework could be a **voluntary standardised declaration of rights holdings**, not the development of a central registry. It would be essentially simple in function, enabling standardised, all-media declaration and discovery of rights information, providing links to licensing, monitoring, usage reporting, payments or other commercial or cultural services associate with rights management.

To ensure the integrity of the network, this declaration option should be enhanced by automated mechanisms to remove duplicates, check data validity⁶⁰³ and identify conflicts between declarations, incentives would need to be defined, mechanisms developed, costs estimated, and liabilities of data providers limited.

Validation and publishing of declarations could be integrated with the assignment of a rights identifier (as described in Annex 5.7) and be managed by agencies typically on a national or supra-national basis⁶⁰⁴, as is the case with the administration of many standard identifier systems (for example, DOI, ISBN, ISRC, ISWC). Agencies could administer declaration and validation processes and maintain publicly accessible and searchable datasets of rights declarations within a federated agency network. National authorities could accept the

⁶⁰³ Validity here refers only to automated checking for formal compliance with the data standards, as would be done by any computer system ingesting data through an API, not to any investigation of the validity of the claims to rights being made. The authority and liability for all such claims rests with the party asserting them.

⁶⁰⁴ Such agencies might or might not be associated with national IPOs. Provided a standard declaration and identifier is agreed there may be various models for its administration according to national or sectoral interests.

(minimal) apparent extra burden for the sake of the resilience and recovery of their creative industries. The workload could be shared between them, benefiting creative and cultural sectors, creators, rightsholders, publishers, CMOs, Independent Management Entities, platforms⁶⁰⁵ and user alike.

Any individual person or organisation would be free to make declarations on their own behalf or others. However, many existing rights data sources (such as the ones in charge of the registration of standards mentioned in annex 5.3, CMOs' and publisher databases, service companies, etc.) would likely be able automatically to provide standard declarations from their existing datasets and processes at very low cost, as occurs now with the production of many standard metadata messages in standards such as ONIX, DDEX or IPTC.

5.6.2. Technology

Digital technologies are available to open and integrate the rights data framework. Of course, integrating them as needed would require a programme of design and development.

There are many existing applicable **technological assets**. Wherever possible the framework should consist in opening and enhancing the existing technology toolbox including technical architecture, standards, and guidelines based on best practices. An important intellectual and network capital is available (see section 2.4). The opening could be fast-tracked through close co-operation with the stewards of the existing rights data framework; most of them are not-for-profit organisations maintaining, developing, and promoting foundational standards and technologies⁶⁰⁶.

Generic technologies and standards, for example, cloud-native applications; digital fingerprinting, watermarking and identification; data exchange and transformation languages; artificial intelligence; ontologies; distributed ledger technologies; high-speed database machines could be deployed to support the generation, integration, declaration, validation and querying of standardised rights data. These same tools can be (and are) used in rights licensing, enforcement, use monitoring and reporting and remuneration, which are in the remit of essential systems surrounding the framework but not part of the co-ordinated framework itself.

The following five current technology developments could play a significant role in an open framework:

- **Non-proprietary content-dependent identification mechanisms**, for example the developments around the International Standard Content Code (ISCC, see section 2.4.), would depend neither on registration authorities nor on commercial companies. They could bridge some gaps between standard content identifiers such as ISRC, ISWC, ISBN or ISAN and digital manifestations of the content they denote and enable the matching of digital objects to be available to anyone.
- **Digital identity wallets** include the developments around self-sovereign identity at the European Blockchain Partnership (EBP)⁶⁰⁷, several commercial products, and

⁶⁰⁵ The benefit to social media and other platforms hosting large volumes of digital content of having trustworthy independent sources, include conflict identification, for establishing rightsholders of content which may be published by anyone on their platforms is potentially very considerable.

⁶⁰⁶ Many standards for identifiers, metadata formats, and data exchanges emerged from the further developments of European-funded initiatives such as <indecs>. <indecs> developed a framework of metadata standards to support eCommerce based on intellectual property with the support of the European Commission. EDItEUR and mEDRA are two others of the many relevant European projects.

⁶⁰⁷ See <https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/ebsi>

the realisation of the recent policy declaration of the European Commission⁶⁰⁸. These would be useable either to identify users or to prove particular personal attributes, typically to access public or private digital services and so eliminating the need for many personal service accounts.

- **Identifier binding protocols**, for example the developments around traceability at the EBP and the research around a Distributed Trusted Rights Framework (see annex 5.3), would bind immutably open identification of content or rights to the identities of their creators or rightholders. Changeable content or rights metadata would be linked to this binding via permissioned links. This immutable binding mechanism is comparable to bookkeeping, where an entry may not be deleted, a new entry must be added to the previous ones to show what happened with a value or to correct an erroneous entry. This allows an auditor to check an accountant's ledgers. A similar mechanism is used for the registration of nominal shares in a company, or for land or company registries.
- Rights metadata contains assertions about who did what, who owns or controls something, what can be done with it and (with usage metadata) what has been done with it. **Verifiable credentials**⁶⁰⁹, for example the developments around verifiable credentials at W3C, diploma at the EBP, and an Open Content Certification Protocol (OCCP)⁶¹⁰ in Germany, would raise trust in such assertions, and in content authenticity, to ensure the accountability of individuals or organisations, even if they must or prefer to remain pseudonymous.
- Research for **intelligent metadata ingestion** may help resolve data quality issues across data siloes⁶¹¹. It could solve some of the challenges of cleaning and completing metadata related to existing catalogues and cultural heritage. It could make use of methods such as the Linked Open Data project at Europeana (see section 2.6.1), the possibility to use pattern recognition, rule-based algorithms, and machine learning (three AI techniques) to curate missing or erroneous rights management information as exemplified by the machine learning developments of Gracenote Works at Nielsen, or the possibility to use AI to generate descriptive metadata as exemplified by the research at the Swedish Royal Library⁶¹². Note though that work of this kind is always likely to be subject to assertions by verified entities as described in the preceding paragraph.

5.6.3. Characteristics

The public **utility** of open rights declarations built upon an open rights data framework could be comparable to the public utility of land or company registries. The role of rights declaration **agencies** could be comparable to the role of patent offices or trademark & design offices.

An open rights data framework on this model would be:

⁶⁰⁸ See https://ec.europa.eu/commission/presscorner/detail/en/ip_21_2663

⁶⁰⁹ "A verifiable credential can represent all of the same information that a physical credential represents. The addition of technologies, such as digital signatures, makes verifiable credentials more tamper-evident and more trustworthy than their physical counterparts", definition from the W3C Verifiable Credentials Data Model 1.0 available at <https://www.w3.org/TR/vc-data-model/#what-is-a-verifiable-credential>

⁶¹⁰ See <https://posth.me/occp/>

⁶¹¹ These processes go beyond from the formal semantic mappings and transformations between defined schemas which are essential for interoperability, and which are subject to formal validations as referenced in Annex 5.3.

⁶¹² See <https://www.kb.se/in-english/research-collaboration/kblab.html>

- Compatible with the Berne Convention as the rights declarations would remain **voluntary**.
- Supporting a series of **interoperable sectoral or jurisdictional policies**; it would not need one governance that fits all.
- Contributing to the development of a genuine **single market for data**, in compliance with the GDPR, and respect of business confidentiality.
- Applying the proven **Once-Only Principle** to reduce administrative burdens by asking individuals and organisations to provide standard information only once.
- Helping define **responsibilities** pertaining to rights data management. This means not only helping realise the potential of the *acquis communautaire* but also paving the way for the implementation of ongoing regulatory initiatives such as the Digital Services Act, Digital Markets Act, or Data Governance Act. Indeed, the proposed **Data Governance Act** aims at fostering the availability of data for use by increasing trust in data intermediaries and strengthening data-sharing mechanisms across the European Union.

5.7. Annex 7: An interoperable network of identifiers and schemas

This annex⁶¹³ outlines the rationale, goals and scope of an Open Rights Data Framework and identifies the 'black hole' in the network which needs to be filled with a Right Identifier and a Rights Declaration.

5.7.1. Context of this proposal

In presenting the preliminary findings of the study "Copyright Data and New Technologies" conducted on behalf of the European Commission in June 2021, the study team concluded with two questions:

- *Can an Open Rights Data Framework (ORDF) help release the digital potential of European Creative industries?*
- *What governance is needed to trust rights management information and what could be the role of the Commission and public authorities?*

The preceding slides had also referenced the need for a **Rights Declaration** at the heart of such an ORDF.

This Annex 7, written in response to these two questions, proposes that an Open Rights Data Framework with a Rights Declaration at its centre is not merely helpful but essential if significant progress is to be made in **interoperability** across diverse supply chains of content and rights information on the internet. It outlines a structure and governance for such a framework based on work already done (much of it funded by the EC), and the role which might be played by the Commission in its establishment.

⁶¹³ Annex 7 has been written by Godfrey Rust, co-author of the <indecs> framework, technical lead of the Linked Content Coalition standards and the RDI project and consultant in the development of many standards including DOI, ISWC, EIDR, ONIX and DDEX. The proposal has the personal endorsement of several other established experts from the content and standards domains: Mark Bide (<indecs>, EDItEUR/ONIX), Michael Healy (ISBN, ISTC, ISNI), Mark Isherwood (DDEX), and Angela Mills-Wade (LCC). Godfrey Rust is the rightsholder of Annex 7.

An ORDF can address three main challenges which are major constraints on the health and growth of the digital content network:

- many Parties have no way of **declaring** their Rights holdings in a way that is automatically understood by computers and enables them to exercise control over them;
- systems often have difficulty **discovering** the holders or status of Rights in an automated way so that licensing, payments or other transactions can be carried out; and
- **conflicts** in claims of Rights holding, whether intentional or otherwise, are widespread and often go undetected and/or unresolved.

Meeting these challenges is the goal of an ORDF, but that is not an end in itself. Better and automated access to accurate rights information will increase the legitimate use of content, whether for commercial, cultural, educational, social or other purposes, and improve the protection of the rights of its creators and other interested parties.

This proposal draws on more than 20 years of work in the area of rights data interoperability by many people going back to the <indecs> project, and on the consensus views of many of the content rights data standards bodies expressed through the analyses of the Linked Content Coalition and their proving through the RDI and ARDITO projects.

5.7.2. What is data interoperability?

Data interoperability is about the automated flow and processing of data among the systems of a network. In an ORDF it means that *computers can automatically process⁶¹⁴ rights data unambiguously even where it is originally expressed in different formats, vocabularies and identifiers in different parts of the network.*

It is, first and foremost, a **semantic**, not a technical problem: it is about meaning. If computers cannot recognise what an identifier or a term represents, they cannot say or do anything about it unambiguously.

The primary building blocks of data interoperability are common **identifiers** and data **schemas⁶¹⁵**. An identifier means that a computer can know what something is, and a schema means that a computer can understand and then often act automatically on things that are being said about it.

Some of these are **standardised** under shared governance for different sectors (such as the message schemas managed by ONIX, DDEX and IPTC, and identifier systems such as ISBN, ISRC, ISWC, ISNI and DOI). Others are **proprietary**: that is, managed by a content provider, intermediary, platform, user or other supply/value chain participant. Functionally there is no difference: if an identifier or schema is in shared use within the network then it should be able to interoperate regardless of whether it is standardised or proprietary.

The best mechanism for interoperability is that everyone in the network uses the same identifiers and schemas, but in reality, and often for good reasons, this only occurs to a limited extent, and usually in particular market sectors (for example, in the book supply

⁶¹⁴ The term “process” here covers all aspects of input, reading, querying, transformation and output of data.

⁶¹⁵ “Schema” is used here as a general term for any data format, standardised or not, in which data may be stored or communicated. DDEX, ONIX and IPTC are important examples of standardised families of message schemas, but every database and metadata interaction between computers uses some schema(s).

chain the ISBN identifier and ONIX product message are more or less ubiquitous and result in extensive interoperability between systems).

So, the mechanisms needed for fuller and cross-media data interoperability are the **linking of identifiers** and the **mapping of schemas** which enable data to be processed across different chains with the necessary translations taking place automatically and in a way that can be trusted: the former is increasingly taking place initiatives like ISNI and EIDR, and the latter was comprehensively demonstrated in the RDI project.

However, to get to the heart of this it is necessary to understand the scope of the ORDF by describing its **entity relationship model**⁶¹⁶: the kinds of things for which identifiers and schemas are needed in an ORDF, and the main connections between them. This allows us to see the critical data problem at the centre of the network.

5.7.3. ORDF: a network of linked Identifiers

Figure 76 below, based on the LCC Rights Reference Model⁶¹⁷, shows the six main types of data entity in an ORDF and the main kinds of links between them. It applies to any market sector and type of content or right:

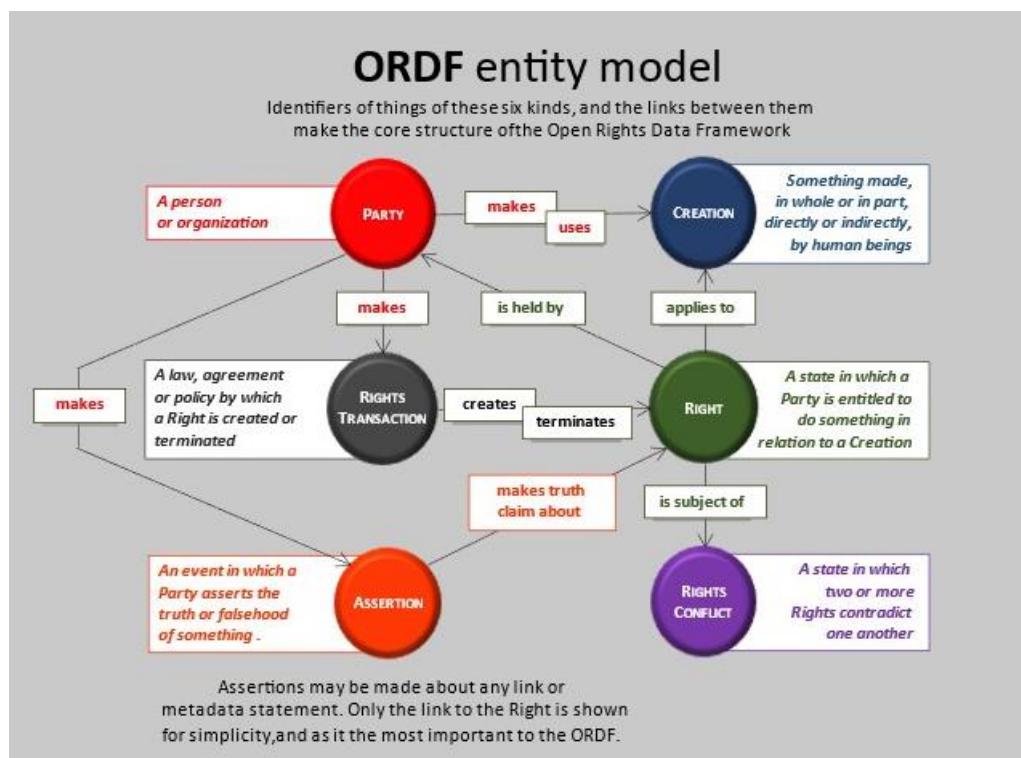


Figure 75: ORDF Entity Model
Source: Godfrey Rust, Rightscom Ltd

⁶¹⁶ An entity relationship model is normally the first stage of any data modelling which defines the scope of the model or system being designed.

⁶¹⁷ <http://www.linkedcontentcoalition.org/phocadownload/framework/The%20LCC%20Rights%20Reference%20Model%20v1.0.pdf>

Each entity in this network must have an identifier ("ID") if computers are going to be able to recognise it unambiguously and process information about it. To a computer, the network needs to look like this if it is to function properly (Figure 77):

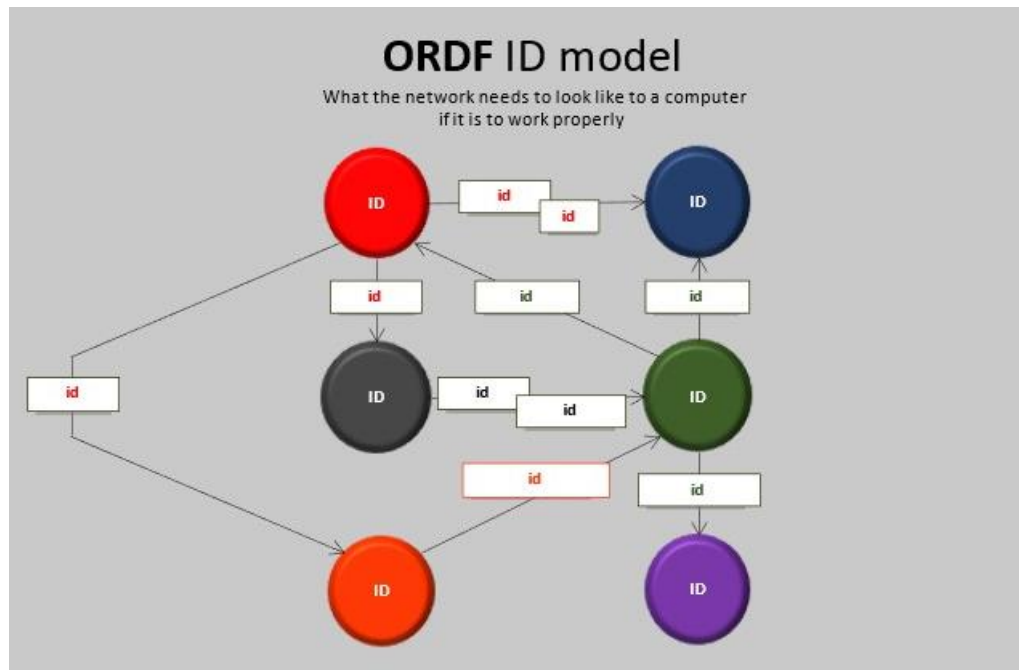


Figure 76: Computer-eye view of an ORDF

Source: Godfrey Rust, Rightscom Ltd

Each of the first three types of entity – **Party**, **Creation** and **Rights Transaction** (which includes all kinds of license) – will normally appear in any rights data model⁶¹⁸. There are large numbers of identifiers and metadata schemas (both standardised and proprietary) in use in supply/value chains for each of these. This is indicated simply below in Figure 78 in the blue shaded area. Whatever happens in relation to an ORDF, these data elements will continue to operate, and new ones will appear: a successful ORDF will not replace these but enable them to interoperate.

⁶¹⁸ Of course, the names used, and the detailed descriptions, will vary.

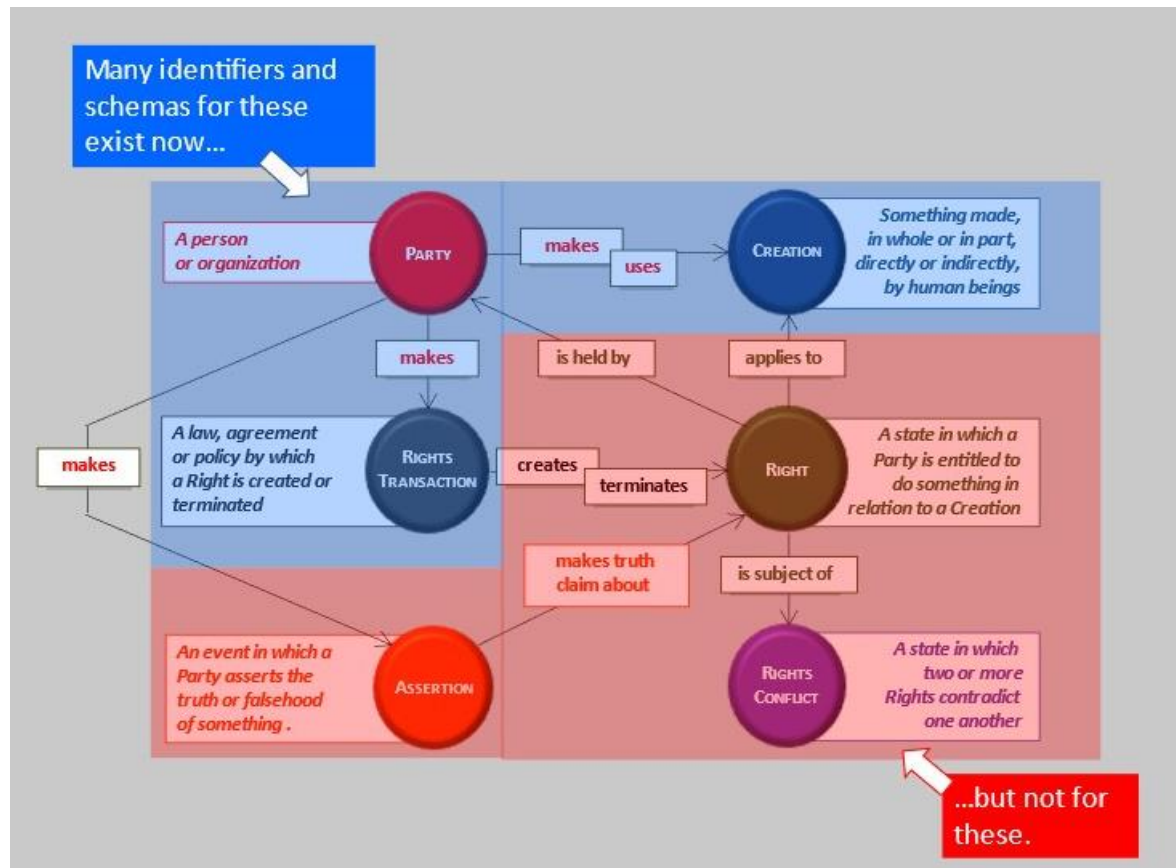


Figure 77: ORDF Entities – public identifiers and schemas

Source: Godfrey Rust, Rightscom Ltd

However, the situation is quite different with the **Right**: there is no public standard Right identifier, and few proprietary ones in use anywhere. So, remarkably, the entity which is being traded through the rights data network – the Right itself – *cannot be simply identified by computers*. As the figure shows, it is the Right which connects to all of the identifiers of the other entities, yet it has no identifier of its own with which to hold them together.

The remaining two entities shaded in red (the **Assertion** and the **Rights Conflict**) suffer from the same problem because they depend for their existence on the Right. The Assertion has nothing to connect to, and the Rights Conflict is an entity which connects two or more things which, in computing terms, do not exist, so these two entities typically do not exist either.

What accounts for this “black hole” in rights data?

5.7.4. Why is the Right “missing”?

Rights are not publicly identified as distinct entities because rights information is typically treated as *metadata belonging to something else*, usually a Creation or a Rights Transaction, such as a licence. Historically this is understandable because these are the contexts in which rights data is typically published and found.

But analysis shows this to be very misleading. The LCC defines a Right, in data terms, as “a state in which someone is entitled to do something with something.”⁶¹⁹ Systems have no problem identifying events and states as data entities with identifiers (orders, deliveries and payments in the supply chain do this all the time), but in general have not in the past done this with Rights.

When a Right is modelled in this way, it is immediately clear that one Right may cover many Creations, and that one Creation may be the subject of many different Rights. Only by recognising this many-to-many relationship (which is abundantly clear, for example, in Rights granted to major platforms or broadcasters to use entire, dynamic “repertoires” of content) can rights data be represented accurately in a way that can support the requirements of the rights data infrastructure. It already happens in some closed systems: it now needs to be brought into the public network.

The work of <indecs>, LCC and the EU’s RDI and ARDITO projects has successfully realised and tested this analysis over a wide range of types of Creation, Rights Transaction and Right. Schemas such as ODRL and RightsML are also consistent with this analysis, although their focus is on the Rights Transaction and so the Right itself had no distinct identifier.

What is needed at the heart of an ORDF is a machine-interpretable **Rights Declaration** – a “digital © Notice” – which can meet the three challenges of an ORDF identified above:

- **Declaring Rightsholdings**

Parties of any kind should be able easily to declare accurate descriptive, attribution and Rights information about Creations of any kind in computer-interpretable forms (if they wish to⁶²⁰).

- **Discovering Rightsholdings**

Parties of any kind should be able easily to access descriptive, attribution and Rights information about Creations of any kind in computer-interpretable forms (where others have chosen to make it available).

- **Detecting Conflicts**

Automated methods should exist to detect conflicts in descriptive, attribution or Rights information, and report them to those who have declared it (but not to resolve conflicts automatically).

It is not that the rights data does not exist (although in many contexts that is also unfortunately true): the data may exist somewhere, but not in the form needed for interoperable declaration, discovery and conflict identification.

The nearest thing we have to a widely used rights declaration at present is the © **Notice**, which in the digital network is often no better than a human-readable clue to where to start to look: it is ambiguous (it uses names, not identifiers), unspecific (it might apply to any time or place), unaccountable (its source is unknown), it may be wrong (rightsholding changes) and it cannot be used for the granular details of rights that are delegated through a chain of agreements.

The <indecs>/LCC/RDF/ARDITO work provides a blueprint for a solution, but we should ask: if this is such a serious and basic problem, why has the market not fixed it so far?

⁶¹⁹ This is not, of course, intended to be a legal definition but a broader generic definition for the purpose of data modelling, and it includes copyright and other content rights as well as any other kind of entitlement arising from policies and agreements.

⁶²⁰ There is no intent anywhere in this proposal to oblige any Party to disclose or publish any information which they have no wish or legal requirement to.

5.7.5. Why is an independent ORDF needed?

Apart from the fact that it exists in something of an “analysis blind spot” as described in the previous section, there are two good reasons why the market has not yet responded to the challenge of a Rights Declaration and Rights Identifier.

First, successful content data interoperability standards such as ONIX, DDEX, IPTC, DOI, EIDR and most of the range of “ISXX” identifiers operate within recognisable content sectors and supply chains whose major players have found it worthwhile to invest in open standards solutions whose impact is immediately clear. The ORDF Rights Declaration must be *cross-sector, multi-functional and all-media* as well as international, and getting major players to step up and invest time and money into setting it something so broad, while they have so many other issues to deal with, is not going to happen without high-level leadership or incentive. However, if a standard mechanism is established and it is made both easy and cheap to participate as a by-product of existing operations, there should be many practical benefits and market opportunities which can make it successful.

Secondly, as the Copyright Infrastructure and New Technologies project has shown, this is an immensely complicated network, and it has been hard “to see the wood for the trees” in terms of a way forward. The <indecs>/LCC/RDI work sets out not only a clear analysis of the required data framework, but crucially a mechanism (though the LCC Entity Model⁶²¹) for supporting the interoperability of other existing identifiers and schemas so that an ORDF Rights Declaration and Identifier can be introduced without competing with or compromising any existing data standards or systems: however, it has needed the right market conditions for that to be taken up at the appropriate level.

The question must be asked: why would people take the trouble to create Right Declarations? There are three main reasons:

- **to make money**, or more money, from their content,
- **to protect** their content from infringement, and
- **to be attributed** for their content.

Each of these requirements is growing daily.

The present lack of any solution at all for a Rights Identifier is also itself an opportunity. If the EC, with others supporting them, were to step forward to lead an ORDF initiative centred on a Rights Identifier and Rights Declaration, it would be moving into unoccupied territory which makes it a competitor or threat to no-one, and initiating a service which no-one is obliged to join but which most network participants can see would be of general benefit if it succeeds. As the ORDF data model shows, this work is not marginal: if successful it may be a catalyst every bit as powerful in the global rights network as standards initiatives like ISBN, DDEX and DOI have been in their own supply/value chains.

5.7.6. The scope of an ORDF

The ORDF must be **neutral** in commercial, cultural, technological, legal, political and other ways, and must support the expression of any kind of Right or Rights Transaction in any kind of content or media. Unless, like other successful cross-domain standard identifiers such as UPC, URI or DOI and supply chain standard schemas like ONIX and DDEX, it is

621

<http://www.linkedcontentcoalition.org/phocadownload/framework/The%20LCC%20Entity%20Model%20v1.0.pdf> Note that an updated version is in progress which enhances but does not significantly change this model.

designed to serve the interest of all legitimate participants in its supply chain (which for the ORDF means, in effect, anyone) it is likely to fail.

The network is complicated and an ORDF should **not reinvent wheels**. Many tools and standards are successfully embedded in supply chains and much technology is designed around them. An ORDF must work with existing schemas and systems both at a technical and a governance level: it will only succeed by offering solutions which are seen to add value to existing operations.

Within a supply chain, data interoperability between systems is achieved by two main methods:

- **universal standards** for identifiers and schemas which everyone uses; and/or
- **mappings** between different identifiers and schemas so that data can be automatically translated from one to the other.

It is likely that a mixture of these two methods will be used in any supply chain, and inevitable that both will be used in an ORDF. There are a large number of different supply chains with their own established standards and practises which intersect and overlap.

Methods are needed for these to interoperate so that data can move out of a particular supply chain and into the network in general.

As far as possible interoperability should be **automatic** and **invisible** to users. An ORDF requires that Rights Declarations may be created in an interoperable form **at the point of entry** into the network (although they can be created at any time). For individual creator-publishers that requires tools which are integrated with the applications which load content to the Web, and for corporate systems it requires tools which enable data within a corporate system to be automatically converted into interoperable forms. Such processes, once set up, should be very cheap or free to use as Rights come in and out of existence. Both of these kinds of process are already commonplace within the network for specific supply chains. An ORDF governance will not build or operate such tools but define standards which will enable their spread and interoperability to the network as a whole.

Such a framework requires **governance** and some **investment** to get off the ground, with its aim like other such standards-based initiatives to become self-supporting through the backing of those who benefit from it. At the same time, participation in its interoperability must be as cheap and easy as possible so that as far as possible its presence becomes invisible to content providers intermediaries and users alike.

Issues of **trust** will be central to its success, and the ORDF governance will play a key role in the certification or validation of methods or services through which identifiers and schemas are mapped or linked to establish interoperability between them.

It is not the task of an ORDF to determine the legitimacy of any individual claim of right or to arbitrate between claims, agreements or jurisdictions, but it will play a key role in enabling the automated identification of conflicts between Rights Declarations for others to resolve. It will also play a role in the **authentication** of the identities of those making Assertions.

5.7.7. Specification: LCC-based Rights Declaration schema

To fulfil its central role, the ORDF Rights Declaration schema must be able to support all of the characteristics of rights management and data noted in the next two sections, and do so for any right, content and media type. While many Declarations will be simple, others

will be complex, and the schema cannot impose arbitrary limitations on completeness. This means it must be able to represent data from any other schema and vocabulary whose users wish to interoperate with it.

This is a very demanding specification but is exactly what was achieved in the EU-funded RDI (Right Data Integration) project in 2015 using the LCC data models. Limited versions of this architecture have since been used in the ARDITO system and form the basis of the system supporting the music industry DDEX data standards.

The longer paper of which this is summary sets out a full draft specification of a Rights Declaration based on the LCC models, to form the basis of the technical work to be done by the ORDF and support the characteristics described in the next two sections.

5.7.8. Requirements: characteristics of rights management

Below are listed many common characteristics of content rights management which may affect rights in any types of content, and so must be supported by an Open Rights Data Framework (ORDF) and its Rights Declaration. These are elaborated in more detail in the fuller version of this paper:

- **Public and private rights data:** no-one is forced to declare anything.
- **One object, many Creations, many Rights:** A digital object commonly includes many Creations and therefore many sets of Rights, which may be layered in different modes (Items, Manifestations, Expressions, Works) or grouped as components side by side.
- **Versions of Creations** may have different Rights and rightsholders.
- **Rights data at the point of publication:** successful standards establish authority and declarations as high up the supply chain as possible.
- **Rightsholders change:** an ORDF must support rules which enable automated updates arising from changes such as terminations and duplicate and conflict detection.
- **Three main groups of Rights:** Use Rights (such as copying), Control Rights (such as licensing), Moral Rights (such as attribution) must be capable of declaration.
- **Attribution:** is needed but does not automatically imply rightsholding: declarations must be explicit.
- **Chains of Rights:** the trail through which Rights are delegated or assigned.
- **Conditions of Rights:** a Right may be dependent on a Pre- or Post-Condition which may be any kind of event or state.
- **Fragmentation of Rights:** all aspects of Rights (time, place, action, purpose, context etc) may be fragmented in any combination as Rights are delegated though a chain.
- **Bundling of Rights:** Rights Transactions often relate to bundles of all sizes ("repertoires" or "catalogues") whose contents commonly change over time.
- **Rights in Rights:** The combination of fragmentation and bundling means Rights are best expressed as "Rights in Rights" not directly "Rights in Creations".
- **Prohibitions and exceptions:** Prohibitions are not routinely needed as Rights do not exist unless they are explicitly permitted, but prohibitions are commonly

expressed as exceptions to permissions (e.g., “in Europe but not Germany or Austria”).

- **Exclusivity:** is a critical factor to enable conflict detection.
- **Assertion of Rights:** the truth or falsehood of a Right may be asserted by any Party, not only or necessarily the rightsholders.
- **Termination and reversion:** when a right is terminated a new Right must be created, which may often be a result of reversion to the previous rightsholder.
- **Rights Conflicts:** are widespread due to errors, outdated information, genuine disagreements and deliberate false claims. Conflicts may be hidden because of bundling or fragmentation.
- **“Free use” licences:** such as Creative Commons may be declared as Rights.
- **Public domain works:** should be explicitly declarable explicitly (by who?).
- **Orphan works:** should be explicitly declarable explicitly.

5.7.9. Requirements: characteristics of rights data

Below are listed aspects of representing Rights in data which must be accommodated by an Open Rights Data Framework (ORDF) and its Right Declaration, to ensure support for the characteristics of Rights management described in 5.3.8. These are elaborated in more detail in the fuller version of this paper:

- **Identifiers:** computers can only communicate unambiguously and automatically if the things they are talking about are denoted by identifiers (names which are unique within their domain).
- **Identifier mapping:** entities often have multiple identifiers of different types: services like ISNI and EIDR create links between them which must be brought into automated use through the ORDF; other services create links between related identifiers which can also come into ORDF.
- **Schema mapping:** authorised mappings are needed from schemas to the ORDF Rights Declaration, and from schema to schema.
- **Value sets, mapping and ontology:** Terms in controlled vocabularies (“Value sets”) are identifiers. All major schemas use value sets and the terms in them must be mapped to one another and the ORDF vocabulary through an ontology to support translation, hierarchical querying and complex one-to-many term mappings.
- **Rights** are expressed as possible events and states. Conditions/constraints on rights may be expressed as any kind of event or state at any level of detail.
- **Completeness:** if a Rights Declaration is incomplete, it will often be incorrect.
- **Time Values:** are critical for Rights and must be represented by identifiers (as ISO 8601 values); times may be Points or Periods and also be qualified by proximity (e.g., “circa”, “before”, “not later than” etc).
- **Measurements:** are critical for Rights and their basic values (numbers) are identifiers; measures may be single values or in ranges and are qualified by units of measure and proximity (e.g., “exactly”, “approximately”, “not more than” etc).
- **Multiple values, AND/OR/NOT** known as Boolean operators: most attributes of Rights (rightsholder, Creation, permitted action, time, place etc) may have multiple occurrences which may be constrained by any of these.

- **Use of variable values:** where one value (e.g., the start time of a Right) is dependent on another (e.g., the time of signing an agreement).
- **Offers:** rightsholders will often wish to publish the terms of licenses on offer along with their Rights Declarations.

5.7.10. Governance of the ORDF

The ORDF should be governed by a new standards body, the core of which should be a **consortium of specialized data standards bodies**⁶²² from commercial, cultural and social sectors, with interoperability as its scope. However, as has been proved recently by the success of DDEX⁶²³, a broad initiative of this kind will only succeed if there is a balanced representation, so rightsholder and user interests, cultural and educational networks, platforms and technology providers and standards must have a voice where they are not already represented through sectoral standards groups. It is also essential that the ORDF governance has the strongest liaison with and support from WIPO/IPOs and other institutions concerned with legal issues in content rights.

ORDF governance should:

- **Develop and manage data standards** for interoperability, including for the Rights Declaration and Right ID.
- **Appoint agencies** to issue Right IDs and manage Rights Declaration validation, conflict detection and publication.
- **Manage trust lists** of participants⁶²⁴.
- **Certify services** for registering schemas, and for services which map schemas, vocabularies and identifiers as being ORDF-compliant.

It should not:

- operate data registries or services dealing with individual items of metadata,
- compel or enforce the declaration of any data,
- mandate the use of particular technologies, beyond formats of standards,
- create new data or identifier standards for sectors.

While ORDF standards may be ISO- or CEN-certified in due course, the ORDF should follow the route of successful content data standards in many sectors and first create its own specialized governance to develop, test and introduce what is needed, in order to have the required focus and momentum. It will require some paid personnel resources, as other content standards bodies have, but it will succeed (as others do) only if it wins the confidence and expert engagement of its constituent members.

⁶²² Such as ONIX, DDEX, IPTC, IDF, ISO "ISXX" agencies, Creative Commons, RDA, CIDOC. The LCC was created on the same basis (with more limited membership) and shows the willingness of such bodies to co-operate for common interests.

⁶²³ DDEX, launched 15 years ago and now with the broadest range of international content data standards of any sector body, is a useful partial model for a successful standards initiative balancing different commercial interests and based on the same extensible data modelling as LCC. This is explained in more detail in the longer paper.

⁶²⁴ See <https://helpx.adobe.com/document-cloud/kb/european-union-trust-lists.html>

If successful, ORDF governance will be a catalyst for interoperability between standards and systems of all kinds within the rights data network and a focal point for the development of solutions for rights-related data and technological interoperability.

ORDF is only likely to succeed if major cross-sector institutions with a wide interest in the health of the content rights network, such as the EU and WIPO, support its setting up. It must become a global initiative and so high-level participation from America and Asia in particular should be canvassed. At the same time, to succeed it will be a balancing act between attracting sufficient early support and not creating something top-heavy and too easily paralysed by the weight of counter-productive self-interest at the outset. In time the ORDF should be funded as far as possible via the organizations providing services which use it and benefit from it.

5.7.11. Implementation of the ORDF

The ORDF cannot be a centralised system: the network will not tolerate a single organisation acting as universal registrant or manager of critical data. The design must be for a federated data network based on standards, not a centrally stored or managed dataset. The ORDF governance should define the essential rules and processes, but the system must as far as possible be technology-neutral and open to anyone to provide services at any point in the network. Routine processes for validating individual Rights Declarations must be fully automated.

Declaration: Any party may create Rights Declarations compliant with ORDF standard(s). Wherever possible these should be created by processes and apps which make them by-products of existing systems for managing rights data or publishing content on the Web and therefore require the minimum cost and effort.

Authentication: Assertions of Rights Declarations must be signed/authenticated by a party (an individual or organization) authorised to do by ORDF governance, which may recognise multiple established methods of digital signing or authentication. The ORDF will not itself assert the truth of any Declaration, only the authenticity of its Asserter, but will have authority to de-authorise parties from making Declarations if they consistently breach ORDF policies.

Validation: Any number of organizations or tools may create Rights Declarations but these must be validated (automatically) as formally correct by a certified **ORDF agency**, of which there may be any number. Such agencies will be part of a network which enables the routine detection and automated update of duplicates and conflicts among all datasets prior to publication (mirroring the validation checks which would be carried out in a single database). Services which query Rights Declarations for any other purpose only require access to validated Declarations.

Identification: Rights Identifiers will be issued via ORDF agencies: whether these are directly issued the agency as part of the validation process or issued by the party generating the Declaration will be determined locally (either method may be valid). Other identification agencies successfully use combinations of these methods.

Conflicts: Conflicts between published Rights will be identified automatically by ORDF agencies in the validation process, the interested Asserters notified automatically and the conflict status set on the Declarations according to policies set by the ORDF. Resolution of conflict is out of scope, and data resolution in the network will occur automatically when one or more Asserter amends/withdraws their Declaration.

Querying: Tools and services which query Rights Declarations to discover Rightsholders need access to all available Declarations, and also provide automated access to certified services (such as ISNI) which have mapped identifiers or controlled values and can enable transformation of these values.

Priority and impact: The Right ID and Rights Declarations should be the initial focus of an ORDF, but if successful this can be expected to lead, as has happened with other content standards initiatives, to all manner of developments in the interests of interoperability, and to the ORDF itself becoming a catalyst for cross-sector standards activities. As indicated by the ORDF data model in Figure 1 above, an ORDF Rights ID and Rights Declaration have the potential to become a major focal point or 'switchboard' for enabling Rights Transactions outside of established sectoral supply chains, making critical automated links between Rights and Parties, Creations and Rights Transactions, as well as providing a focal point for the detection and resolution of rights conflicts of all kinds.

The **development of tool and services** to create Rights Declarations from existing or new data and issue Right IDs, to validate, de-duplicate and manage conflicts in them, to query them and then use the results to drive all manner of other processes such as licensing, tracking or usage reporting is a matter for the marketplace. However, critical mass for the ORDF will only come with initial support by body(s) such as the EC to bring together the interested parties to create the interoperable standards, policies and specifications based on the models outlined here and described in detail in the longer version of this paper, and the funding of some initial systems to prove and establish them. This is a challenge, certainly, but there is no other broad and coherent approach on offer to bring some order out of the ever-increasing confusion of the content rights network.

5.8. Annex 8: Exemplary potential benefits of an Open Right Data Framework

An open rights data framework would be adopted if it would **effectively support use cases and applications** providing accurate rights information at the right time so much so that these applications could address industry pain points (see section 2.5) more efficiently: better, faster, and at a lesser expense. The study team explored the potential benefits of an open rights data framework with a few representatives from the music, publishing, and television sectors. A distributed network of rights management information based on a trustworthy, interoperable, accessible, and as comprehensive as possible rights data framework would, among others, address the four following issues.

5.8.1. Addressing inaccurate & slow payments to songwriters

Figure 78 summarises the current data flows and counterflows in the case of music streaming. Music flows red from the left to the right, from the songwriters' brains to the listeners' ears. Remuneration flows green from the right to the left, from the subscribers' pockets, or advertisers' budgets, to the artists' bank accounts. At each step from the left to the right, music – that is data – *carries* rights management information – that is metadata. At each step from the right to the left, metadata *carries* money. On the songs side, the streaming service reports usage, then a CMO or a processing hub, such as ICE Services, identifies what is due and invoices, then the service pays money. These data flows are complex – much more complex than this diagram suggests. Accordingly, they are slow, costly, and prone to inaccuracies.

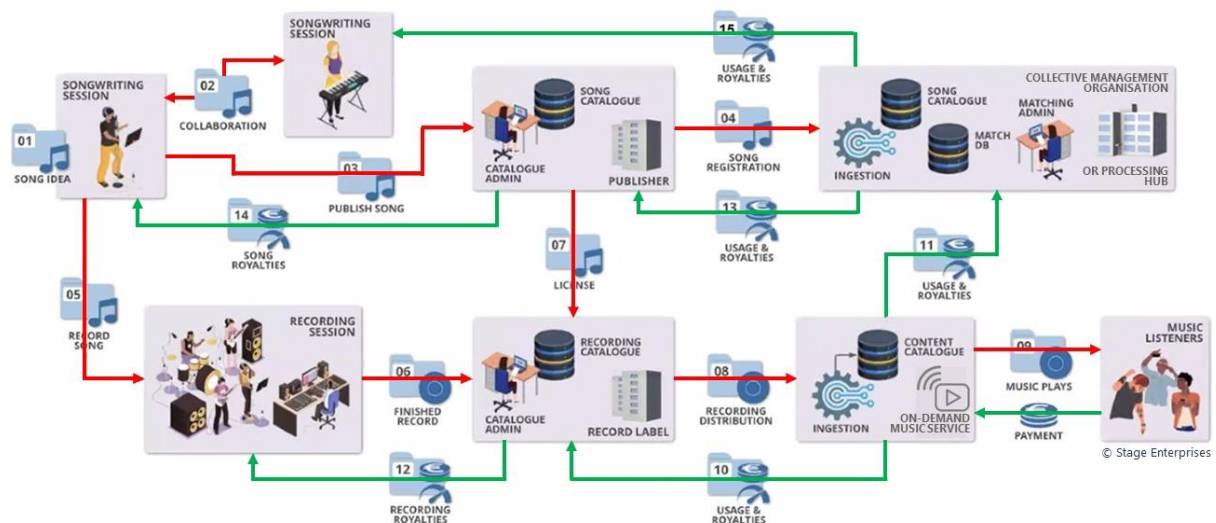


Figure 78: Current dataflows of music streaming

Source: Philippe Rixhon Associates and Stage Enterprises

In the case of music streaming, one could use an identifier binding protocol to bind immutably content-dependent identification of songs or recordings with identities of songwriters, performers or rightsholders. One could do it as close as possible to the moment of creation. One could attach, through permissioned links, sets of dynamic rights metadata, which could be securely augmented, step by step, when the song or its recording would move from the left to the right.

One could develop a data management system to **prevent** the mismatch between ISRCs and ISWCs to occur. It would be built on two premises:

- The enforcement of the moral right of attribution at Step 09; communication to the public would only be possible if the song is attributed to all its rightsholders,
- An identifier binding protocol, linking work identifiers (and subsequently other subject matter identifiers) with rightsholder identities, and rights management information such as ownership splits between rightsholders – from the moment of creation (Step 01 and 02).

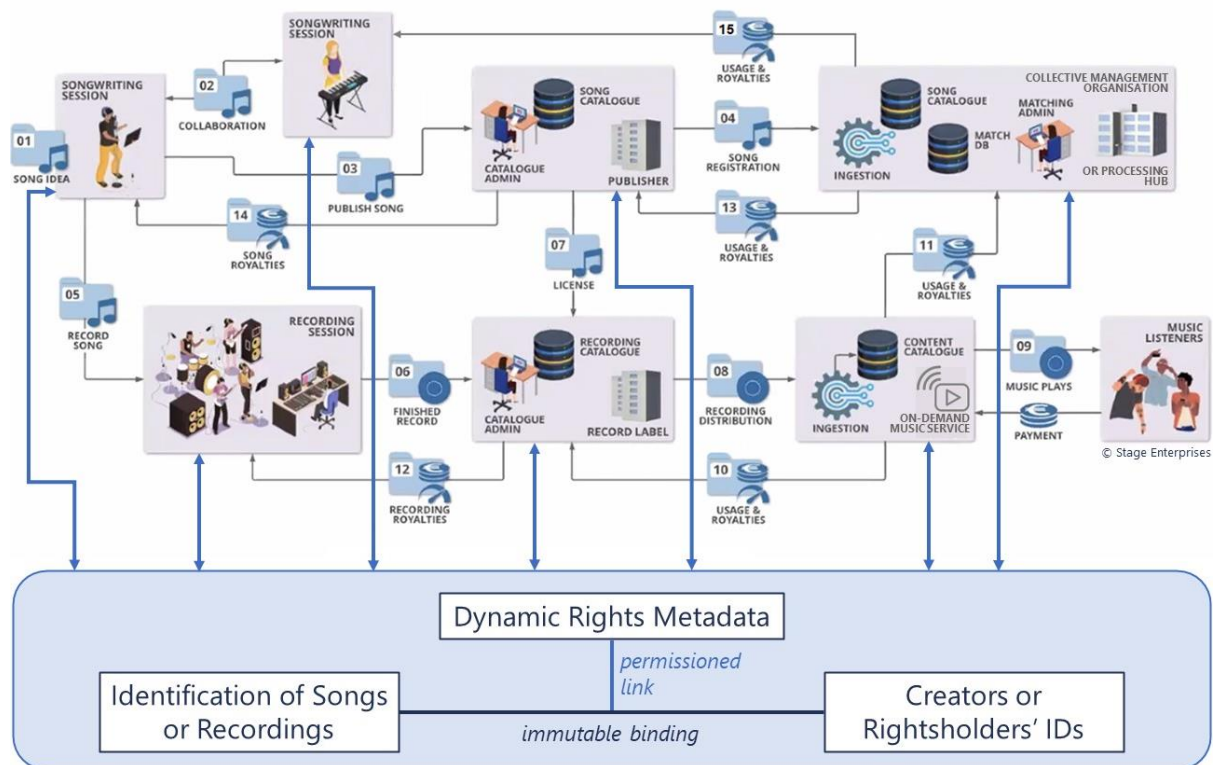


Figure 79: An attribution and verification middleware

Source: Philippe Rixhon Associates and Stage Enterprises

An open rights data framework would allow a middleware to replace the monodirectional red and green arrows by bidirectional blue connectors. The middleware would provide interaction services for software applications via an event-driven and standards-based messaging engine. The MovieLabs Digital Distribution Framework in the film industry, the Digital Data Exchange and Cis-Net in the music industry, and the ONIX standards in the publishing industry are in a sense all examples of such a middleware. An identifier binding protocol would strengthen them.

5.8.2. Addressing the high costs of duplication

Duplication of rights management information within or across databases is a source of errors. On a distributed network based on an open rights data framework, services could emerge to prevent duplications, or to deduplicate or match existing records.

Non-Proprietary Content-Dependent Identification mechanisms (NPCDI) could be used to check a registration against the distributed network of rights management information, flag, and de-duplicate duplicate registrations.

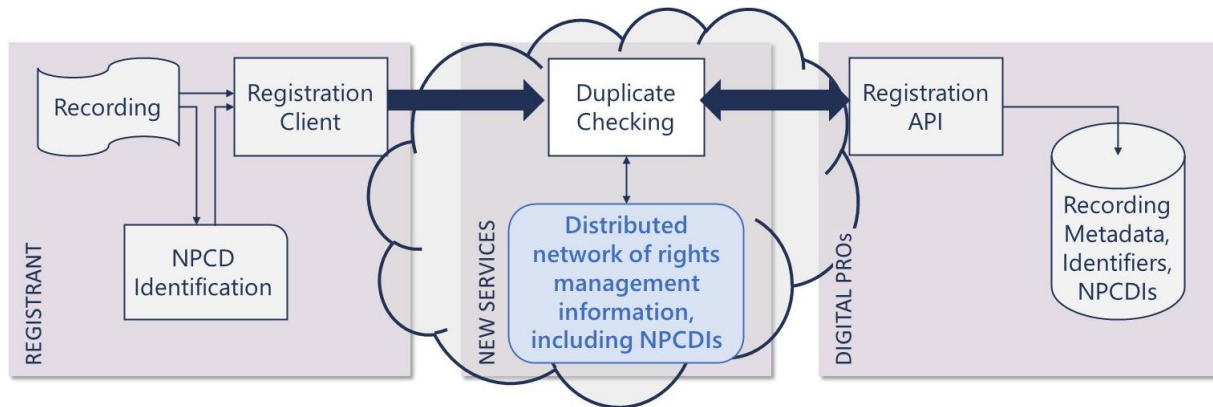


Figure 80: Possible new deduplication services

Source: Philippe Rixhon Associates and County Analytics

The pre-screening would allow deduplication against multiple data sources. It could be used to diffuse identities created at Performance Rights Organisations (PROs, a synonym for Mechanical Licensing Collective / MLCs, and a subset of Collective Management Organisations / CMOs), e.g., International Standard Recording Codes (ISRCs) or local codes.

5.8.3. Addressing inefficient markets through micro-licensing

Automated rule-based micro-licensing could help press publishers exploit their newly granted neighbouring right. Therefore, one would need expert systems. A knowledge engineer would interview a human expert and – assisted by machine learning – would build a knowledge base, containing copyright regulations and business practices of a specific creative sector. This knowledge base would be interpreted by an inference engine, that would communicate with a non-expert user through an appropriate user interface. This would be a human-to-human system.

Once tested and stable, this expert system would be compiled into an automated rule-based micro-licensing engine. That would be a system-to-system tool. On one side, a system would input the selected image, its metadata, and facts and context around the required licensing. On the other side, the micro-licensing engine would produce a machine-and-human-readable micro-licence. This licensing process could be simple, accurate, fast, transparent, and affordable. At least, if we would have the necessary metadata.

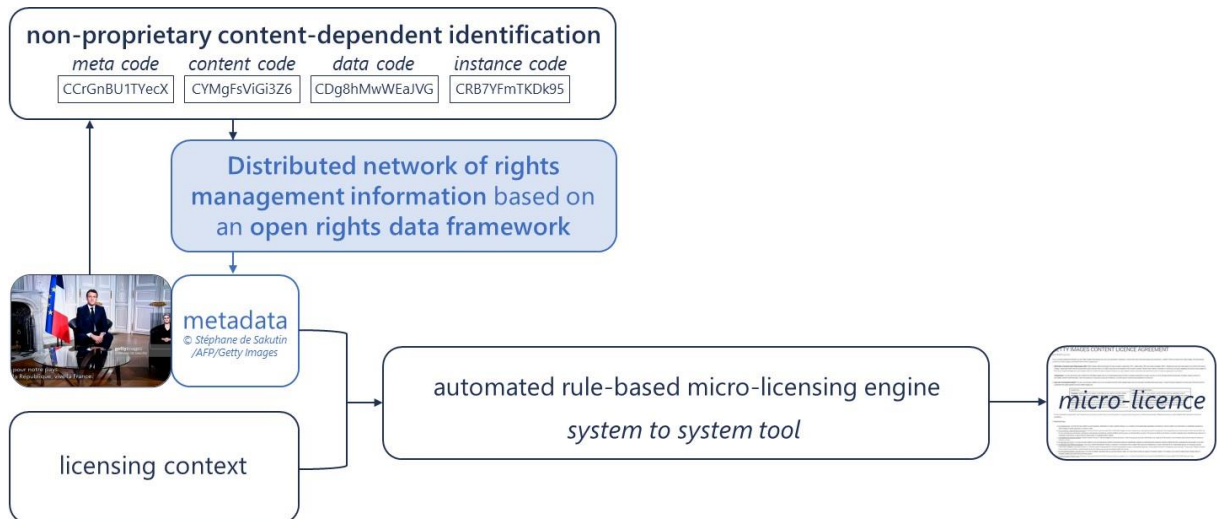


Figure 81: Retrieving metadata through open content-dependent identification

Source: Philippe Rixhon Associates

If metadata would not be attached to the image, one could analyse the *DNA* of the image, for example by using a non-proprietary content-based identification tool. It would produce a series of cryptographic hashes – from abstract and persistent characteristics on the left to concrete and volatile attributes on the right.

With that *DNA*, one could then query the distributed network to retrieve verified rights metadata concerning the image, relying typically on an identifier binding protocol and verifiable credentials. Then, one could reconcile image and metadata, and finally, fire the micro-licensing engine.

Traditional media companies estimate that they have lost billions of Euros in advertising revenues to the online platforms. One would need only a fraction of that to build the systems outlined here and enable flows of content and counterflows of remuneration based on reliable, exhaustive and current metadata.

5.8.4. Addressing misappropriation of film and television programmes

The combination of digital watermarks and non-proprietary content-dependent identification could help protect film content and TV channels against misappropriation.

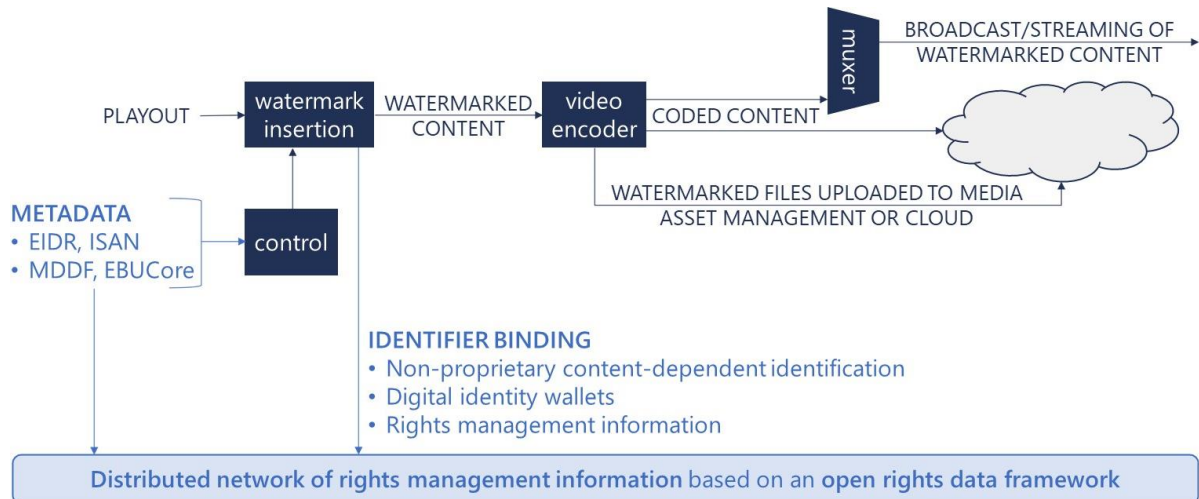


Figure 82: Marking the TV programme

Source: Philippe Rixhon Associates, EBU and Mediaset

One could continue with more opportunities arising from the existence of an open rights data framework. Again, the avenue is here is to open the existing framework, not to define a new one. Opening the framework requires to strengthen it to make it more trustworthy, and more independent.

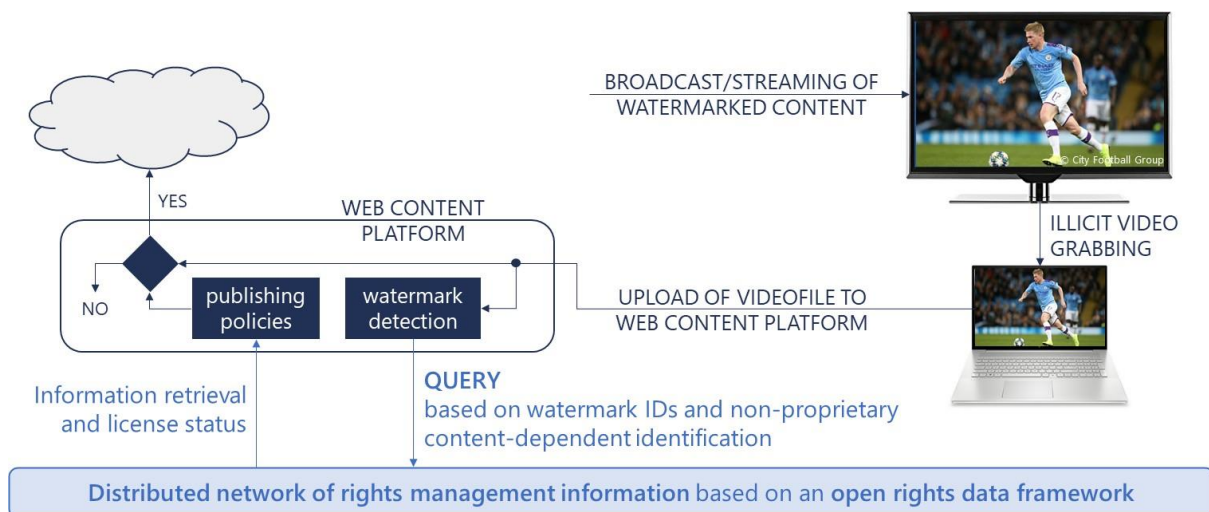


Figure 83: Enforcing rights

Source: Philippe Rixhon Associates, EBU and Mediaset

During the study, the team had the opportunity to discuss how such an open rights data framework could dramatically enhance the current watermarking systems used to detect illicit video grabbing and enforce the rights of commercial broadcasters. Large online platforms and a few commercial service providers are already building solutions. As rights data management must be neutral to business models, an open framework would create a space for many solutions. It could benefit all stakeholders in the content ecosystem, also the online platforms and the commercial service providers.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696 or
- by email via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications at: <https://publications.europa.eu/en/publications>. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: <http://eur-lex.europa.eu>

Open data from the EU

The EU Open Data Portal (<http://data.europa.eu/euodp/en>) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

