Monitoring the Evolution and Benefits of Responsible Research and Innovation (MoRRI)

Analytical report on the dimension of open access

Ingeborg Meijer, Xavier Potau, Clifford C Tatum, Rodrigo Costas Comesana, Bea Mahieu

Sub-task 2.5, analytical report, Deliverable D2.4

Executive Summary

In this report, Open Access will be reviewed from three separate angles:

- The general concept of open science from a policy perspective
- The Open Access instrument for publications
- Developments in Open data

The general concept of open science from a policy perspective: In the past years, the Commission repeatedly noted that there was a limited uptake of Open Access. A lack of policy coordination and/or framework conditions impeded the free movement of research activities and knowledge, hindering access to publicly funded research results and knowledge transfer. The Open Access pilot initiative within the mainstream Framework Programme itself (in addition to the ERC) was launched by the EC in 2008. Since then, a clause on Open Access was present in FP7 grant agreements for areas participating in the pilot requiring that Beneficiaries *shall deposit an electronic copy of the published version or the final manuscript accepted for publication of a scientific publication relating to foreground published before or after the final report in an institutional or subject-based repository at the moment of publication.*

The policy movement to enhance Open Access to research outputs led to the launch of a series of projects, aimed at fostering and enabling the Open Access of research outputs in Europe. One of them is the OpenAIRE repository that contributed to the uptake of OA at European level. Progress towards OA in individual member states has also reached a tipping point with 50% of the papers published in 2011 and 40% of those published worldwide between 2004 and 2011 available in some form of open access. The results from the latest OA study commissioned by DG RTD confirm the momentum both in OA publishing and in the establishment of OA policies and repositories in European Member States.

The Open Access instrument for publications: Assessment of Open Access publishing is complicated by the growing diversity of what counts as Open Access, the copyright restrictions for when a publication can be made openly accessible, and the lack of clear and consistent identification of Open Access publications in bibliographic data. There are two formal operational paths to access through Open Access journals and self-archiving in repositories, subsequently referred to as Gold Open Access and Green Open Access.

- The golden OA route/road: When a paper is published through this route, it is immediately available to anyone without the need for a subscription or viewing fee. For fully open access journals, the business model of the journal is usually based on a) Processing charges on the side of the authors (APC); b) Direct contributions from the journal owners, more common in case of journals owned by funders, institutions such as Universities, or professional bodies/associations; and c) A combination of both.
- The green OA route/road (also called 'self-archiving'): In the green road of OA publishing, authors publish in traditional subscription-based journals and an additional copy of the paper (usually called the post-print, a final peer-reviewed copy without the format and branding of the journal) is stored in an open institutional or subject-based repository. An embargo period is established, usually of between 6-12 months, in which the publication is only available through subscription to the publisher. After the embargo period is over, the general public can access the publication from the open repository.

- Hybrid Open Access generally refers to the situation whereby authors can pay to make their articles in subscription journals openly accessible on the Web. In Hybrid journals it is up to the authors to decide whether to pay APC so that their publication is freely available from the publisher's website. In a hybrid journal, the traditional method of access through subscription coexists with individual open access publications.

The important and on-going debate in the assessment of OA is whether OA publishing yields increased citation impact. While there are conflicting reports in the debate about an Open Access citation advantage (OACA), heightened attention to this issue has increased our understanding about Open Access publishing more generally.

Developments in Open data: The other big challenge for OA is re-uniting publications with public funding and to interlink publications with research data. Such improved interlinking would allow for the investigation of research results, beyond the limits of project-based funding, and provide the data needed for the exploration of longer-term results of public research funding. Data and datasets are central for empirically oriented science and scholarship. They can be very diverse (e.g. archaeological, biological, genetic, economic, mathematical, astronomic, etc.) and once collected, the same data can be used by a variety of researchers from different institutes, disciplines and organisations to produce new results. Open data and data sharing have proven to create many benefits, such as stronger open science; a higher efficiency in the use (and reuse) of scientific resources; the possibilities of expanding new research lines as well as other users; the acceleration of the scientific progress; better science through the possibilities of verifying, refuting or refining scientific results; improvements of measurement and data collection methods; protection against faulty data, etc. Despite these main benefits, there are cultural and technical issues that hamper widespread data curation, sharing and use.

- Cultural aspects: Perceptions and cultural issues must develop not only to data sharing activities but also towards the establishment of cultures of acknowledging the (re)use of data, for example through citations. However, scholars perceive important barriers regarding data sharing such as the lack of recognition and reward for data curation and data sharing, the lack of time and lack of funding for data sharing activities; the feeling of 'losing' control over the data by the creators of the data; the possible misuse of the data by others; the exposure of potential errors, as well as the potential lack of acknowledgement to the creators of the data.
- Technical and policy aspects: Technical developments necessary include the development of standards and platforms for proper data sharing (e.g. software systems, data storage, data management, data compatibility, metadata standards, data identification, persistent identifiers, or granularity and versioning of datasets, but also the development of policies of tracking and evaluating data sharing contributions, as well as considering them during hiring, tenure and promotion decisions, as the considering of parts of the research budgets are attributed to data sharing.

The above is the result of an extensive literature review on Open access as presented in chapter 2, and the relevant vocabulary as summarized in chapter 3.

Chapter 4 presents an overview of EC funded projects and other activities on Open Access and Open Data, some of which are still ongoing. Taking these studies as a source to examine the suitability and availability of data and indicators of the benefit of Open Access to RRI, is discussed in detail in chapter 5 and 6.
In general, the studies that are reviewed are explorative in nature; the focus is on perceptions of scientists, attitudes and cultural behaviour rather than on actual data.

From a data-perspective, the studies analyse feasibility, coverage, metadata and implementation issues of Open Access and Open Data. Data availability from repositories may not be the main issue since for publications bibliometric methods are available, and consolidated data sets such as the Web of Science from Thomson Reuters or Scopus from Elsevier are systematically collected. Regarding Open Data, the actual data availability is relatively scarce even though huge repositories are available in some scientific fields. Open Data is also a relatively young area of interest, and it is beneficial that a worldwide effort has started to increase, harmonise, and integrate Open data activities for the benefit of science.

Gaps in data availability on Open Access are mainly related to:

- Lack of open access functionality in databases such as Scopus
- Difficult and labour intensive retrieval of publications from Google scholar with little options for aggregate analysis
- Technical issues, and other theoretical problems regarding the OA transition phase not yet solved by bibliometricians.

Gaps in data availability on Open Data are mainly related to:

- *Structural scarcity of data*: basically the lack of incentives to share the data as well as the lack of culture in acknowledging (citing) datasets implies that the data sharing events are low and therefore posing a challenge for the development of indicators based on them.
- *Diversity and dispersion of data repositories and data venues*. There are multiple and diverse data repositories, that vary from disciplinary to institutional, etc.

In chapter 7, twelve possible indicators on open access and open data are presented, which focus primarily on context, input and output.

Table of Contents

Executive Summary 2
Table of Contents 5
1. Introduction - analytical and empirical aspects of Responsible Research and Innovation 6
2. Results of the literature review on open access 7
2.1 Review of core literature 7
2.2 Synthesis of literature review on Open Access 9
3. Functional vocabulary of open access – definitions and terminology 19
4. Review of existing empirical knowledge of open access 23
4.1 EC studies and projects in the area of open access and open data 23
4.2 Other empirical studies on the dimension of open access 29
5. Assessment of data availability on open access 36
5.1 Data availability across open access categories 36
5.2 Availability of quantitative and qualitative data 40
5.3 Availability of data across analytical levels included in the intervention logic model 41
5.4 Availability of data at different levels of aggregation 42
6. Data selection for RRI monitoring – reflections of current data gaps and required data collection 44
7. Early thoughts on open access indicators 47
8. Ideas bank 64
9. References 65
10. Appendix – literature review 67
10.1 Review guidelines 67
10.2 Review reports 68

1. Introduction - analytical and empirical aspects of Responsible Research and Innovation

This report is one out of a series of six reports, each targeting a separate dimension of Responsible Research and Innovation (RRI). The six dimensions include 'Citizen engagement and participation of societal actors in research and innovation', 'science literacy and scientific education', gender equality', 'open access to scientific knowledge, research results, and data', 'research and innovation governance' and research and innovation ethics'. The six reports collectively form the main output of Task 2 of the 'Monitoring the Evolution and Benefits of Responsible Research and Innovation' (MoRRI) project, and they are informed by the results of the literature review on RRI and its conceptual components which was performed as Task 1 of the project.

The six reports emerging from Task 2 specifically address analytical and empirical issues relating to each of the RRI dimensions. Each report aims to:

- Provide an operational understanding of the RRI dimension it targets
- Present existing empirical information about the RRI dimension
- Assess data availability and specify analytical levels and degrees of aggregation of available material

The reports will provide a platform for subsequent definition of metrics and indicators for the RRI dimensions in Task 3. The report at hand specifically focuses on the dimension of 'Open Access'.

The report is structured in accordance with the main aims of Task 2 and also provides an outlet for the results of Task 1. In chapter 2, results from the literature review are presented. These provide a background for the following chapters. Chapter 3 is concerned with the development of an operational understanding of open access. The objective is to provide a functional vocabulary of open access by clarifying important analytical components and definitions of open access. This chapter includes specification of the relationship and borderlines between the open access dimension and the other five dimensions of RRI. Chapter 4 accounts for existing empirical information on open access. It is based on a review of selected studies funded by the European Commission, along with review of evidence from other empirically oriented studies, which are considered particularly relevant for the open access dimension.

In chapter 5, availability of existing data on open access is assessed. Following the scheme outlined in the MoRRI proposal, this chapter specifically considers the availability of data on open access relating first to its characteristics in terms of the intervention logic model, i.e. data describing the context, input, output, and outcome of ethics. More specifically, **context** relates to the environment and overall situation in a country; **input** to the activities carried out, measures taken, structures created or resources provided to address what is done in order to address issues of RRI and whether it is done in a systematic manner; **outputs** to the immediate or direct results of activities and **outcomes** relate to the achievements (MoRRI Proposal 2014:64). Second, availability of data are described according to the level of aggregation of these data, distinguishing data that describe the global level, the national level, the regional level, the institutional level, the programme/project level and the individual level.

Reflecting the findings in chapter 5, chapter 6 considers issues relating to data gaps and assesses the overall need for primary data collection to fill gaps. Finally chapter 7 provides early thoughts on the development of indicators and metrics for open access, which will be the objective of Task 3.

2. Results of the literature review on open access

This chapter includes a list of the core literature on open access selected for review (approximately 10-15 papers have been reviewed for each RRI dimension), as well as a synthesis of the literature review on this dimension. The literature review was performed in Task 1 of this project. The synthesis summarises the main conceptual elements of the targeted dimension, and forms the background for the succeeding chapter about the 'functional vocabulary' for the dimension.

2.1 Review of core literature

The objectives of the literature review (Task 1) is to:

- Review the state of knowledge regarding RRI
- Define the policy context of RRI in Europe and elsewhere
- Give a comparative assessment of RRI dimensions, weighing-up advantages, disadvantages and available options
- Conduct a preliminary assessment of the availability of empirical evidence on the dimensions
- Finalise the definitions and properties of the RRI key dimensions
- Finalise the definition and properties of additional factors that may be relevant for the monitoring tasks.

In order to meet these objectives and provide useful input to the thematically and methodologically strongly related aims of Task 2 and other ensuing project tasks, the approach to the literature review was designed in close cooperation with the dimension and task leaders. In a first step, the five dimension leaders were asked – based on their long-standing experience in their respective fields – to select 10 to 15 key publications in each key RRI-dimension for detailed review. Second, a review template was designed in order a) to ensure a systematic analysis of the selected literature and b) to cover all relevant aspects and information required in Tasks 1 and 2. Before it was rolled out to the individual reviewers, the template was subject to a pre-test.

In the next chapters, Tasks 1 and 2 are completed for the Open Access dimension. In this context, Open Access is reviewed from three separate angles:

- The general concept of open science from a policy perspective
- The Open Access instrument for publications
- Developments in Open data

For Open access, the following key publications were selected and reviewed (see review templates in chapter 8. A list of additional relevant publications can be found in chapter 8 as well.

Policy documents and reports

- European Commission. Towards better access to scientific information: Boosting the benefits of public investments in research. COM(2012) 401 final.
- European Commission. 2014 Communication from the Commission. Towards a thriving data-driven economy. COM(2014) 442 final
- The Finch Report (2012): Accessibility, sustainability, excellence: how to expand access to research publications.

- Tim Davies, 2013, Open Data – Barometer, 2013 Global Report, World Wide Web Foundation and Open Data Institute.
- Van den Eynden, V. and Bishop, L. (2014). Sowing the Seed: Incentives and motivations for sharing research data, a researcher's perspective. A Knowledge Exchange Report.
- Genova, F. et al. (2014) The Data Harvest. How sharing research data can yield knowledge, jobs and growth. A Special Report by RDA Europe.

The Open Access publication model

- Amyot, D., Deschamps, P., Nicol, A., Rebout, L., & Roberge, G. (2014). Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels — 1996 – 2013 (p. 54). Montreal.
- Archambault, Eric, Didier Amyot, Philippe Deschamps, Aurore Nicol, Françoise Provencher, Lise Rebout, and Guillaume Roberge. 2014. "Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels—1996–2013." *Rapport, Commission Européenne DG Recherche & Innovation; RTD-B6-PP-2011-2: Study to Develop a Set of Indicators to Measure Open Access*
- Björk, Bo-Christer, Mikael Laakso, Patrik Welling, and Patrik Paetau. 2014. "Anatomy of Green Open Access." *Journal of the Association for Information Science and Technology* 65 (2): 237–50. doi:10.1002/asi.22963.
- Carpenter, Todd. 2013. "Progress Toward Open Access Metadata." *Serials Review* 39 (1): 1–2. doi:10.1016/j.serrev.2013.02.001.
- Craig, Iain D., Andrew M. Plume, Marie E. McVeigh, James Pringle, and Mayur Amin. 2007. "Do Open Access Articles Have Greater Citation Impact?: A Critical Review of the Literature." *Journal of Informetrics*, The Hirsch Index, 1 (3): 239–48. doi:10.1016/j.joi.2007.04.001.
- Laakso, Mikael, and Bo-Christer Björk. 2013. "Delayed Open Access: An Overlooked High-impact Category of Openly Available Scientific Literature." *Journal of the American Society for Information Science and Technology* 64 (7): 1323–29. doi:10.1002/asi.22856.
- Swan, Alma. 2010. "The Open Access Citation Advantage: Studies and Results to Date". Technical Report. http://eprints.ecs.soton.ac.uk/18516/.

Developments in Open data

- Costas, R., Meijer, I., Zahedi, Z., & Wouters, P. (2013). The value of research data - Metrics for datasets from a cultural and technical point of view. A Knowledge Exchange Report (pp. 1–48).
- Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 368(1926), 4023–38. doi:10.1098/rsta.2010.0165
- Dallmeier-Tiessen, S., Darby, R., Gitmans, K., Lambert, S., Suhonen, J., Wilson, M., … Coordination, A. (2012). Compilation of results on drivers and barriers and new opportunities.
- Farhan, H., Alonso, J., Davies, T., Tennison, J., Heath, T., & Berners-lee, T. (2013). Open Data Barometer, 1–45.
- Fienberg, S. E., Martin, M. E., & Straf, M. L. (1985). Sharing Research Data. Washington: National Academy Press.

- Lemke, A. A., Wolf, W. A., Hebert-Beirne, J., & Smith, M. E. (2010). Public and biobank participant attitudes toward genetic research participation and data sharing. Public health genomics, 13(6), 368–77. doi:10.1159/000276767
- Piwowar, H. A., Becich, M. J., Bilofsky, H., & Crowley, R. S. (2008). Towards a data sharing culture: recommendations for leadership from academic health centers. PLoS medicine, 5(9), e183. doi:10.1371/journal.pmed.0050183

The guidelines for the review process and the findings of the individual reviews are documented in the Appendix to this report. These three different areas or angles of Open Access have different types of documents as main sources of evidence. For example, information on the general concept of open science from a policy perspective is mostly found in study reports published or commissioned by Governments and official Government documentation, while information on the Open Access publication model and developments in Open data can be found through a combination of scientific peer-reviewed publications and other grey literature.

2.2 Synthesis of literature review on Open Access

The synthesis of the reviewed literature has been conducted in order to provide a concise overview of the key dimension, its policy context, main definitional elements and functional vocabulary, most important claims about impacts, and relationships to other key dimensions of RRI.

Historically, open science relates to the need to build a publicly recognised reputation. The scholarly tradition of open knowledge was turned into a procedure for establishing knowledge claims that could be evaluated and recognised by peers and then utilised by others. Knowledge was considered a public good, and likewise a publication (any kind) as well. Since then, propertisation of knowledge occurred through copyright imposed by the academic publishing market, which may not be congruent with competition law. And now policy aims to return to the general concept of open science: to "Enhance open circulation of knowledge across national borders, including knowledge transfer".

Policy in the European Context

The EU is currently the world's largest producer of publicly funded scientific knowledge, measured by publications. In addition to the conventional wisdom narratives on the 'European Paradox' (the supposed lack of European effectiveness in converting new knowledge into socio-economic benefits), the conversion of knowledge is dependant on its quality and availability, and sufficient absorptive capacity in the receiving end. In the past years, the Commission repeatedly noted that there was a limited uptake of Open Access. A lack of policy coordination and/or framework conditions impeded the free movement of research activities and knowledge, hindering access to publicly funded research results and knowledge transfer. A diversity of national policies, legal requirements and practices regarding knowledge transfer as well as open access to scientific publications and scientific data adversely affected the wider dissemination, access to and use of knowledge created with public funds[1].

In addition to this lack of homogeneity in policy and practices, historically open access has also suffered from an information asymmetry problem between researchers, funders and publishers. Initially, researchers were wary of the quality of works published in open access and were not sufficiently aware that open access is not necessarily in conflict with publishers' copyright provisions. Publishers had to adapt their business models, pricing and sales policies as well as redefine their views on

[1] Study in support of an ex-ante impact assessment of post 2010 ERA policies. Final Report to the European Commission DG-Research. Technopolis Group.

copyright issues. Finally, the main barriers for funders were to overcome the lack of awareness regarding the opportunities of open access.

The first notable international calls for support to the open access movement happened in the 2001-2003 period. During this period, three major international declarations were put forward by the international scientific community: the Budapest initiative in February 2002; the Bethesda Statement on Open Access Publishing in June 2003 and the **Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities in 2003**[2]. The Berlin Declaration had a particular impact in Europe, as it was backed by hundreds of European researchers and institutions.

As a result, the situation and international calls of the community for the adoption of Open Access prompted a policy response, both from the European Commission as well as at the national level in the main Western European knowledge economies. For the past nearly two decades, open access concepts have been introduced progressively into Europe's overarching research guidelines and as of late it has been elevated to policies such as the Innovation Union, the Digital Agenda for Europe and the European Research Area.

In December 2006 the ERC published a **Scientific Council Statement on Open Access**, which stressed the importance of the dissemination of high-quality scientific results. More Communications followed in 2007[3] and the analysis of the situation concluded that there was a need for policies mandating the public availability of funded research outcomes. In 2007 the **ERC followed up with guidelines** requiring that all peer-reviewed publications from ERC funded research projects were deposited into an appropriate repository with an open access within 6 months of publication. The ERC also urged the need to make primary data available.

In 2010, the Commission adopted the **Europe 2020** Flagship Initiatives **Innovation Union**[4] and the **Digital Agenda for Europe**[5]. Both Communications make reference to **Open Access (OA)** as a means to work towards achieving the Europe 2020 objectives. They underline the importance of promoting better access to the results of publicly funded research, and introduce open access as the general principle for projects funded by EU research framework programmes. In addition, Open Access is today part of the **European Research Area strategy**[6]. One of the main objectives of the ERA is to enhance knowledge circulation across Europe and beyond. The first action under such objective is to *Enhance open circulation of knowledge across national borders and open access to research outputs (publications and data) by researchers and society at large*[7,8].

In the framework programme, Open Access was initially supported through specific sets of projects within **FP6 Science and Society (SaS)** and **FP7 Science in Society (SiS)**. The main aim of these programmes was to raise the political importance of

[2] Giancarlo F. Frosio. Open Access Publishing: A Literature Review. CREATe Working Paper 2014/1. 2014

[3] 2007 Communication on scientific information in the digital age: access, dissemination and preservation (COM(2007)56) and 2007 Commission Communication on setting out approaches for a common European framework for knowledge transfer COM(2007) 182 final

[4] COM(2010) 546 final

[5] COM(2010) 245 final/2

[6] COM(2012) 392 final

[7] European Commission (2010b), Access to scientific information in the digital age, ERA Core Group – Concrete and implementable objectives, Final 19.5.10

[8] Rogers, Steve (2009), Building the European Research Area, Briefing for Máire Geoghegan-Quinn, Commissioner-Deesignate for Research, Innovation and Science.

science in society at the European level. However, the Open Access idea did not feature as prominently in FP6. Instead, the initial SiS programme focused on the dissemination of scientific information to the general public, in line with the first Science and Society action plan of 2002. This resulted in a push towards communication of scientific activities through public information channels and capacity building for scientific journalists. Dissemination, use and access rights of scientific publications were not explicitly addressed until the beginning of FP7, when the main challenges faced by researchers were identified[9]. The main objectives of the FP7 SiS programme, as identified in the FP7 SiS interim evaluation were the following:

- Increase the availability of research data and publications of publicly funded research in Europe via Open Access
- Strengthen the use of creative common licenses or similar mechanisms
- Increase awareness of the importance of Open Access among researchers
- Support the access of all existing decentralised repositories via a unified interoperable European search portal

The SiS interim evaluation already indicated that the programme was having major success in getting open access principles piloted in FP7 and that this would facilitate embedding these principles into **Horizon 2020**[10]. Additionally, SiS funding, although insufficient to effect a systemic change, was decisive to support unpartisan studies that brought together stakeholders with very different (and sometimes confronted) positions, such as publishers, libraries, research councils and research organisations.

The actual **Open Access pilot initiative** within the mainstream Framework Programme itself (in addition to the ERC) was launched by the EC in 2008. The OA pilot required peer-reviewed research in the pilot areas and resulting from research funded by the FP7 programme to have an Open Access. Since August 2008, a clause on Open Access was present on FP7 grant agreements in the areas that took part in the OA pilot. These areas were energy, environment, health, information and communication technologies, research infrastructure, science in society and socio-economic sciences and humanities. This clause was referred to as the Special Clause 39:

> *Beneficiaries shall deposit an electronic copy of the published version or the final manuscript accepted for publication of a scientific publication relating to foreground published before or after the final report in an institutional or subject-based repository at the moment of publication.*

The agreements usually conceded an embargo period, i.e. a lag of 6 or 12 months (depending on the FP7 area) between publication of the journal and the moment from which authors were required to provide Open Access to the research article. The embargo period allows scientific publishers to ensure a profit on their investment (by charging for journal subscription), while assuring Open Access to funded research articles once the embargo period has lapsed. The EC pilot exercise follows the *green road model* of OA, where grant recipients are expected to deposit peer-reviewed

[9] Archibugi D., Ampollini I., Basili C. The Contribution of Science and Society (FP6) and Science in Society (FP7) to a Responsible Research and Innovation. A Review.
[10] Technopolis and Fraunhofer (2012). Interim evaluation & assessment of future options for Science in Society Actions. Final Report.

research articles or final manuscripts resulting from their projects into an online repository[11].

The policy movement to enhance Open Access to research outputs led to the launch of a series of projects, aimed at fostering and enabling the Open Access of research outputs in Europe (see section 3.2 for a list of such studies). One of them, funded under the EC Research Infrastructures programme, is the **OpenAIRE** pilot action. OpenAIRE aimed to support the implementation of OA by providing the needed infrastructure of an EC-supported OA repository, so that researchers in institutions without open repositories could fulfil the OA mandate and **external repositories** could also be linked and their information harvested.

By guaranteeing Open Access to funded research, both ERC and the European Commission intended to improve the dissemination of high-quality scientific results. The aim is to increase the EU's return on research and development investment and increase the efficiency of scientific discovery.

With the introduction of the open access pilot in FP7, the status of Open Access was effectively elevated to that of a European Policy Measure itself, in addition to being one of the research topics supported by the framework programme. The Open Access pilot ran until the end of FP7 and covered approximately 20 % of the FP7 research budget.

Going forward, the Commission proposed to make open access a general principle and all beneficiaries are required to deposit and ensure open access to scientific publications. Additionally, Open Access is now an integral component of the **Responsible Research and Innovation** concept[12]. In order to be responsible, research and innovation must be transparent and accessible. The latest report commissioned by DG RTD on open access validates the literature on the societal benefits of OA with tangible data on its uptake by researchers, publishers and funders, concluding:

> *Greater societal benefits may result from the fact that OA reduces the digital divide, increases transparency and accountability, levels disparities and facilitates participation and results in better informed citizens (Davis, 2009; Herb, 2010; ICTP, 2008; OASIS, nda).*[13]

The Open Access publication model

Assessment of Open Access publishing is complicated by the growing diversity of what counts as Open Access, the copyright restrictions for when a publication can be made openly accessible, and the lack of clear and consistent identification of Open Access publications in bibliographic data. A recently published metadata standard for Open Access holds some promise for improving both human and machine identification of Open Access publications (Carpenter 2013). Here too, stakeholders involved in the new standard were unable to agree on a precise definition of Open Access. Instead, the standard specifies metadata elements for *free to read* and *license reference*, the latter of which should point to copyright information publically accessible on the Web (NISO 2015).

[11] In addition to this pilot, FP7 rules of participation also allow all projects to have open access fees eligible for reimbursement during the time of the grant agreement (*golden road*).

[12] Options for Strengthening Responsible Research and Innovation. Report of the Expert Group on the State of Art in Europe on Responsible Research and Innovation. DG RTD, 2013.

[13] Julie Caruso, Aurore Nicol & Eric Archambault. Open Access Strategies in the European Research Area. 2013. Science Metrix study for DG Research and Innovation

The Budapest Open Access Initiative (BOAI) established two operational paths to access through Open Access journals and self-archiving in repositories, subsequently referred to as Gold Open Access and Green Open Access (Bailey, 2005). Hybrid Open Access generally refers to the situation whereby authors can pay to make their articles in subscription journals openly accessible on the Web (Björk 2012). A more in-depth discussion of the different models of OA can be found in section 3 of this report.

An important and on-going debate in the assessment of OA is that OA publishing yields increased citation impact. While there are conflicting reports in the debate about an Open Access citation advantage (OACA), heightened attention to this issue has increased our understanding about Open Access publishing more generally. Numerous bibliometric studies claim that Open Access publishing results in a significant increase in citations. In these studies the size of advantage varies widely based on a variety of issues, such as disciplinary differences, methodological approaches, variation in how Open Access is defined, and difficulty in determining when an article is made openly accessible (Swan 2010). Craig et al. (2007) are critical of the early OACA analyses on the basis of methodological factors, arguing that advanced bibliometric techniques were not being used. They credit Eysenbach (2006) with introducing the importance of "author characteristics (reputation, prior citation history, lifetime publication count, country, funding organization, etc.) as confounding variables" in OACA analysis. Craig et al. (2007) also credit Moed (2007) for drawing attention to the standard bibliometric practice of using of fixed time windows when calculating citations for each article (p.246). In addition to methodological concerns, other confounding factors have been shown to influence citation frequency, such as: early exposure to draft versions of a manuscript (Moed 2007), self-selection bias, whereby an author may choose Open Access for only her best publications (Kurtz M. et al. 2007), the availability at multiple access points (Xia et. al 2010), and physical proximity of researchers (Lee et al. 2010).

A recent European study (Archambault et al. 2014) sheds some additional light on the OACA situation. The authors claim "OA papers were between 26% and 64% more cited on average for any given year than all papers combined, whereas non-OA received between 17% and 33% fewer citations (based on a sample size of at least 10,000 papers any given year) (p. iii). However, of particular interest is the variation of citation behaviour found among different OA types. For example, they find that the "citation advantage is derived almost exclusively from the Green and Other OA portion, as Gold OA is associated with a citation disadvantage on average for all fields except for physics & astronomy (p. 20). This highlights the point that Gold OA are still journals relatively young on average and thus need additional time to build a reputation, whereas Green OA is potentially more representative of all publishing channels.

Open Data and Public Sector Information policy

The concept of Open Data is also being applied to **Public Sector Information (PSI)**, that is, data generated by Government and other public sector bodies (see vocabulary in Section 3 of this report). In 2009, Tim Berners-Lee, one of the inventors of the World Wide Web and an initiator of Linked Data, suggested a 5 star framework with which to qualify the re-usability of Open Data and Public Sector Information[14]:

- 0 – Data not available under an open licence, even if it is available on-line
- ⋆ - Available on the web (whatever format) but with an open licence, readable by the human eye

[14] See http://5stardata.info

- ★★ - Available as machine-readable structured data (e.g. excel instead of image scan of a table)
- ★★★ - as (2) plus non-proprietary format (e.g. CSV instead of excel)
- ★★★★ - Data in the web, as opposed to on the web. Use of open standards from W3C (RDF and SPARQL) to identify things, so that people can point directly at specific data points
- ★★★★★ - All the above, plus: Data is linked to other people's data in order to provide context

This framework has increasingly become the standard for public sector bodies to assess their degree of 'openness' with public sector information. Principles for the storage of Open Government Data were also put forward in 2007. Under these, data is considered 'open' if it meets the following 8 criteria[15]: Complete, primary, timely, accessible, 'machine-readable', non-discriminatory, non-proprietary, and license-free. Finally, the community of research infrastructures dealing with large datasets is behind the idea of making scientific data FAIR from the point of view of 'Findability, Accessibility, Interoperability, and Reusability'[16].

There are several notable policy milestones related to Open Data and Public Sector Information. The EU has stressed the goal of opening PSI as a resource for innovation and for addressing societal challenges for many years. Access to documents held by the three major institutions (Parliament, Council and Commission) was already regulated in Regulation no. 1049/2001. In 2004, OECD Science Ministers signed a Declaration on Access to Research Data from Public Funding[17], stating that all publicly funded archive data should be made publicly available to enhance scientific progress. In 2007, the OECD published its Principles and Guidelines for Access to Research Data from Public Funding[18] as a soft-law recommendation.

In December 2011, the EC presented a comprehensive Open Data package, revamping and broadening its support for Open Data and PSI. The next big challenge for OA is re-uniting publications with public funding and to interlink publications with research data. Such improved interlinking would allow for the investigation of research results, beyond the limits of project-based funding, and provide the data needed for the exploration of longer-term results of public research funding.

In December 2012, the EU Open Data Portal was launched, providing data held by the EC and other public bodies, and encouraging its re-use and accessibility in machine-readable formats. In June 2013, the EU endorsed the G8 Open Data Charter[19], committing to the implementation of the activities in the G8 Collective Action Plan for Open Data. Also in late 2013, the Global Open data Initiative declaration[20] fostered by the community and presented its own set of principles, with a focus on the bottom-up and emphasising accountability issues.

---

[15] See http://opengovdata.org

[16] See http://datafairport.org/

[17] Declaration on Access to Research Data from Public Funding. 30 January 2004 - C(2004)31/REV1. Available at: http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=157

[18] OECD Principles and Guidelines for Access to Research Data from Public Funding. Available at: http://www.oecd.org/sti/sci-tech/38500813.pdf

[19] G8 Open Data Charter and Technical Annex. Available at: https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex

[20] See http://globalopendatainitiative.org/declaration/

Table 1 G8 Open Data Charter and Global Open Data Initiative declaration principles

G8 Open Data Charter principles	Global Open data Initiative declaration principles
Release open data by default Ensure high quality and quantity of data Make data usable by all Release data for improved governance Release data for innovation	Make data open by default Make the process people-centered (or "put the users first") Provide no-cost access Put accountability at the core Improve the quality of official data Enact legal and political reforms to create more open, transparent and participatory governance

In **Horizon 2020** the Commission has launched a **pilot on open access for scientific data** collected during the course of some of the projects and taking into account legitimate concerns related to the grantee's commercial interests, privacy and security. Although at present time depositing research data is not compulsory, it is strongly encouraged by the Commission. This pilot will cover to around €3b, or 20% of the overall Horizon 2020 budget in 2014 and 2015. One of the areas that will participate in the pilot is Science with and for Society[21].

Within the EC, DG CONNECT[22] takes responsibility for developing PSI policy and its work has evolved into a multifaceted initiative that covers all aspects of the re-use of PSI and research data, with the following overarching objectives:

- The creation of data value chain friendly policy and legal environment
- Building of a Multilingual (Open) Data infrastructure
- Supporting Research and Innovation

The **PSI Directive** is now being transposed into legislation by different Member States, a process that will be concluded by July 2015.

Although still a new concept, the take-up of Open Data initiatives and the implementation of PSI policies in European countries is being carried out by platforms and scoreboards such as: the European PSI platform[23], run by DG Connect; and the Open Data Barometer[24], run by the Open Data Institute and the World Wide Web Foundation. Open Data and PSI are, at present time, fast-moving fields. New reports, initiatives and working groups are emerging, running consultations at different levels, and publishing findings. Permanent working groups for several aspects of Open Data and PSI are established for example in the Research Data Alliance[25].

[21] Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. Version 16 December 2013
[22] Communication on Open Data (COM(2011)882); A revision of the Decision governing the re-use of Commission's own information (2011/833/EU); Revision of the Directive on the re-use of public sector information (2013/37/EU)
[23] http://www.epsiplatform.eu
[24] http://opendatabarometer.org
[25] See https://rd-alliance.org/groups

The Open Data and Data Sharing dimensions

Data and datasets are central for empirically oriented science and scholarship. Thus data can be very diverse (e.g. archaeological, biological, genetic, economic, mathematical, astronomic, etc.) and once collected, the same data can be used by a variety of researchers from different institutes, disciplines and actors to produce new results. As a result, open data and data sharing are currently considered to be at the core of good and efficient science and scholarship, which has opened a debate about the need of openly shared data that are the result of research, particularly when this research is funded with public support. The increasing awareness about data sharing in the scientific community is reflected in the profusion of reports and publications discussing the issues surrounding data sharing (Costas, Meijer, Zahedi, & Wouters, 2013).

Data sharing and Open data are topics whose relevance goes back to the 80's with the 1985 report by Fienberg, Martin, & Straf (1985) who already pointed out some of the benefits, problems, controversies, and main challenges related with sharing research data. Some of the conclusions of this report are still relevant, such as the need for developing guidelines on data sharing, the need of involving multiple stakeholders in discussions about the need of policies and standards for accessing, classifying, documenting and archiving data. More recently, other comprehensive documents dealing with the problems related to data sharing and data citation have been published (Costas et al., 2013). In this report an extensive study on the state of the art of data sharing and data metrics was presented (in addition to interviews with stakeholders and technical analysis of existing data repositories) and one of the main outcomes was the identification of a circular paradox (in other contexts defined as the "data citation vicious circle") consisting of scholars that do not share data because they feel that they are not rewarded by it, implying that the volume of data publications (or datasets shared) is currently still small.

Some of the main topics that have been discussed in the most important literature selected could be summarized as follows:

Discussions on benefits and conceptual issues

The main benefits derived from open data and data sharing have to do with a stronger open science (Fienberg et al., 1985), a higher efficiency in the use (and reuse) of scientific resources (Piwowar, 2011), the possibilities of expanding new research lines as well as other users (Piwowar, Day, & Fridsma, 2007) and the acceleration of the scientific progress (Piwowar, Becich, Bilofsky, & Crowley, 2008). Other benefits (Fienberg et al., 1985) include better science through the possibilities of verifying, refuting or refining scientific results, improvements of measurement and data collection methods, protection against faulty data, etc.

Rewards and citation advantage

One of the most important benefits from the point of view or rewards provided by the promotion and development of open data and data sharing activities is that they can function as a source of scientific recognition (Costas et al., 2013). An important issue that has been pointed as a strong advantage of data sharing from a reward point of view is the increase of the citation rates of primary publications (Piwowar et al., 2008, 2007). However, the generalized lack of recognition and reward systems for data sharing contributions has been also a point of concern in the research around the topic (Piwowar et al., 2008) suggesting the importance of establishing incentives for scholars and stakeholders (Piwowar et al., 2008). In this line, the main incentives from the point of view of scholars are (Van den Eynden & Bishop, 2014):

- Direct benefits: for the researchers, for science itself and for the different disciplines.

- Norms: that can come from the scholar or the disciplines.
- External drivers: including the policies and expectations from research funders and publishers.

Perceptions and cultural issues

Related with the previous is also the body of literature around perceptions and cultures regarding data sharing and open data. Thus, for example (Piwowar et al., 2008) has highlighted the importance and relevance of having a transition towards a culture of biomedical data sharing, provided recognition incentives are resolved. However, perceptions and cultural issues must develop not only to data sharing activities but also towards the establishment of cultures of acknowledging the (re)use of data, for example through citations (Costas et al., 2013). In this regard (Tenopir et al., 2011) have identified important barriers regarding data sharing that are perceived by the scholars. These barriers include particularly the lack of time and lack of funding for data sharing activities. Other barriers include the feeling of 'losing' control over the data by the creators of the data, the possible misuse of the data by others, the exposure of potential errors, as well as the potential lack of acknowledgement to the creators of the data (Cragin, Palmer, Carlson, & Witt, 2010).

Policy and technological developments

A fourth body of literature can be identified with the quest for solutions to the most important challenges related with data sharing and open data. These can be divided in two main groups: policy and technological needs. In the first group we can mention the need of reward systems and career structures (Arzberger et al., 2004) that include data sharing and data metrics. These rewards systems would also need necessary technical developments, including the development of standards and platforms for proper data sharing (e.g. software systems, data storage, data management, data compatibility, metadata standards, data identification, persistent identifiers, or granularity and versioning of datasets (Brase, Farquhar, Gastl, Gruttemeier, & Heijne, 2009; Dallmeier-Tiessen et al., 2012; Groves, 2010), but also the development of policies of tracking and evaluating data sharing contributions, as well as considering them during hiring, tenure and promotion decisions, as the considering of parts of the research budgets are attributed to data sharing (Piwowar et al., 2008).

Barriers and incentives in the context of Open Science

While institutional change and careers are issues related to the governance dimension, specific action points are very much related to the area of Open Science. The over-arching benefits and contributions of Open Access and Open Data are nowadays generally well understood by Member States and its funders and researchers.

However, institutions are usually expected to bear the burden of most aspects of the practical implementation of funder mandates, and to provide the necessary tools to researchers and other stakeholders related to the institution's library services. The citation advantage of Open publications is not a sufficient incentive for researchers to bear on their own the burdens of making their data open. Data managers are becoming more important members of research groups and research institutions, but their roles are not generally well defined and/or well resourced.

As a result, while the community may agree to the general principles, implementation of Open Science principles is not always smooth. Several barriers were identified in the practical implementation of Open Data as part of the FP7 SiS RECODE project[26]:

[26] Policy RECommendations for Open Access to Research Data in Europe (RECODE) project. Policy recommendations for open access to research data. Available at: http://recodeproject.eu

- Lack of financial support for long-term data preservation and curation
- Burden of evaluating and maintaining the quality, value and trustworthiness of research data
- Training of researchers and other relevant stakeholders, which is also very dependant on the needs and knowledge levels between and within disciplines, established research cultures and the pace of technological developments
- Awareness-raising on the opportunities and limitations of open access to research data

Added to these institutional Challenges is the burden of incentivising compliance with funder and national mandates, while giving tools (e.g. training, infrastructure, technical assistance) on how to go about publishing and re-using research data. The availability of proper incentives and support mechanisms (for researchers by their institutions and funders, and for institutions by their funders), will have an impact in the way that researchers and institutions will be able to capitalise on knowledge available in Open Access and Open Data archives.

A recent study from the UK's Expert Advisory Group on Data Access (EAGDA) for the Wellcome Trust[27] provides more insight into the challenges and missing incentive mechanisms in this area. In summary, there are three areas where funders and institutions should concentrate their efforts to eliminate these barriers:

- Appropriately resourcing data management and sharing activities
- Appropriately recognising and valuing the contributions of those who generate and share high quality data and datasets
- Appropriately supporting the cultural change required, leveraging on the predisposition of early-career researchers

[27] EAGDA. Establishing incentives and changing cultures to support data access (2014). Available at: http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/EAGDA/WTP056496.htm

3. Functional vocabulary of open access – definitions and terminology

Building on the results of the literature review, the purpose of this chapter is to arrive at a functional vocabulary of open access. The intention of the chapter is to present the definitions and terminology related to open access that allows for an empirical and practical approach to the concept of open access. The functional vocabulary will be the basis for the subsequent exploration of empirical studies and data on open access. Importantly, this chapter will consider the borderlines and relationship between the open access dimension and the remaining five dimensions of RRI.

When we are talking about the Open Science dimension of Responsible Research and Innovation, it is important to define and clarify the most important concepts, in order to avoid some of the more common confusions around the topic.

Open Access (OA) and Open Data (OD)

Open access is the idea of making research results freely available to anyone that wants to access and re-use them. One of the main drivers of the OA idea is to make publicly funded research accessible to the general public.

In the academic sense, the term Open Access referred originally to the provision of free access to peer-reviewed academic publications. Presently, the term also encompasses the free access to the research data that underpins publications or research projects, also referred on its own as Open Data (OD). Open Data is usually distributed with requirements of attribution and share-alike (copies or adaptations of the data need to be shared using the same principles as the source).

Open Access publishing models

There are several complimentary publishing 'roads' or 'routes' that publications can take to arrive at an Open Access state. The two most common are the 'green road' and the 'golden road':

The golden OA route/road (also called Open Access publishing): When a paper is published through this route, it is immediately available to anyone without the need for a subscription or viewing fee. This is the main route to publishing of Open Access and Hybrid journals. For fully open access journals, the business model of the journal is usually based on:

- Processing charges on the side of the authors (APC)
- Direct contributions from the journal owners, more common in case of journals owned by funders, institutions such as Universities, or professional bodies/associations.

As a result, in golden OA article processing charges (APC) are usually paid to the publishers by the author, their institutions or research funders. Therefore, the golden road shifts the model of publishers from that of copyright (exploitation based on exclusivity and access) to that of being paid for archiving and organising the peer-review process.

Funders at the national level such as the UK and the Netherlands are increasingly embracing this mode of OA. At European level, Horizon 2020 guidelines state that publications need to be made available to the general public at publication time. Author Processing Charges (APCs) that are incurred by beneficiaries are eligible for reimbursement during the duration of the project, and can be waived partially after it. A breach of this mandate can result in the grant being reduced.

Hybrid journals: Open Access journals can be created from scratch using this model (e.g. PLOS ONE) or may have evolved from subscription-based journals to Open

Access journals (e.g. Nature Communications). Journals also do not need to be fully Open Access or subscription-based and, in practice, many operate under a Hybrid model. In Hybrid journals, it is up to the authors to decide whether to pay APC so that their publication is freely available from the publisher's website. In a hybrid journal, the traditional method of access through subscription coexists with individual open access publications.

The green OA route/road (also called 'self-archiving'): In the green road of OA publishing, authors publish in traditional subscription-based journals and an additional copy of the paper (usually called the post-print, a final peer-reviewed copy without the format and branding of the journal) is stored in an open institutional or subject-based repository. An embargo period is established, usually of between 6-12 months, in which the publication is only available through subscription to the publisher. After the embargo period is over, the general public can access the publication from the open repository.

As the field of Open Access changes rapidly, we can observe new variations and classifications of Open Access publishing models. Table 2 shows one of such classifications, mentioning also fully distributed and bottom up models of publication (e.g. Diamond OA) and those outside of established channels (e.g. Rogue OA).

Table 2 Types of Open Access

Type	Description
Green OA	Full text (draft or published) manuscripts self-archived in a repository and/or accessible from personal, institutional, or subject websites after an agreed embargo
Gold OA - Journal	Open access journals with immediate free access, some of which (e.g. PLoS) operate on an author pays model
Gold OA - Article	(Also referred to as Hybrid OA) author pays the article processing costs (APC) to make articles published in a subscription based journal that are
Delayed OA - Green	Publisher specifies an embargo period (e.g. 6, 12, 18, or 24 months), after which a published article may self-archived in as open access repository
Delayed OA - Journal	Subscription-based journals whereby all or selected published articles are converted to open access after a specified period (e.g. 6, 12, 18, or 24 months)
Diamond OA	Researchers themselves organise the peer-review process. This results in a non-commercial, non-for-profit academic publishing model where it is free for researchers to publish and read.
Transient OA	Freely available on the web during a finite period (e.g. journal promotion); content changes in repositories and/or websites (e.g. updated or deleted manuscripts)
Restricted OA	Sample restrictions: access requires registration and/or membership in a group; limited use, such as read-only (not downloadable or not sharable; metadata not available for aggregation and/or analysis
Rogue OA	(also refereed to as Robin Hood OA) – Published manuscripts posted on websites or self-archived in repositories in conflict with licensing agreements and/or copyrights; may also contribute to transient OA

Source: Adapted from Archambault et al 2014, with contributions from Laakso and Björk 2013, and Fuchs and Sandoval 2013.

In addition to the broad categories of Gold, Green, and Hybrid modes of Open Access, multiple versions of a manuscript may exist due to variations in publishers' licensing

agreements. These agreements typically specify how, when, and under which conditions a manuscript may be openly accessible on the web. For example, a publisher may allow Green Open Access through self-archiving in an institutional repository. However, publishers' copyright restrictions differ on the stage of manuscript development that may be openly accessible, thus assigning different rights to different versions of the text. Commonly specified version types include the submitted manuscript (before peer review), the accepted manuscript (peer-reviewed but not formatted), and an exact copy of the published manuscript (Björk, et al. 2013). This creates the possibility that the Open Access version of a manuscript is substantively different from the published version. In such instances, it is unclear whether the Open Access version has been sufficiently validated through the quality control measures such as peer review.

Another variation is delayed access, which is applied as an embargo period, after which a copy of the publication may be self-archived or the publisher may remove access restrictions on the journal website. Embargo periods are generally specified as a delay of 6, 12, 18, or 24 months after publication, with 12 months being the most common embargo (Laakso and Björk 2013). The EC pilot exercise followed the green road model, with embargo periods depending on the nature of research (**Error! Reference source not found.**Table 1).

Table 3 Embargo periods, 7th Framework programme

	6 months	***12 months***
ERC	All areas except SSH: as soon as possible up to 6 months	Social Sciences and Humanities
FP7	*Thematic areas:* Health, Energy, Environment (including Climate Change), and Information & communication technologies (Cognitive Systems, Interaction and Robotics) *in the activity:* Research infrastructures (e-infrastructures)	*in the thematic area:* Socio-economic Sciences and the Humanities *in the activity:* Science in Society

Source: OpenAIRE website (https://www.openaire.eu)

For Green Open Access, it is thus left to authors and institutions to track and manage a variety of self-archiving policies, which in itself has been shown to be a barrier to Open Access (Davis and Connolly 2007). However, this kind of administrative overhead is largely absent from subscription journals that convert articles to Open Access after a specified delay (e.g. 12 months). In addition, a bibliometric analysis of 'delayed access' journals found journal and article impact factors higher than comparable averages from both subscription journals and direct (no delay) Open Access journals (Laakso and Björk 2013).

Public Sector Information (PSI)

Public Sector Information can be defined as the range of information that public sector bodies (and bodies governed by public law) collect, produce, reproduce and disseminate in many areas of activity while accomplishing their institutional tasks. PSI can include data about social, economic, geographical and demographical issues. Other types of PSI include data on weather conditions, traffic, tourism flows, business information and even multimedia libraries of state-owned media organisations.

This term is sometimes used interchangeably with the term Open Data. However, the two terms are different and should not be confused, as PSI does not imply that the information is free from copyright ownership, intellectual property rights, access restrictions, or processing and re-use charges. Part of the confusion between PSI and

OD comes from the fact that western economies are evolving towards a model of public sector information that allows its free access and re-use (sometimes referred as 'open government data'). The 2003 EC Directive on the re-use of Public Sector Information[28] encourages Member States to make available for re-use as much PSI as possible. The PSI directive was subsequently revised in June 2013 and it is to be transposed into national legislations by July 2015. The 2013 revision of the PSI Directive puts more emphasis on the requirements to Public Sector Bodies to allow the re-use of the information they create and collect as well as extending the scope of PSI to include public sector museums, libraries (including those of universities) and archives. The amendments also state that charges for the re-use of information should be reasonable, and provide provisions to arbitrate complaints. Open Data generated from Public Sector Information is sometimes referred to with the term **Open Government Data (OGD).**

Vocabulary in Open Data

From the Open Data and Data sharing point of view, some reports (Costas et al., 2013; Fienberg et al., 1985) have presented a summary of functional vocabulary on the Data sharing dimension:

Data sharing has been defined as the "voluntary provision of information from one individual or institution to another for purposes of legitimate research" (Fienberg et al., 1985) or simply "the release of research data for use by others" (Borgman, 2012). Data sharing requires the systematic *collection*, *curation* and *dissemination* (or *publication*) of data.

Data publication, referring to the publication of data mirroring the scientific publication model.

Data citations can be defined as formal citations included in the reference list of published articles to data resources. Thus, the concept of data citation is linked to the idea that datasets should be published just as other kinds of scholarly products and being considered relevant from scholarly and funding policy perspectives.

Data metrics, mainly related with data publication and data citation. It could refer to data publication and data citations (but also data 'altmetrics') as signals of use of data. For data metrics to build up, data sharing is a prerequisite.

An important source of functional vocabulary comes from the models suggested for data publication and data metrics (Costas et al., 2013). One of the models is related to the publication model. This model seeks to define discrete, well-described datasets, ideally with some level of quality assurance or peer review. In this regard we can talk about **standalone publication** (when data are published as a standalone dataset, with no requirement for a co-existing standard journal article describing the data, and the data archive provides systems which provide a data description document as the citable item). Alternatively, there are also **data journal publications**, published in **data journals** specialized in publishing **data papers**. A data paper is a journal publication whose primary purpose is to describe data (providing information on the what, where, why, how and who of the data). As such, a data paper contains facts about data, as found in conventional research articles.

[28] Directive 2003/98/EC, also known as the 'PSI Directive'.

4. Review of existing empirical knowledge of open access

In this chapter, which constitutes the bulk of the report, focus is turned to empirical studies in the area of open access. It presents the results of Sub-task 2.2 and Sub-task 2.3, which reviews the state of knowledge regarding the RRI dimensions, including empirical knowledge emerging from EC funded studies on the RRI dimensions. Results specifically for the open access dimension are presented in this report.

The chapter is divided into two parts. First, a selection of five EC studies (from the Framework Programme FP) with particularly rich empirical information on open access is reviewed (4.1). Second, a selection of other studies that equally hold rich information on open access is presented schematically (4.2). The aim of the review of EC studies is to 1) specify the questions concerning open access, to which the studies provide (partial) answers, 2) tentatively identify the indicators that may be harvested from the reviewed studies, 3) assess whether the information contained in the studies relate to the context, input, output, or outcome of open access following the intervention logic model, 4) specify the analytical level of the information, distinguishing between global, national, and sub-national (regional, institutional, programme/project and individual) levels, and 5) specify whether the studies provide quantitative or qualitative data. For the extensive list of other relevant empirical studies in 4.2, the aim is to summarize the sources of information, the analytical level at which information is presented, and the key focus of the studies, in order to pave the road to subsequent qualified selection of existing indicators of open access in Task 3 of the MoRRI project.

These specifications of the studies holding empirical information about open access will be used as the background for assessing the overall availability of empirical information on open access in the succeeding chapter.

4.1 EC studies and projects in the area of open access and open data

A number of commission projects have explored the dimension of open access. For the purpose of this report, 5 FP projects are reviewed which are considered particularly relevant for the open access dimension in terms of identifying empirical data for further analysis. These projects are listed in Table 4, below.

Table 4 Commission studies for review

Proposal Call	Project Acronym	Project Title	Project Start Date	Project End Date	Sources
FP7-SCIENCE-IN-SOCIETY-2012-1	RECODE	Policy RECom-menda-tions for Open Access to Research Data in Europe	01-02-2013	31-01-2015	http://recodeproject.eu/ http://cordis.europa.eu/project/rcn/106728_en.html?isPermaLink=true Reports: Sveinsdottir et al. (2013): Deliverable D1: Stakeholder Values and Ecosystem. RECODE. Available at: http://recodeproject.eu/wp-content/uploads/2013/10/RECODE_D1-Stakeholder-values-and-ecosystems_Sept2013.pdf Bigagli et a. (2014): Deliverable D.2.1:Infrastructure and technology challenges. RECODE. Available at: http://recodeproject.eu/wp-

Proposal Call	Project Acronym	Project Title	Project Start Date	Project End Date	Sources
					content/uploads/2014/04/D2.1-Infrastructure-and-technology-challenges.pdf Finn et al. (2014): Deliverable D3.1: Legal and ethical issues in open access and data dissemination and preservation. RECODE. Available at: http://recodeproject.eu/wp-content/uploads/2014/05/D3.1-legal-and-ethical-issues-FINAL.pdf Noorman et al. (2014): Draft Deliverable D4.1: Institutional barriers and good practice solutions. RECODE. Available at: http://recodeproject.eu/wp-content/uploads/2014/09/RECODE-D4.1-Institutional-barriers-FINAL.pdf RECODE policy recommendations for open access to research data – summary booklet D5 – Guidelines for different stakeholder groups on supporting open access to and preservation of research data *(Submitted January 2015)* D6 – Using existing open access networks to support policy harmonisation across Europe
FP7-SCIENCE-IN-SOCIETY-2008-1	SOAP	Study of open access publishing	01-03-2009	28-02-2011	http://project-soap.eu/ Report: Periodic Report Summary 2 – SOAP http://cordis.europa.eu/result/rcn/55371_en.html Periodic Report 1 – SOAP http://cordis.europa.eu/publication/rcn/14993_en.html Final Report Summary – SOAP http://cordis.europa.eu/result/rcn/55370_en.html
FP7-SCIENCE-IN-SOCIETY-2013-1	PASTEUR-4OA	Open Access Policy Alignment Strategies for European Union Research	01-02-2014	31-07-2016	http://www.pasteur4oa.eu/
FP7-INFRASTRUCTURES-2011-2	OpenAIRE-plus	2nd-Generation Open Access Infrastructure for Research	01-12-2011	31-12-2014	http://cordis.europa.eu/project/rcn/100079_en.html https://www.openaire.eu/

Proposal Call	Project Acronym	Project Title	Project Start Date	Project End Date	Sources
		in Europe			
FP7-ICT-2007-2	AEGIS	Open Access-ibility Every-where: Ground-work, Infrastruc-ture, Standards.	01-09-2008	31-08-2012	http://cordis.europa.eu/project/rcn/88209_en.html http://www.aegis-project.eu/

RECODE - Policy RECommendations for Open Access to Research Data in Europe

The FP7 led project RECODE (2013-2015) brings together existing networks, communities, projects and various stakeholders to identify and assess main challenges within the 'open access, data dissemination and preservation sector'. Furthermore, against the backdrop of good practices, general 'recommendations for a policy framework to support open access to European research data, will be produced'. The main objectives with the RECODE project are:

- To **reduce stakeholder fragmentation** in the area of open access to and dissemination and preservation of research data through: the identification of relevant networks and the facilitation of dialogue and collaboration between these networks
- To identify **stakeholder values and inter-relationships** in order to identify synergies and areas of conflict and promote collaboration on shared problems and solutions
- To identify **gaps, tensions and good practice solutions** for infrastructural and technological, legal, ethical, institutional and policy issues relating to the sharing of data
- To use **five case studies** to examine each of these areas across disciplinary boundaries
- To use stakeholder collaboration exercises to **identify and promote over-arching good practice policy solutions**
- To produce a **set of guidelines** that identify, promote and disseminate good practice solutions for the **sharing of scientific data** to stakeholders **across the open access** and data dissemination and preservation landscape (recodeproject.eu)

The project was finalised in January 2015. The consortium developed targeted guidelines for funders, research institutions, data centres and publishers on supporting open access to and preservation of research data. The final recommendations were presented at a conference attended by more than 200 participants from 40 countries. These provide relevant material and valuable insights for further data review in terms of indicator development. Six sub-deliverables have been published (see Table 5). They address various main challenges such as infrastructural or technological barriers to open access implementation and indicate possible solutions, best practice recommendations, among others, as well as the guidelines and a summary booklet for policy.

Table 5 Examples of open access indicators retrieved from RECODE

Guiding question	Indicator potential	Analytical level (intervention logic model)	Analytical level (aggregation)	Data classification and methods
What are the main values, motivations and barriers to Open Access among key stakeholders?	• Structural obstacles • Stakeholder attitudes	Input	Institutional National	Document analysis 5 case studies (including 29 interviews) Validation workshop
What infrastructural and technological barriers to open access can be identified? How can these be mitigated?	Structural and technological obstacles	Input	Institutional National European	qualitative, quantitative and document review met (5 key barriers identified)
Which legal and ethical issues relevant to open access can be identified?	• Formalized procedures/legislation: (intellectual property rights, including copyright, trade secrets and database rights, privacy and data protection ,open access mandates)	Input	Institutional National European	Literature review, case study interviews
Which challenges do institutions face to with regard to implementing open access?	• Data sharing infrastructure	Input		Literature review, case studies
Do specific open access platforms exist, nationally and transnationally?	• Networks • stakeholders	Output	National European	Qualitative and quantitative data
How can a general policy framework to support open access to European research data be constructed?	• Guidelines/recom-mentations	Output	European level	Qualitative and quantitative data (final project outcome)

SOAP - Study of open access publishing

The SOAP project, carried out between 2009 and 2011, gathered extensive world-wide information on open access publishing for key stakeholders such as the European Commission, publishers, libraries and research communities. The objective was for the stakeholders to gain knowledge about and assess which open access model/s would prove most beneficial in the transition towards open access publishing (project-soap.eu). The main objectives of SOAP were stipulated as follows:

The SOAP project described and analysed open access publishing. It compared and contrasted business models. Such an approach allowed for a better understanding of the marketplace as well as the opportunities and risks associated with open access publishing.

The SOAP project conducted a large-scale survey that investigating the European Research Area (ERA) scholars' requirements for scientific publishing. The survey's findings uncover what researchers as authors are willing to trade off (and what they are not) in the transition to open access publishing (project-soap.eu)

The vast amount of stocktaking word-wide data collected, through large-scale online surveys and state of the art analysis (Final Report Summary – SOAP), generated a very rich foundation for further analysis in terms of indicator development within the open access dimension (see Table 6, below).

Table 6 Examples of open access indicators retrieved from SOAP

Guiding question	Indicator potential	Analytical level (intervention logic model)	Analytical level (aggregation)	Data classification and methods
How can the landscape of open access journals be characterised?	• Number of articles/papers • Measures taken • Prevailing (business) models • Size of operations	Input	Global	Inventory of the state of the art, large-scale online survey of researchers' opinions and attitudes (a total of 53 890 responses)
What barriers to open access can be identified?	• Structural obstacles • Stakeholder attitudes • Gaps between offer and demand	Input	Global	Large-scale online survey of researchers' opinions and attitudes (a total of 53 890 responses)
What main drivers can be identified for publishing in open access journals?	• Typology of attitudes	Input	Global	Separate survey of scientists who published in open access journals

PASTEUR4OA - Open Access Policy Alignment Strategies for European Union Research

The PASTEUR4OA project was initiated early 2014 and will continue to 2016. Following the European Commission's Recommendation to Member States of July 2012 on the need to advance the 'access to and preservation of scientific information', PASTEUR4OA 'aims to help develop and/or reinforce open access strategies and policies at the national level and facilitate their coordination among all Member States.' Furthermore, a range of more specific objectives included in the project will result in:

- The identification of Key Node organisations throughout Europe and in accession/associated states, on the basis of their institutional profile, record and ability to influence policymaking, and the development of a network of expert organisations
- The development of a programme for engaging policymakers

- A Europe-wide project meeting of national experts
- Establish the foundations of a Knowledge Net by the end of the project through continuous engagement of the Key Node organisations.
- Recording policies and policy types in order to develop a policy typology
- Policy analysis: effectiveness and growth
- A mapping of existing policies to policymakers
- Development of advocacy materials
- The identification of policymakers in the MS and accession/associated states
- Policymaker engagement
- The project's final conference which will bring together Key Node members and policymakers, as well as provide the opportunity for a wider presentation of project achievements

Due to the initial phase of the project, no results have so far been produced. Nonetheless, the stock-taking exercise of current policies within the area of open access and the development of open access strategies, among others stated objectives, could potentially feature into a characterization of the open access dimension in terms of indicator development.

Table 7 Examples of open access indicators retrieved from PASTEUR4OA

Guiding question	Indicator potential	Analytical level (intervention logic model)	Analytical level (aggregation)	Data classification and methods
Which national/institutional based open access strategies/policies exist	Policies at the level of governments, research institutions	Input	National Institutional (across Europe)	Stock-taking analysis

OpenAIREplus -2nd-Generation Open Access Infrastructure for Research in Europe

The OpenAIREplus project (2011-2014) aimed to construct a '2nd-Generation Open Access Infrastructure' by continuing the work and the outcomes of its predecessor project OpenAIRE. The main objective of the OpenAIRE project was to 'support the implementation of Open Access in Europe' and to help realise the Open Access Policy set forth by the ERC Scientific Council Guidelines for Open Access as well as the Open Access pilot initiated by the Commission. The OpenAIRE project carried out between 2009 and 2012, resulted in 'an interoperable and validated network of more than 520 repositories and OA journals, integrating more than 9 million OA publications and 1,000 datasets, with 50,000 organizations and 30,000 projects from two funders. It has identified over 100,000 FP7 publications from about half the 26,000 FP7 projects, and offers literature-data integration services' (www.openaire.eu).

The OpenAIREplus project extends this work in several ways, for instance by facilitating 'access to the entire Open Access scientific production of the European Research Area, providing cross-links from publications to data and funding schemes'. Furthermore 'deposited articles and data will be openly accessible through an enhanced version of the OpenAIRE portal, together with any available relevant information on associated project funding and usage statistics' (www.openaire.eu; http://cordis.europa.eu/project/rcn/100079_en.html).

The vast amount of data produced through OpenAIRE and OpenAIREplus could provide an extensive data foundation for the collection of relevant statistics on open access resources across Europe as well as provide valuable data for further analysis.

Table 8 Examples of open access indicators retrieved from OpenAIREplus

Guiding question	Indicator potential	Analytical level (intervention logic model)	Analytical level (aggregation)	Data classification and methods
How many products are published in open access resources?	Total number of: • articles • papers • A/P – change over time	Output	European (EU27 and beyond)	A vast amount of publication repositories (9,189,509 publications and 6,290 datasets from 571 repositories and OA journals)
To what extent do researchers use open access platforms?	• Number of researchers • Change over time	Output	European (EU27 and beyond)	A vast amount of publication repositories (9,189,509 publications and 6,290 datasets from 571 repositories and OA journals)

4.2 Other empirical studies on the dimension of open access

In addition to the EC (FP) funded studies identified and reviewed above, a number of other studies offer relevant empirical information on issues related to open access (Table 8) and on open data (Table 9) in research and innovation contexts. For each entry, the analytical level in terms of aggregation is specified, and a brief note on the key focus of the study is provided.

Table 9 Other empirical studies on open access

Source	Type of source	Analytical level (aggregation)	Key focus
European Commission. Special Eurobarometer 401. Responsible Research and Innovation (RRI), Science and Technology	Report	National EU-level data disaggregated by socio-economic profile	Survey results Perception study
Eric Archambault, Didier Amyot, Philippe Deschamps, Aurore Nicol, Lise Rebout & Guillaume Roberge Proportion of Open Access Peer-Reviewed Papers at the European and World Levels—2004-2011 August 2013	Report	National Thematic Field	Bibliometrics Number and evolution of OA publications
Julie Caruso, Aurore Nicol & Eric Archambault Open Access Strategies in the European Research Area August 2013	Report	National	Benchmark Numbers of open access mandates, strategies and repositories
European Commission. Directorate-General for Research and Innovation. Report on the online survey on scientific information in the digital age held from July-September 2011.	Report	Data is given at EU-level, although questions were categorised per country, dataset might be available on request	Survey Perception study
European Commission. Directorate-	Report	EU-level, some responses	Survey

Source	Type of source	Analytical level (aggregation)	Key focus
General for Research and Innovation. Survey on open access in FP7.		are disaggregated per FP7 area	Perceptions on feasibility and difficulty of implementation of OA in researchers' workflow
Björk, et al. 2010. "Open Access to the Scientific Journal Literature: Situation 2009." PLoS ONE 5 (6): e11273.	Scientific article	International	proportion of peer reviewed scholarly journal articles that are openly accessible
Gargouri et al. 2012. "Green and Gold Open Access Percentages and Growth, by Discipline." In *Proceedings of 17th International Conference on Science and Technology Indicators*, edited by Eric Archambault, Yves Gingras, and Vincent Larivière. Montréal: Science-Metrix and OST.	Scientific article	International	Comparison of Green and Gold OA on the basis of proportion of OA and growth rate
Laakso, and Björk. 2012. "Anatomy of Open Access Publishing: a Study of Longitudinal Development and Internal Structure." *BMC Medicine* 10 (1): 124	Scientific article	international	volume of articles published in full immediate OA journals, while observing shifts in OA revenue models, publisher types and relative distribution among disciplines
Björk, et al. 2014. "Anatomy of Green Open Access." Journal of the Association for Information Science and Technology 65 (2): 237–50	Scientific article	International	meta-analysis of previous studies of Green OA combined with new data/analysis
Davis et al. 2008. "Open Access Publishing, Article Downloads, and Citations: Randomised Controlled Trial." *BMJ* 337 (jul31 1): a568–a568	Scientific article	US - 11 journals published by the American Physiological Society	OA citation advantage using hybrid OA and randomized trial
Eysenbach 2006. "Citation Advantage of Open Access Articles." *PLoS Biol* 4 (5):	Scientific article	International contributions to a single journal: Proceedings of the National Academy of Sciences	OA articles are cited earlier and are, on average, cited more often than non-OA article
Kurtz et al 2005. "The Effect of Use and Access on Citations." *Information Processing & Management*, Special Issue on Infometrics, 41 (6): 1395–1402	Scientific article	Publication records from two NASA Astrophysics Data System (ads.harvard.edu) and from the ArXiv e-print archive	The effect of access on citation
Laakso, and Björk 2013. "Delayed Open Access: An Overlooked High-impact Category of Openly Available Scientific Literature." *Journal of the American Society for Information Science and Technology* 64 (7): 1323–29	Scientific article	International	Citation advantage from delayed access journals
Lee et al. 2010. "Does Collocation Inform the Impact of Collaboration?" *PLoS ONE* 5 (12):	Scientific article	USA: Articles published by Harvard investigators	The effect of collaboration on citation
Moed 2007. "The Effect of 'open Access' on Citation Impact: An Analysis of ArXiv's Condensed Matter Section." *Journal of the American Society for Information Science and Technology* 58 (13): 2047–54.	Scientific article	International: ArXiv repository	The effect of early access on citation

Source	Type of source	Analytical level (aggregation)	Key focus
Xia et al. 2011. "Multiple Open Access Availability and Citation Impact." *Journal of Information Science* 37 (1): 19–28	Scientific article	International: 20 top library and information science journals	The effect of multiple access points on citation

Table 10 Other empirical studies on open data

Source	Type of source	Analytical level (aggregation)	Key focus
Carlos Iglesias. A year of Open Data in the EMEA region. Topic Report No. 2013 / 12. European Public Sector Information Platform. December 2013	Report	National	Observatory Prevalence and impact of open data initiatives around the world
Aurore Nicol, Julie Caruso, & Éric Archambault Open Data Access Policies and Strategies in the European Research Area and Beyond August 2013	Report	National	Benchmark Open data portals
Commission Staff Working Document: Impact Assessment Accompanying the document Commission Recommendation on access to and preservation of scientific information in the digital age {C(2012) 4890 final} {SWD(2012) 221 final}	Impact Assessment	EU-level National (only for some Key indicators on OA publishing in MS)	Appraisal of policy options and some cost-benefit analysis of adopting an OA policy for Europe
DG CONNECT European Public Sector Information (ePSI) Platform. http://www.epsiplatform.eu	Scoreboard Repository	National	Appraisal of status of PSI directives Survey (crowdsourced) Case studies of PSI and open data
Tim Davies et al. Open Data Barometer. 2013 Global Report. World Wide Web Foundation	Report Scoreboard	National	Profiling of countries with respect to the state of the art of PSI Impact assessment of PSI Also, assessment of quality of different government datasets
Tim Davies et al. Exploring the Emerging Impacts of Open Data in Developing Countries. World Wide Web Foundation	Report	--	Impact assessment Case studies focusing on developing countries
Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., & Moorman, D. (2004). Promoting Access to Public Research Data for Scientific, Economic, and Social Development. Data Science Journal, 3(November), 135–152.3(November), 135–152.	Scientific article	Global	General introduction on open data and data sharing
Borgman, C. L. (2012). The Conundrum of Sharing Research Data. Journal of the American Society for Information Science and Technology,	Scientific article	Global	General discussion on open data and

Source	Type of source	Analytical level (aggregation)	Key focus
63(6), 1059–1078. doi:10.1002/asi			data sharing
Brase, J., Farquhar, A., Gastl, A., Gruttemeier, H., & Heijne, M. (2009). Approach for a joint global registration agency for research data. Inforamtion Services & Use, 29, 13–27. doi:10.3233/ISU-2009-0595	Scientific article	Global	Technical discussion on registration of datasets and data citations
Costas, R., Meijer, I., Zahedi, Z., & Wouters, P. (2013). The value of research data - Metrics for datasets from a cultural and technical point of view. A Knowledge Exchange Report (pp. 1–48).	Report	Global	Comprehensive report on data metrics and data repositories
Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 368(1926), 4023–38. doi:10.1098/rsta.2010.0165	Scientific article	Global	Discussion on the feasibility of data sharing implementations for small science disciplines. Focus on institutional repositories.
Dallmeier-Tiessen, S., Darby, R., Gitmans, K., Lambert, S., Suhonen, J., Wilson, M., … Coordination, A. (2012). Compilation of results on drivers and barriers and new opportunities.	Report	Global. Europe	Development of a conceptual mode on process, context, drivers, barriers and enables of data sharing.
Fienberg, S. E., Martin, M. E., & Straf, M. L. (1985). Sharing Research Data. Washington: National Academy Press.	Book	Global	Fundational analysis on benefits and problems of data sharing
Groves, T. (2010). The wider concept of data sharing : view from the BMJ. Biostatistics, 11(3), 391–392. doi:10.1136/bmj.b3928.G	Scientific article	Global	Reflections on data sharing.
Kaye, J., Heeney, C., Hawkins, N., & Vries, J. De. (2009). Data sharing in genomics — re-shaping scientific practice. Nature Reviews, 10(May), 331–335.	Scientific article	Gobal	Reflections on data sharing(with a particular focus on genomics)
Knoppers, B. M., Harris, J. R., Tassé, A. M., Budin-ljøsne, I., Kaye, J., & Deschênes, M. (2011). Towards a data sharing Code of Conduct for international genomic research. GenomeMedicine, 3, 46.	Scientific article	Global	Code of conduct on data sharing in genomics
Krumholz, H. M. (2012). Open science and data sharing in clinical research: basing informed decisions on the totality of the evidence. Circulation. Cardiovascular quality and outcomes, 5(2), 141–2. doi:10.1161/CIRCOUTCOMES.112.965848	Scientific article	Global	Reflections on data sharing in the biomedical fields
Piwowar, H. A. (2011). Who shares? Who doesn't? Factors associated with openly archiving raw research data. PloS one, 6(7), e18657. doi:10.1371/journal.pone.0018657	Scientific article	Global	Analysis of the factors and determinants for data sharing activities.
Piwowar, H. A., Becich, M. J., Bilofsky, H., & Crowley, R. S. (2008). Towards a data sharing culture: recommendations for leadership from academic health centers. PLoS medicine, 5(9), e183. doi:10.1371/journal.pmed.0050183	Scientific article	Global	Analysis of attributes of data sharing systems. Recommendations are set forth.
Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is	Scientific	Global	Discussion on the Data sharing citation

Source	Type of source	Analytical level (aggregation)	Key focus
associated with increased citation rate. PloS one, 2(3), e308. doi:10.1371/journal.pone.0000308	article		advantage
Robinson-García, N., Jiménez-Contreras, E., & Torres-Salinas, D. (2015). Analyzing data citation practices according to the Data Citation Index. Journal of the Associatoin for Information Science and Technology, 1–20.	Scientific article	Global	Analysis of the Data Citation Index
Ross, J. S., & Krumholz, H. M. (2013). Ushering in a new era of open science through data sharing: the wall must come down. JAMA, 309(13), 1355–6. doi:10.1001/jama.2013.1299	Scientific article	Global	Discussion on the benefits of data sharing
Ross, J. S., Lehman, R., & Gross, C. P. (2012). The importance of clinical trial data sharing: toward more open science. Circulation. Cardiovascular quality and outcomes, 5(2), 238–40. doi:10.1161/CIRCOUTCOMES.112.965798	Scientific article	USA	Discussion on the importance of data sharing for clinical trial research
Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., … Frame, M. (2011). Data sharing by scientists: practices and perceptions. PloS one, 6(6), e21101. doi:10.1371/journal.pone.0021101	Scientific article	Global	Analysis of the practices and perceptions of scholars on data sharing
Torres-Salinas, D., Martín-Martín, A., & Fuente-Gutiérrez, E. (2014). Analysis of the coverage of the Data Citation Index – Thomson Reuters : disciplines , document types and repositories. Revista Española de Documentación Científica, 37(1), 1–6.	Scientific article	Global	Analysis of the coverage of the Data Citation Index.
Van den Eynden, V., & Bishop, L. (2014). Sowing the seed: Incentives and motivations for sharing research data, a researcher's perspective (p. 48).	Report	Gobal. Europe	Analysis on incentives and motivations for data sharing from the point of view of researchers

The policy reports and other sources listed above provide an overview of recent empirical work carried out on the topics of Open Access and Open Data. Both topics are nowadays at the forefront of science policy changes and, as a result, the landscape and availability of up-to-date information is quickly evolving, with new relevant studies in the area being published on a monthly basis. Additionally, it has been found that Open Access and Open Data scholars engage frequently in many forms of less structured discussion than that offered by studies and other scholarly communications, such as blogs and social media. As a result, some data points and empirical information may be difficult to track, as it is shared more on an ad-hoc basis.

Empirical information on Open Access Policy

Information focusing on policy developments and the benefits and impacts of policy regarding open access is plentiful, although empirical data is usually harder to come by. For example, historical accounts and literature reviews of open access policy give good overviews of the rationale towards the policy shift supporting open access. Additionally, these sources provide a more or less clear timeline of the main policy agreements and changes, both at European and Member State level.

Also in the area of Open Access, other reviews or qualitative benchmark studies look at the different degrees of implementation of OA policies in different countries, and offer qualitative information on the benefits. Some of the EC FP7 studies on Open Access commented in section 4.1 also provide qualitative information at national level (e.g. the pasteur4oa project contains country case studies focusing on the state of OA and OD policies).

Perception studies and stakeholder consultations

Outside of FP7, the empirical work carried out by the European Commission to date has mostly focused on perception studies, surveys and consultations of the research community. An example is the RRI Eurobarometer survey, which analysed the perception and support of the general public towards the overarching idea of open access. More detailed consultations to the scientific community were carried out in order to gauge the support for different aspects of the open access pilots and subsequent mandates in the Framework Programmes. These more specialised consultations included all the stakeholders, such as scientists, research institutions, funders and publishers. The results of these consultations are reported back in an aggregated way (although the consultation survey identifies responses according to their country).

Empirical information on the benefits of Open Access

Quantitative information and indicators on Open Access and the use benefits usually revolves around the amount of open access journals and open access publications available from researchers of a specific country. In this area, the most comprehensive study is the one carried out by Archambault et al. at ScienceMetrix in 2013. The study also contains information on implementation and take-up of open access policies at the national level. Finally, it also touches briefly on the aspect of Public Sector Information (PSI).

Empirical information on Open Data

In general, quantitative empirical information on the implementation, impact and benefits are more plentiful in the area of Open Data than for Open Access, although those in OD focus for the most part on Public Sector Information (PSI), rather than on Research Data. Main sources of quantitative information on implementation and impacts of PSI and Open Government Data at the national level can be found in the European PSI platform[29], run by DG Connect; and the Open Data Barometer[30], run by the Open Data Institute and the World Wide Web Foundation. Quantitative information on Research Data is scarce, and is mostly limited to general counts of OD repositories and OD datasets available in repositories such as Open Aire[31].

Table 10: Overview of the focus of the studies

Type of study	No.	Analytical level	No.	Key focus	No.
Scientific article	27	National	7	Perception	5
Report	11	EU level	10	Strategy/Policy	7
Book	1	Global	28	Discussion/barriers	17
Scoreboard/ Repository	3	USA	3	Infrastructure/ guidelines	4
Project FP	5			Data/metrics	18

Summarizing the focus of the studies (see Table 10), a number of reports and scientific articles cover initiatives and studies at the European and global level mainly, indicating that at the national – Member state – level the initiatives are not so

[29] http://www.epsiplatform.eu
[30] http://opendatabarometer.org
[31] https://www.openaire.eu/overall-stats/statistics/infra-monitoring

prominent. Regarding the key focus of the studies, there is still a lot discussion on how to organise OA and OD, overcoming infrastructural and organisational barriers. Also there is some attention for perceptions of scientists. Actual translation of policy into guidelines is scarce. Regarding metrics, the other empirical studies have been oriented towards publications and citation analysis for the OA publications, but the metrics and (data)citation orientation for Open data is in a starting phase. From the overview, it is clear that potential indicators derived from the FP projects are mainly input oriented (8 times), as compared to output oriented (4 times).

5. Assessment of data availability on open access

Based on the review and presentation of empirical studies on open access above, this chapter provides an overall assessment of data availability on open access for purposes of indicator development. The chapter discusses the issue of data availability in terms of 1) the extent to which the empirical studies provide relevant information across the categories of open access which were identified in the functional vocabulary, i.e. the extent to which the guiding questions that the studies address satisfactorily capture the contents of open access as defined in operational terms, 2) the balance and availability of quantitative and qualitative data respectively, 3) the extent to which available information address the four analytical levels specified in the intervention logic model, and 4) the availability of data at different levels of aggregation.

5.1 Data availability across open access categories

In the context of the MoRRI project Open Access is understood as activities in Open access of publications or as activities in Open Data. In general, the studies that are reviewed are explorative in nature; focus on perceptions of scientists, attitudes and cultural behaviour rather than on actual data. From a data-perspective, the studies analyse feasibility, coverage, metadata and implementation issues of Open Access and Open Data. Data availability from repositories may not be the main issue since for publications bibliometric methods are available, and consolidated data sets such as the Web of Science from Thomson Reuters or Scopus from Elsevier are systematically collected. However, the value of measurements of Open Access to assess RRI remains to be settled (see below). Regarding Open Data, the actual data availability is relatively scarce even though huge repositories are available in some scientific fields. Open Data is also a relatively young area of interest, and it is beneficial that a worldwide effort has started (see below) to increase, harmonise, and integrate Open data activities for the benefit of science.

Open Access

To address the variability of circumstances associated with Open Access publishing, recent studies invert the assessment of Open Access publishing, from top-down queries of a bibliometric dataset to bottom-up testing whether a publication is an Open Access publication (van Leeuwen et al. 2015). This approach involves random sampling of a given publishing domain, harvesting full-texts from the Internet, and analysis of available metadata from harvested manuscripts (Björk et al. 2010; Gargouri et al 2012). While this approach circumvents much of the variability noted above, it is nevertheless dependant on the presence and quality of metadata. Building on this technique, a recent European study provides the latest estimates for OA publishing (Archambault et al. 2014). Methodological advances in this study include increased sample size (1.25 million records tested for OA status) and a refined harvester with 99.1% "retrieval precision" (accuracy of finding full-text OA articles among the sample set of publication records) and 86.4% recall rate, or percentage of "true positives" among the set of found OA texts (p. 7). Accuracy of these findings is a measure of the frequency that the harvester correctly retrieves full text articles. In a pilot study, the harvester was tuned to minimize occurrence of finding full text articles incorrectly identified as open access (false positive) and to minimize occurrence that openly accessible articles were not found (false negative) (pp. 7-8).

The random sample of 1.25 million records is taken from Scopus data of published articles indexed between 1996 and 2013, which, according to the authors, provides broader international and disciplinary coverage than other relevant databases. This

sample is comprised of two subsets: (1) 250,000 records collected for articles published during the period 1996-2013, and used to analyse historical evolution of OA generally and with regard to different types of OA (p. 10) and (2) 1,000,000 records were collected for articles published during 2008-2013, and used for "deep" analysis of OA output from the European Research Area as a whole, from each of the member countries, and for comparative analysis across 22 scientific fields (p. 25).

Like previous studies, Archambault et al. (2014) employ an operational definition of OA for the harvesting: freely available (on the internet) scientific papers published in peer-reviewed journals. Operational definitions were also developed for OA types, thus simplifying the composition of harvested texts to 4 types: Green OA, Gold Journals OA, and Other OA (p 5). A summary finding from this study is, as of April 2014, more than 50% of the scientific papers published in 2007, 2008, 2009, 2010, 2011, and 2012 can be downloaded for free on the Internet" (p. ii). Extensive detail and breakdown of percentages and trends is provided for countries, scientific fields and type of OA, and is presented in table format throughout the report and as appendices.

Open Data

Most of the research on open data and data sharing is focused on case-studies based on desk research, interviews or surveys. However, particularly in (Costas et al., 2013) a special effort has been made in order to systematize available data sources for the analysis of data sharing and the development of data metrics. Thus, the report presents several initiatives and organizations focused on standardisation and harmonisation of data and measures. Among them we can mention the Research Data Alliance (RDA) (Genova et al., 2014), which aims to promote the international cooperation and infrastructure that scientific data sharing requires. The International Council for Science (ICSU), particularly the ICSU World Data System, which aims at a transition from existing standalone services to more worldwide integrated communities for scientific data. Similarly, the ICSU Committee on Data for Science and Technology (CODATA) has as its mission to strengthen international science by promoting scientific and technical data management and use. CODATA is concerned with all types of quantitative data resulting from experimental measurements or observations in the physical, biological, geological and astronomical sciences. Particular emphasis is given to data management problems common to different scientific disciplines and to data used outside the field in which they were generated. The general objectives are the improvement of the quality and accessibility of data, as well as the methods by which data are acquired, managed and analysed; the facilitation of international cooperation among those collecting, organizing and using data; and the promotion of an increased awareness in the scientific and technical community of the importance of these activities. From it side the

The Australian National Data Service (ANDS) is also an organization that seeks 'more researchers reusing more data more often'. In order to make this possible, ANDS is building the Australian Research Data Commons (ARDC). The ARDC is a combination of the shareable Australian research collections, the descriptions of those collections including the information required to support their re-use, the relationships between the various elements involved (the data, the researchers who produced them, the instruments that collected them and the institutions where they work), and the infrastructure needed to enable, populate and support the commons. This combined information can then be used to help people discover data in context. ANDS will build a set of interlinked web pages and make them available for harvesting by web search engines at Research Data Australia (which is part of the Research Data Alliance described above.

Another relevant organization is Researchfish, which provides a comprehensive research outcomes and impact evaluation system. Researchfish is an online facility that enables research funders to track the impacts of their investments, and researchers to easily log the outcomes of their work. The web portal connects funding organisations, researchers and universities - linking up individual research outcomes with multiple funder contributions. This enables instant and thorough reporting, so that funding organisations can track the progress of research and ensure value for money, whilst saving time and money on complex administration. A simple to use permissions-based interface with a comprehensive reporting capability replaces lengthy and expensive data cleaning, organising and analysis. Huge amounts of data across a vast range of researchers or awards can be combined and transformed into comprehensive graphs, to show meaningful results. For researchers, the portfolio provides a simple to use environment for associating research outcomes with awards and satisfying funder reporting requirements. CV Builder draws upon reported outcomes and maintains a current profile for use with funders, publishers, press or others. For research organisations there is the opportunity to view all awards held by its researchers and it allows to monitor funder-required submissions by researchers at their organization. Currently, over 90 funding and research organisations are using the system, tracking outcomes on over 42,000 awards. Over 35,000 researchers are now reporting through the system in over 40 countries with a 95% compliance rate. In this way, over £40 billion in awards are being tracked in various currencies. The Research Councils UK (RCUK) is actively working with Researchfish Ltd to implement better interoperability with the research information systems used by Research Organisations (ROs). Both technically, and from a user perspective Researchfish offers advantages with regard to Open access and open data, alongside other research outputs such as engagement activities (relevant for the dimension Public engagement), research tools and methods, Intellectual property and licensing, software and technical products, research databases and models to mention just a few.

In terms of concrete data about data sharing and data citations, Costas et al (2013) highlighted a series of capital sources with interest for the study and development of metrics on open data and data sharing activities. A total of 35 different repositories have been checked. All the repositories analysed offer "open" access to their data, however for 12 of them (34%) registration or some restrictions are involved (e.g. only for research and academic purposes, only for academic users, etc.). Most of the repositories are of a medium to large size, with numbers of datasets ranging from ~5,000 to 11 million records. Not all repositories mention the content that they offer. The majority of repositories (74%) show some kind of identifiers. They present a broad diversity in types of identifiers, and only 8 repositories (23%) present DOIs for their datasets. Below, we describe 3 data repositories and 2 tools for citation analysis in more detail, including access, usage, validation and metrics characteristics (see table 11). Quantitative data are presented when available.

GigaScience (http://www.gigasciencejournal.com/). This is an integrated online open-access open-data journal co-published in collaboration between BGI Shenzhen (the largest genomic organisation in the world) and BioMed Central, focus on biological and biomedical research. It aims to encourage data dissemination, organisation, understanding, and use from the entire spectrum of life and biomedical sciences. The journal has a novel publication format: one that links standard manuscript publication with an extensive database (providing DOI assignment to every dataset) that hosts all associated data and provides data analysis tools and cloud-computing resources.

Pangaea (http://www.pangaea.de/) (Elsevier) is an information system operated as an Open Access library aimed at archiving, publishing and distributing georeferenced data from earth system research, linking primary data related to articles in earth and environmental science journals. Pangaea is open to any field of earth system research,

enabling a bibliographic citation of datasets with identification via DOI and a widespread distribution through portals, library catalogues and search engines. Pangaea started a collaboration with Elsevier, a publisher, to interconnect the diverse elements of scientific research. Elsevier articles at ScienceDirect are now enriched with graphical information linking to associated research datasets that are deposited at Pangaea.

Dryad (http://datadryad.org/) is both an international repository of data underlying peer-reviewed articles in the basic and applied biosciences, and a membership organisation, governed by journals, publishers, scientific societies, and other stakeholders. Dryad welcomes data submissions related to published, or accepted, scholarly publications, in particular for tables, spreadsheets, and all other kinds of data that do not have another discipline-specific repository. Dryad also welcomes the involvement of journals, editors, publishers, authors and others who support data archiving. Authors may submit data files associated with their publications. Editors and journals can facilitate their authors' data archiving by setting up automatic notifications to Dryad of accepted manuscripts, streamlining the authors' process for depositing data. Dryad is developed by the National Evolutionary Synthesis Center and the University of North Carolina Metadata Research Center, in coordination with a large group of Journals and Societies. The National Evolutionary Synthesis Center is a joint effort of Duke University, the University of North Carolina, and North Carolina State University.

DataCite (http://datacite.org/) is an international not-for-profit organisation formed in London on 2009. The aims of DataCite are to establish easier access to research data on the Internet, to increase acceptance of research data as legitimate, citable contributions to the scholarly record, and to support data archiving that will permit results to be verified and re-purposed for future study. DataCite seeks to support researchers by helping them to find, identify, and cite research datasets with confidence (i.e. discoverability of datasets), to support data centres by providing persistent identifiers for datasets, workflows and standards for data publication (i.e. helping to solve to problem of identification and traceability of datasets); and to support journal publishers by enabling research articles to be linked to the underlying data. DataCite also contributes to assign persistent identifiers to datasets, by developing an infrastructure that supports simple and effective methods of data citation, discovery, and access. DataCite is leveraging the Digital Object Identifier (DOI) infrastructure, which is well-established and already widely used for identifying research articles (although they also keep an open approach by considering also other identifier systems). In this regard, all DataCite DOIs resolve to a public landing page that contains information about the associated dataset and a direct link to the dataset itself.

Data Citation Index. This is a Thomson Reuters' product result of the collaboration with some of the most important research libraries and digital repositories (e.g. California Digital Library, Protein Data Bank, Pangaea, UK Data Archive, etc.) in order to design a single source of data discovery for the sciences, social sciences, and arts and humanities. The Data Citation Index claims to fully index a significant number of the world's leading data repositories of critical interest to the scientific community, including over two million data studies and datasets. The records for the datasets, which include authors, institutions, keywords, citations and other metadata, are then connected to related peer-reviewed literature indexed in the Web of Knowledge. Thus, it is possible to track and count the citations that an individual dataset has received in the scientific literature. The Data Citation Index intends to solve three of the major issues that frustrate and discourage researchers from submitting their data to repositories: Discovery: as a database the Data Citation Index allows the user to search by different parameters and thus be able to retrieve and discover datasets that

could be of their interest. Attribution: each result page contains a how to cite this resource link with a recommended citation format. This could help to establish citation conventions that could also help to get better data metrics. Thomson Reuters is partnering with researchers to recommend and standardize how citation should be collected and cited for datasets. Measurement: as a result of the linkages from data to literature it is possible to calculate the number of citations that a dataset has received and thus being able to "assess" the use of the dataset by other researchers.

Table 11: Access, usage, validation & metrics" features of a selection of datarepositories (from Costas et e. 2013)

Title	Possibility of search/browse	Data availability & Data Type/format	Data validation	Presence of Resource identifier	Metrics Available	Standard citation format
GigaScience (http://www.gigasciencejournal.com/)	Yes	Biological data producers, integrates manuscript publication with a database contains both genomic and non-genomic datasets.	yes	DOI	not found	yes
Pangaea (http://www.pangaea.de/)	yes	Data Publisher for Earth & Environmental Science. Various scientific data types are archived with detailed description in a relational database. The description of each data set is always visible and includes the principle investigator (PI) who may be asked for access.	yes	DOI	not found	yes
Dryad (http://datadryad.org/)	Yes	Over 2615 data packages and 7172 data files, associated with articles in 189 journals	yes	DRYAD DOI	yes	yes

5.2 Availability of quantitative and qualitative data

Open Access

Publication indices can be retrieved from commercial suppliers, such as Elsevier's Scopus or Thompson Reuters' Web of Science, or from the freely avaialble Google Scholar. In addition, there are a number of data resources associated with the domain of Green OA:

- Directory of Open Access Journals (DOAJ): directory includes 10,383 Journals, 6,199 searchable at Article level, 136 Countries, and 1,862,682 Articles (accessed 30 March 2015) http://doaj.org

- Directory of Open Access Repositories (OpenDOAR): includes a "descriptive list" of 2,860 open access repositories (accessed 30 March 2015) http://www.opendoar.org/

- Registry of Open Access Repositories (ROAR): directory includes a database of 3,924 open access repositories, browse-able by Country, Year, Repository Type, Institutional Association, and Repository Software (accessed 30 March 2015) http://roar.eprints.org

- Registry of Open Access Repositories Mandatory Archiving Policies (ROARMAP): includes over 600 open access mandates organized by the following types: Funder (73), Funder and research organisation (53), Multiple

research organisations (8), Research organisation (e.g. university or research institution) (464), and Sub-unit of research organisation (e.g. department, faculty or school) (69). (accessed 30 March 2015) http://roarmap.eprints.org

- SHERPA/RoMEO maintains an index of publishers' copyright policies concerning Green OA: tracks over 22,000 journals in local database and searches external databases such as Zetoc, DOAJ, and Entrez. RoMEO also searches the *Zetoc*, *DOAJ*, and *Entrez* databases for additional journals. (accessed 30 March 2015) http://www.sherpa.ac.uk/romeo/

Open Data

The majority of studies analysed are either focused on more conceptual discussions as well as identifying barriers and are not really based on data. The more empirical ones are based on data from databases such as PubMed, or Google Scholar (Piwowar, 2011), specific samples of publications (e.g. microarray clinical trials, Piwowar et al., 2007) and more recently studies based on the Thomson Reuters Data Citation Index have also started to be developed (Robinson-García, Jiménez-Contreras, & Torres-Salinas, 2015; Torres-Salinas, Martín-Martín, & Fuente-Gutiérrez, 2014). Probably the best sources of data about open data and data sharing are the different data repositories currently existing. In (Costas et al., 2013) an extensive and thorough discussion on diverse data repositories has been systematically presented. Some of the main characteristics described for these repositories can be summarized as follows:

- All the repositories offer "open" access to data (which is for example not possible through the Data Citation Index, which only covers citations to the covered datasets). .
- The age of the data repositories varies, some of them having existing from before the year 2000.
- Searching and filtering possibilities are presented in most of them, allowing for the "discovery" of data.
- Different formats of data are recorded in these data repositories (e.g. photographs, maps, surveys, etc.).
- Although more than half (57%) of the repositories have some level of validation of the data deposited, this is not common for all existing repositories.
- They still present a relatively low level of standardization regarding the presence of DOI and other identifiers for the datasets deposited.
- Repositories seem to have the technology to accrue metrics about the use and reception of the repositories, but the presence of metrics is not present across all repositories.

All the previous indicates an important development of infrastructures for data depositing and repositories. These repositories can be good sources for the analysis of the presence, evolution of data sharing initiatives across disciplines and scholarly actors.

5.3 Availability of data across analytical levels included in the intervention logic model

Following the MoRRI proposal, indicators will be considered for different levels or phases of the 'logic model' of Open Acess and Open Data activities. These levels

include the 'Context', i.e. the overall environment for Open Access and Open Data, ' Input', i.e. the Open Access and open Data activities that are carried out, measures taken, structures created or resources provided to stimulate sharing of knowledge and data across Europe, 'Outputs', i.e. the immediate or direct results of such activities like data citation and citation of openly accessible knowledge in publications, and the 'Outcomes' i.e. the mid- and long term achievements and consequences of open access and open data sharing of knowledge with society.

The empirical information that emerges from the studies presented above mainly concern the 'input' level and the Context in which this is happening. However, also the direct results of such activities namely tracking the use of open access publications and data can be assessed.

In terms of outcome, more debate will follow on the actual benefits of Open Access publications beyond citation impact. For instance outcomes in terms of other people than scientists using the knowledge, and for Open Data, a measurable effect of lack of redundancy of collecting data across different Member States

For the continued work in MoRRI, it is useful to explore further the extent to which the studies might contribute to the development of indicators for the 'impact/benefit' side of the project, i.e. activities related to Tasks 6-8 of the project.

5.4 Availability of data at different levels of aggregation

With regard to the matter of different levels of aggregation of the available empirical data, a distinction was made between OA/OD data at the global level, the national level, and the sub-national level, the latter including regional, institutional, and individual level data.

Open Access

Publications are a globally available source and through the characteristics of a publication (address, DOI, funding acknowledgement) a wide range of analytical levels relation to geographical (global, national, regional, institutional) scope can be assessed. This however does not mean that similar practices exist across different science fields. Also the availability of data may differ across fields, i.e. in the social sciences and humanities a journal publication may not be the preferred avenue of sharing knowledge or for communication with peer scientists.

Open Data

From the point of view of open data and data sharing activities, the best sources of information are the repositories and deposited data as well as traces on their use (e.g. citations or altmetrics on the usage and reuse of the data). Some of the reviewed works have also studied actors and stakeholders on open data through surveys, however these studies are case-studies and there are not data systems that systematically collect this type of data.

Regarding the similarities of the data publication model with the scientific publication model of articles (Costas et al., 2013), similar levels of aggregation can be outlined:

- *Item level* (e.g. datasets). This is the lowest aggregation level, focusing on the individual data items and events surrounding them (e.g. datasets, data publications, citations and 'altmetric' mentions of the datasets, etc.)

- *Actor level* (e.g. data creators, data users, institutions, countries, etc.). This level focuses on the study of the different actors related with the production and impact of the data items. Thus actors can be the authors or creators of the dataset (as well as the users of the datasets), but also their institutions and countries. Thus a similar model of aggregation levels and analytical units as existing for scientific publications can be also established for the analysis of open data and data sharing activities.
- *Disciplinary level*. Given the resemblance with the publication model, a disciplinary aggregation level can be also established for the analysis of datasets and data sharing.
- *Venue level*. The existence of multiple data repositories and data journals also allows for the analysis of the venues of publication of the data. Thus metrics and indicators could be also calculated for repositories, data journals, or data sharing platforms and initiatives.

6. Data selection for RRI monitoring – reflections of current data gaps and required data collection

The purpose of this chapter is to assess data gaps and provide reflections on the need for primary data collection in order to mitigate data gaps, based on the contents and results of the previous chapter as well as on the list of promising indicators constructed in chapter 7.

The summary table 6.1.1 below, capturing the contents of chapter 7, serves as a basis for assessing the potential to develop new indicators based on existing empirical material.

Table 6.1.1. Summary table capturing the contents of chapter 7

INDICATOR	**ANALYTICAL MODEL (Logic model)**		**ANALYTICAL LEVEL (aggregation)**		**UNIT OF ANALYSIS**		**NUMBER OF OBSER-VATIONS**	**TIME SERIES**		**YEAR OF DATA, MOST RECENT**
	CONTEXT INPUT OUTPUT OUTCOME	(1) (2) (3) (4)	GLOBAL NATIONAL REGIONAL INSTITUTIONAL PROGRAMME/ PROJECT INDIVIDUAL	(1) (2) (3) (4) (5) (6)	COUNTRIES INSTITUTIONS INDIVIDUALS PUBLICATIONS OTHER (PLEASE SPECIFY)	(1) (2) (3) (4) (5) (6)		Y N	(1) (2)	
Indicator 1	1		EU		1		3	2		2013
Indicator 2	3		1		1		4	1		2013
Indicator 3	2/3		1		1		3	2		2013
Indicator 4	1		EU		1		4	2		2012
Indicator 5	3		EU		3		2	2		2012
Indicator 6	3		EU		3		4	2		2012
Indicator 7	1		1		1		4	2		2013
Indicator 8	2		EU		1		1	2		2011
Indicator 9	3		EU		1		1	2		2011
Indicator 10	3		EU		1		1	2		2011
Indicator 11	3		1		1,2,3,4		1	possible		2013
Indicator 12	4		1		1,2,3,4		1	possible		2013

Problems in identifying Open Access publications

Getting comprehensive information on Open Access publications is still difficult. Open Access publishing is strongly supported from a EU policy perspective, and three databases for bibliometric analyses are available, suggesting that data gaps in OA publications should be minimal. This however, is not the case. A certain level of comprehensiveness is needed if one wants to calculate metrics. Google Scholar provides access to publications. It is feasible to retrieve 50-70% of publications without paying, but it is a non-automated and labour intensive process. It is impossible to do any aggregate statistical analysis or calculate any metrics.

The Scopus database provides a desktop version for amateur users, but this version does not contain an open access functionality, and the Scopus database does not provide an Open Access label to publications. When paying a fee for Scopus or for the Web of science, access to the full database is possible. Then it is possible to couple the DOAJ list with the database and retrieve the publications that are declared Open Access (only Gold and Hybrid).

The Web of Science database has labelled OA publications in their database. But when comparing the labelled publications and the publications retrieved through coupling, there is a big difference in numbers, which can be explained by the fact that the DOAJ list of journals declare publications OA retrospectively. This means that publications that were once non-OA, will end up as OA. Including other technical complications, e.g. by using DOIs, the current situation is that Open Access is in a transition phase. Due to these technical issues, and other theoretical problems not yet solved by bibliometricians, it is not recommended to use commercial datasets (eg WoS and Scopus) to calculate impact figures as an indicator for evaluation purposes.

As a result of the above, along with other challenges in identification of OA publications, state of the art methods for assessment of OA include testing a random sample of publication records for free accessibility on the web (e.g Archambauldt et al 2014). Limitations of this approach include incomplete coverage of research output in the main science index systems (van Leeuwen 2013) and inconsistent and/or nonexistent standards for identification of a publication's OA status. Random samples are often based on WoS or Scopus datasets, which both have limited coverage in especially the humanities and social sciences. In addition, the harvester method relies to some extent on the quality of metadata in harvested publications.

Barriers in Open Data

From the point of view of open data and data sharing, the main data gaps can be summarized as follows:

- *Structural scarcity of data*: basically the lack of incentives to share the data as well as the lack of culture in acknowledging (citing) datasets implies that the data sharing events are low and therefore posing a challenge for the the development of indicators based on them (cf. Costas et al., 2013). This means that even when counting with the best resources to collect indicators on data sharing the numbers will be low[32], thus threatening the validity and reliability of any indicator based on data metrics..
- *Diversity and dispersion of data repositories and data venues*. As shown by Costas et al (2013) there are multiple and diverse data repositories, that vary from disciplinary to institutional, etc. (35 different data repositories were analysed in Costas et al, 2013). This causes a strong dispersion of sources that can be used for the development of data metrics, complicating the data

[32] New studies are corroborating this low data citation and altmetric activity around research data (http://arxiv.org/abs/1501.03342).

collection; and at the same time the diversity of repositories, with different policies, structures, available meta data, etc. causing that the integration and condensation of data about data metrics becomes problematic.

Feasible (partial) solutions to the previous problems could be to work with comprehensive data repositories and data sources for metrics, including here sources such as the Data Citation Index, Data Cite or Dryad (as well as some of the initiatives presented before) as these are probably some of the most extensive, complete and standardized sources useful for the development of data metrics. Also, the consideration of repositories from the list reported by Costas et al (2013) could be necessary for case studies on diverse disciplines and geographic areas (e.g. DataFirst for Africa, PDB (Protein Data Bank) for the area of proteins and DNA research, or GenBank for genetics research).

More and more, research funders require research data management plans in funding applications. This is the case in Horizon 2020, but also Member States include such plans I there funding requirements. However, only when research data management gets compulsory more open open data can be expected. In the meantime, at the member state level research data management plans could serve a s proxy indicator for open data. Then also the institutional barriers need to be addressed (see: http://recodeproject.eu/wp-content/uploads/2014/09/RECODE-D4.1-Institutional-barriers-FINAL.pdf)

7. Early thoughts on open access indicators

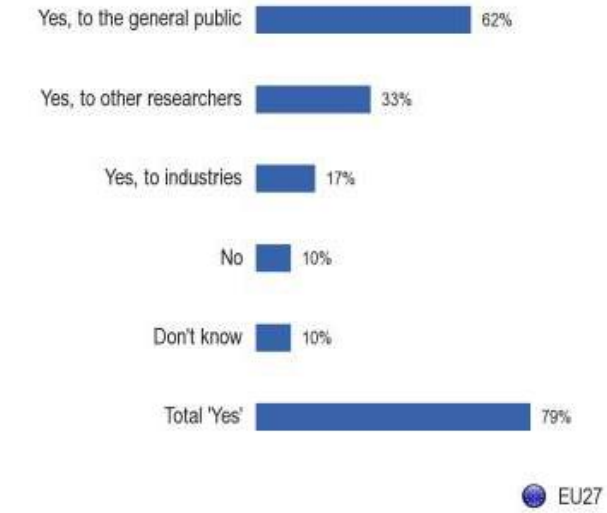
This chapter provides a space for compiling promising indicators based on existing empirical information identified throughout the report. Indicators have been selected on the basis of relevance and data availability. The intention is to prepare for the ground for Task 3, in which the selection of existing indicators and the development of new ones will take place.

Table 7.1.1: Potential indicator for OA, no. 1

Information Item	**OA1**
Name of indicator	Public perception of online free availability of the results of the publicly funded research
Brief description	The indicator showcases what is the public perception of online free availability of the results of the publicly funded research in the EU Member States. Data are collected on the EU-level, but can be disaggregated by individual Member States or by various socioeconomic profile (gender, age, level of education, attitude to science).
Analytical level (logic model)	Context-related
Analytical level (aggregation)	EU-level, Country level, Socioeconomic profile level
Qual / Quant	Quantitative
Source of data	Indicator presented at European Commission. Special Eurobarometer 401. Responsible Research and Innovation (RRI), Science and Technology
Date	2013
Time-series	No
Measurement level	Ordinal
Unit of analysis	Countries, socioeconomic profile
Coverage	28 Member States
Attributes	Results of publicly funded research should be made available online free of charge (MULTIPLE ANSWERS POSSIBLE): • Yes, to the general public • Yes, to other researchers • Yes, to industries • No • Don’t know

Table 7.1.2: Data presentation, OA1

QD17. Do you think that the results of publicly funded research should be made available online free of charge?



QD17 Do you think that the results of publicly funded research should be made available online free of charge?

	Yes, to the general public	Yes, to other researchers	Yes, to industries	No	Don't know	Total 'Yes'
EU27	62%	33%	17%	10%	10%	79%
CY	**80%**	47%	32%	*2%*	8%	90%
EL	**78%**	38%	23%	*3%*	9%	88%
UK	**75%**	29%	20%	*7%*	8%	85%
IE	**74%**	39%	24%	*6%*	9%	85%
ES	**71%**	29%	15%	*6%*	12%	82%
MT	**71%**	21%	16%	*4%*	20%	77%
SI	**68%**	32%	18%	*12%*	7%	81%
FI	**67%**	45%	32%	*6%*	3%	90%
LV	**66%**	22%	20%	*7%*	6%	87%
NL	**65%**	35%	15%	*13%*	4%	83%
CZ	**64%**	30%	11%	*7%*	8%	85%
EE	**64%**	26%	13%	*11%*	7%	82%
LU	**64%**	42%	28%	*16%*	5%	79%
PL	**64%**	26%	14%	*11%*	13%	76%
LT	**62%**	26%	16%	*7%*	15%	78%
SE	**61%**	46%	27%	*13%*	4%	83%
DK	**59%**	36%	24%	*18%*	5%	77%
FR	**59%**	50%	27%	*14%*	6%	80%
DE	**58%**	27%	*13%*	16%	12%	72%
PT	**58%**	32%	18%	*8%*	17%	75%
SK	**58%**	27%	13%	*9%*	11%	80%
IT	**56%**	30%	11%	*8%*	11%	81%
BE	**55%**	47%	21%	*14%*	2%	84%
AT	**55%**	36%	*17%*	18%	7%	75%
BG	**54%**	18%	10%	*3%*	30%	66%
RO	**49%**	24%	12%	*9%*	24%	66%
HU	**47%**	**47%**	24%	*11%*	9%	81%
HR	77%	21%	13%	8%	7%	85%

Highest percentage per country	*Lowest percentage per country*
Highest percentage per item	Lowest percentage per item

QD17 Do you think that the results of publicly funded research should be made available online free of charge?

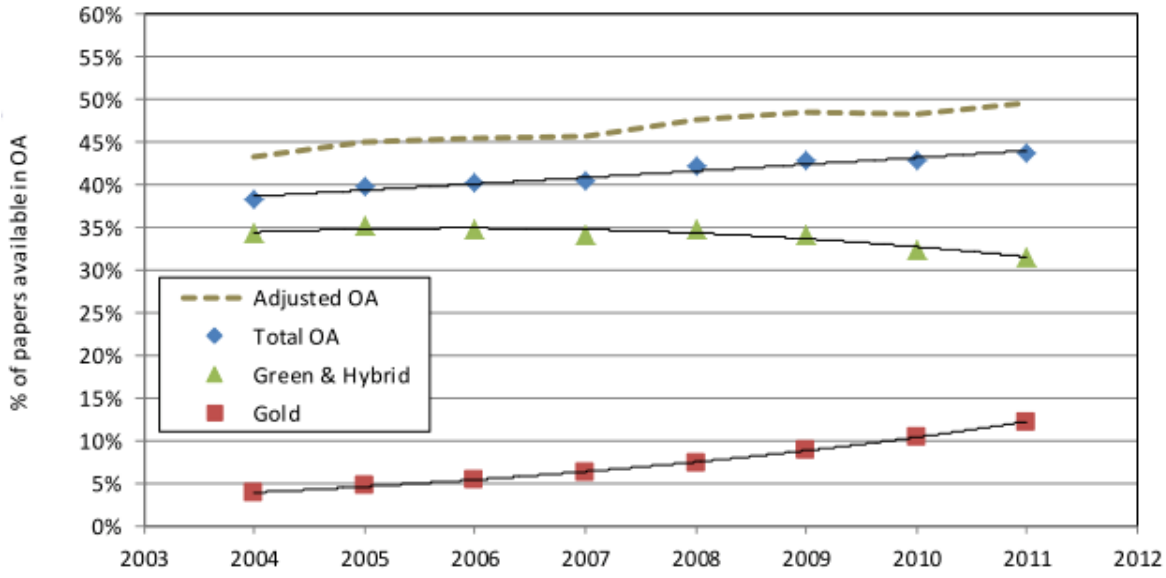
	Yes, to the general public	Yes, to other researchers	Yes, to industries	No	Don't know	Total 'Yes'
EU27	62%	33%	17%	10%	10%	79%
Sex						
Male	63%	34%	18%	11%	9%	81%
Female	61%	31%	16%	10%	12%	78%
Age						
15-24	67%	31%	17%	10%	7%	83%
25-39	66%	35%	18%	9%	7%	83%
40-54	63%	35%	18%	10%	8%	82%
55 +	57%	30%	16%	12%	16%	72%
Education (End of)						
15-	55%	28%	14%	11%	19%	70%
16-19	63%	32%	18%	10%	10%	80%
20+	65%	37%	18%	11%	6%	83%
Still studying	68%	34%	18%	9%	6%	85%
Interested in science						
Total 'Interested'	68%	37%	19%	9%	5%	85%
Total 'Not interested'	56%	28%	14%	12%	16%	72%
Informed about science						
Total 'Informed'	68%	38%	20%	10%	5%	85%
Total 'Not informed'	59%	30%	15%	11%	14%	75%
Studied science or tech						
Total 'Yes'	67%	36%	19%	9%	6%	85%
No	58%	30%	15%	11%	14%	74%
Influence of science on society						
Total 'Positive'	65%	35%	18%	10%	8%	82%
Total 'Negative'	57%	28%	18%	14%	9%	77%

Table 7.1.3: Potential indicator for OA, no. 2

Information Item	**OA2**
Name of indicator	Freely available peer-reviewed papers
Brief description	The indicator shows the proportion of freely available peer-reviewed papers in Scopus over 2004 – 2011. The indicator uses a method of calculation developed by Science-Metrix. The data can be disaggregated by scientific fields and countries (the data are available only for 2008 – 2011).
Analytical level (logic model)	Output-related
Analytical level (aggregation)	Global level, Country level, Field level
Qual / Quant	Quantitative
Source of data	Eric Archambault, Didier Amyot, Philippe Deschamps, Aurore Nicol, Lise Rebout & Guillaume Roberge Proportion of Open Access Peer-Reviewed Papers at the European and World Levels—2004-2011 August 2013
Date	2013
Time-series	2004 – 2011 (No for country-level and field-level data)
Measurement level	Ratio

Unit of analysis	Countries, fields (aggregated from individual article level primary data)
Coverage	28 Member States + other ERA Member States + USA, Japan, Canada and Brazil
Attributes	• Per cent of freely available peer-reviewed papers, 2004-2011 • Proportion of OA per field, 4-year non-weighted sampling, 2008-2011 • Number of papers indexed in Scopus available in OA, 2008-2011 • Proportion of OA per country, 4-year non-weighted sampling, 2008-2011

Table 7.1.4: Data presentation, OA2



Field	Papers	Green & Hybrid		Gold		OA		OA Growth	
		Papers	%	Papers	%	Papers	%	Trend	Index
Agriculture, Fisheries & Forestry	6,142	1,655	27 ± 1	1,033	17 ± 0.9	2,688	44 ± 1		1.13
Biology	7,031	2,749	39 ± 1	1,224	17 ± 0.9	3,973	57 ± 1		1.08
Biomedical Research	12,571	6,377	51 ± 0.9	1,346	11 ± 0.5	7,723	61 ± 0.8		1.04
Built Environment & Design	998	237	24 ± 3	32	3 ± 1	269	27 ± 3		0.99
Chemistry	13,399	2,642	20 ± 0.7	1,275	10 ± 0.5	3,917	29 ± 0.8		1.39
Clinical Medicine	42,806	15,479	36 ± 0.5	4,658	11 ± 0.3	20,137	47 ± 0.5		1.17
Communication & Textual Studies	1,108	168	15 ± 2	63	6 ± 1	231	21 ± 2		1.06
Earth & Environmental Sciences	4,700	1,771	38 ± 1	339	7 ± 0.7	2,110	45 ± 1		0.95
Economics & Business	3,950	1,583	40 ± 2	191	5 ± 0.7	1,774	45 ± 2		1.02
Enabling & Strategic Technologies	12,714	2,567	20 ± 0.7	1,061	8 ± 0.5	3,628	29 ± 0.8		1.11
Engineering	11,264	2,366	21 ± 0.8	269	2 ± 0.3	2,635	23 ± 0.8		1.03
General Arts, Humanities & Social Sciences*	8,220	2,250	27 ± 0.2	137	2 ± 0.0	2,387	29 ± 0.2		1.08
General Science & Technology	2,760	720	26 ± 2	1,036	38 ± 2	1,756	64 ± 2		1.28
Historical Studies	1,590	374	24 ± 2	124	8 ± 1	498	31 ± 2		0.99
Information & Communication Technologies	5,620	1,993	35 ± 1	584	10 ± 0.8	2,577	46 ± 1		1.01
Mathematics & Statistics	4,318	2,091	48 ± 1	310	7 ± 0.8	2,401	56 ± 1		0.95
Philosophy & Theology	732	170	23 ± 3	38	5 ± 2	208	28 ± 3		1.03
Physics & Astronomy	15,028	5,824	39 ± 0.8	698	5 ± 0.3	6,522	43 ± 0.8		0.98
Psychology & Cognitive Sciences	3,027	1,204	40 ± 2	172	6 ± 0.8	1,376	45 ± 2		1.10
Public Health & Health Services	4,688	1,640	35 ± 1	589	13 ± 0.9	2,229	48 ± 1		1.21
Social Sciences	5,170	1,374	27 ± 1	489	9 ± 0.8	1,863	36 ± 1		1.00
Visual & Performing Arts*	6,572	609	9 ± 0.1	231	4 ± 0.1	840	13 ± 0.1		1.00
Total*	**160,000**	**53,072**	**33 ± 0.2**	**15,538**	**10 ± 0.1**	**68,610**	**43 ± 0.2**		1.08

Notes: *In order to decrease the margin of error for the fields of General Arts, Humanities & Social Sciences, and Visual & Performing Arts, the Scopus population of papers were used instead of a sample. However, this does not affect the error margin for the total as this was based on sample data only.

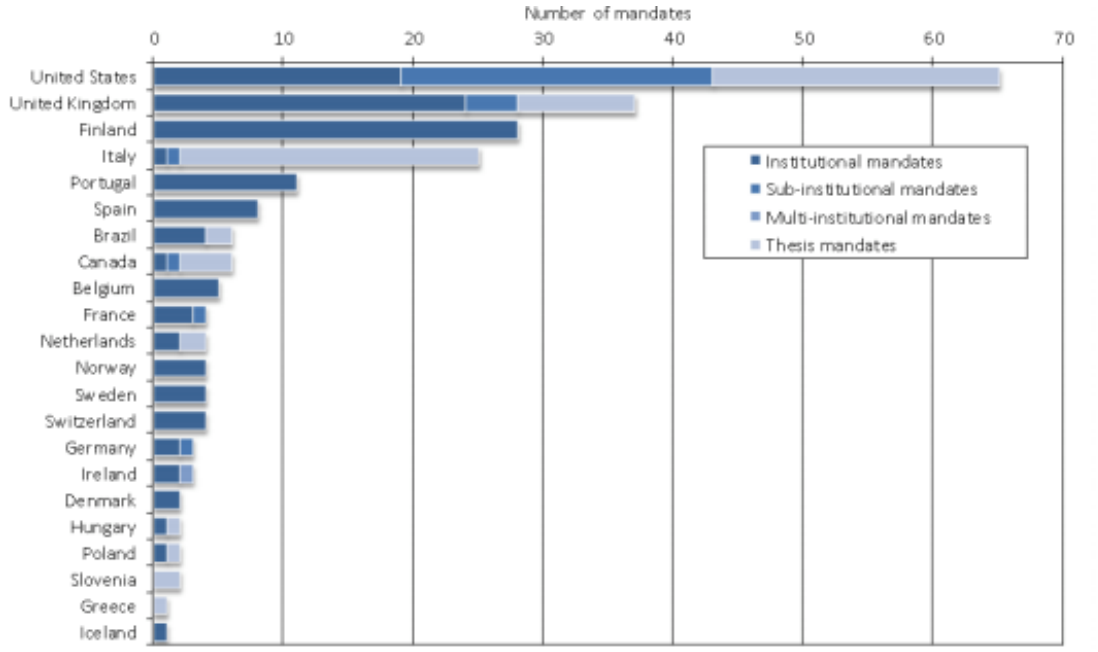
Group	Country	Papers in Sample	Green & Hybrid		Gold		OA	
			Papers	%	Papers	%	Papers	%
EU28	Austria	1,349	545	40 ± 3	105	8 ± 1	650	48 ± 3
	Belgium	2,088	939	45 ± 2	126	6 ± 1	1,065	51 ± 2
	Bulgaria	293	91	31 ± 5	21	7 ± 3	112	38 ± 6
	Croatia	403	147	36 ± 5	95	24 ± 4	242	60 ± 5
	Cyprus	88	35	40 ± 11	3	3 ± 4	38	43 ± 11
	Czech Republic	1,252	411	33 ± 3	130	10 ± 2	541	43 ± 3
	Denmark	1,392	633	45 ± 3	103	7 ± 1	736	53 ± 3
	Estonia	161	63	39 ± 8	23	14 ± 6	86	53 ± 8
	Finland	1,178	488	41 ± 3	80	7 ± 1	568	48 ± 3
	France	7,959	3,205	40 ± 1	472	6 ± 0.5	3,677	46 ± 1
	Germany	10,531	4,026	38 ± 0.9	673	6 ± 0.5	4,699	45 ± 0.9
	Greece	1,336	452	34 ± 3	116	9 ± 2	568	43 ± 3
	Hungary	709	279	39 ± 4	49	7 ± 2	328	46 ± 4
	Ireland	833	358	43 ± 3	75	9 ± 2	433	52 ± 3
	Italy	6,094	2,389	39 ± 1	421	7 ± 0.6	2,810	46 ± 1
	Latvia	71	23	32 ± 11	8	11 ± 8	31	44 ± 12
	Lithuania	228	88	39 ± 6	35	15 ± 5	123	54 ± 7
	Luxembourg	37	13	35 ± 17	2	5 ± 9	15	41 ± 17
	Malta	23	8	35 ± 21	4	17 ± 17	12	52 ± 22
	Netherlands	3,759	1,936	52 ± 2	263	7 ± 0.8	2,199	58 ± 2
	Poland	2,474	707	29 ± 2	326	13 ± 1	1,033	42 ± 2
	Portugal	1,047	479	46 ± 3	97	9 ± 2	576	55 ± 3
	Romania	734	289	39 ± 4	75	10 ± 2	364	50 ± 4
	Slovakia	368	127	35 ± 5	41	11 ± 3	168	46 ± 5
	Slovenia	358	110	31 ± 5	50	14 ± 4	160	45 ± 5
	Spain	5,461	2,074	38 ± 1	604	11 ± 0.8	2,678	49 ± 1
	Sweden	2,301	922	40 ± 2	181	8 ± 1	1,103	48 ± 2
	United Kingdom	11,781	5,100	43 ± 0.9	728	6 ± 0.4	5,828	49 ± 0.9
	Total EU28	53,622	20,204	37 ± 0.4	4,192	8 ± 0.2	24,396	45 ± 0.4
EFTA	Iceland	85	39	46 ± 11	3	4 ± 4	42	49 ± 11
	Liechtenstein	6	1	17 ± 38			1	17 ± 38
	Norway	1,159	494	43 ± 3	110	9 ± 2	604	52 ± 3
	Switzerland	2,642	1,214	46 ± 2	208	8 ± 1	1,422	54 ± 2
	Total EFTA	3,830	1,705	45 ± 2	319	8 ± 0.9	2,024	53 ± 2
Candidate	Turkey	2,873	657	23 ± 2	598	21 ± 1	1,255	44 ± 2
	Macedonia	39	15	38 ± 16	11	28 ± 15	26	67 ± 16
	Total Candidate	3,303	672	23 ± 2	608	21 ± 1	1,280	44 ± 2
	Israel	1,376	640	47 ± 3	92	7 ± 1	732	53 ± 3
Total ERA		59,852	22,085	37 ± 0.4	5,009	8 ± 0.2	27,094	45 ± 0.4
Others	United States	41,740	20,894	50 ± 0.5	2,535	6 ± 0.2	23,429	56 ± 0.5
	Japan	9,703	3,264	34 ± 0.9	804	8 ± 0.5	4,068	42 ± 1
	Canada	6,676	2,885	43 ± 1	411	6 ± 0.6	3,296	49 ± 1
	Brazil	4,224	876	21 ± 1	1,799	43 ± 1	2,675	63 ± 1
World		160,000	53,072	33 ± 0.2	15,538	10 ± 0.1	68,610	43 ± 0.2

Table 7.1.5: Potential indicator for OA, no. 3

Information Item	**OA3**
Name of indicator	Institutional perception of OA strategies
Brief description	The indicator is a set of questions to institutions' (universities and other research performing organisations) on their perception of open access strategies.
Analytical level (logic model)	Input-related / Output-related
Analytical level (aggregation)	Country level
Qual / Quant	Quantitative
Source of data	Julie Caruso, Aurore Nicol & Eric Archambault Open Access Strategies in the European Research Area August 2013

Date	2013
Time-series	No
Measurement level	Ratio
Unit of analysis	Countries (aggregated from institutional level primary data)
Coverage	EU28 + USA, Japan, Canada, Brazil
Attributes	• Institutional, multi-institutional, sub-institutional, and thesis mandates within the ERA and in selected countries • Types of repositories used to archive open access scholarly publications • Incentives used to promote open access archiving and publication of scholarly publications

Table 7.1.6: Data presentation, OA3



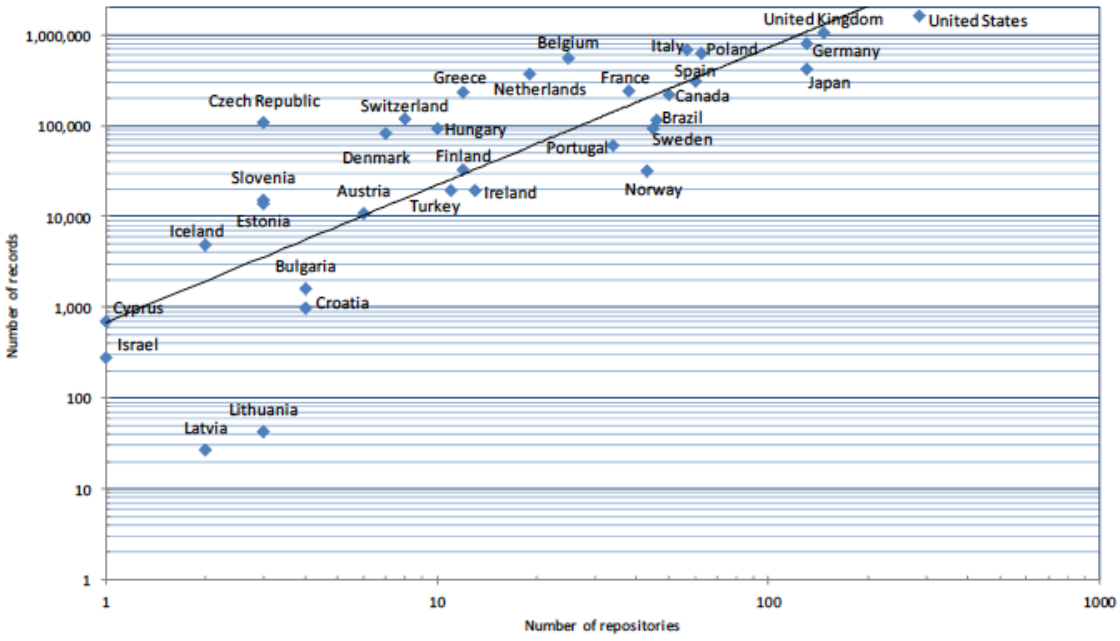
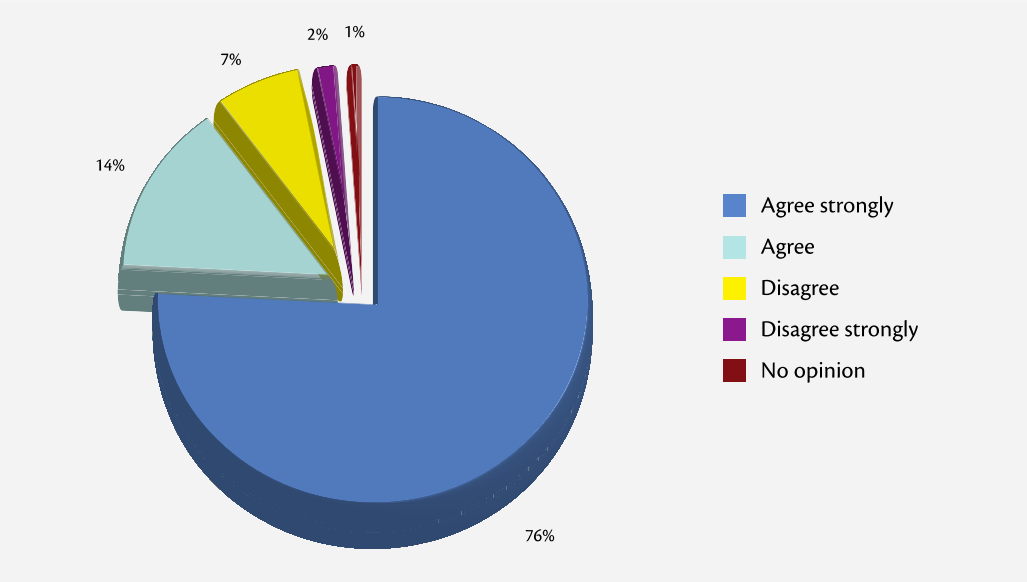


Table 7.1.7: Potential indicator for OA, no. 4

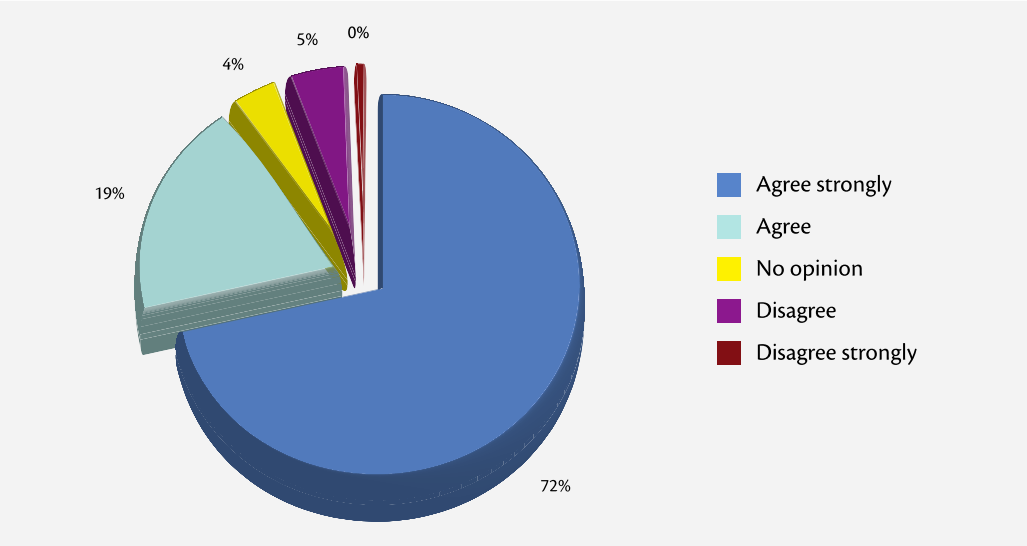
Information Item	OA4
Name of indicator	Stakeholders' perception of access to digital resources
Brief description	The indicator is formed by a set of questions to various stakeholders (national governments, regional and local governments, research funding organisations, university/research institutes, libraries, publishers, international organisations, individual researchers, citizens and respondents identified as 'other', among which there were NGOs, industries, charities, learned societies and scientific and professional associations) on their perception of access to digital resources.
Analytical level (logic model)	Context-related
Analytical level (aggregation)	Country level
Qual / Quant	Quantitative
Source of data	European Commission. Directorate-General for Research and Innovation. Report on the online survey on scientific information in the digital age held from July-September 2011.
Date	2012
Time-series	No
Measurement level	Ordinal
Unit of analysis	Countries (aggregated from individual level primary data)
Coverage	28 Member States
Attributes	• Should publications resulting from publicly funded research be available OA • Does OA increase access to and dissemination of scientific publications • Can OA coexist with the traditional publication system • Preferred way in which public policy can increase OA to scientific publications

Table 7.1.8: Data presentation, OA4

Do you think that publications resulting from publicly funded research should, as a matter of principle, be available free of charge to readers on the Internet (i.e. open access mode)?



Do you think that open access can increase access to and dissemination of scientific publications?



Do you think that open access to scientific publications can coexist with the traditional scientific publication system?

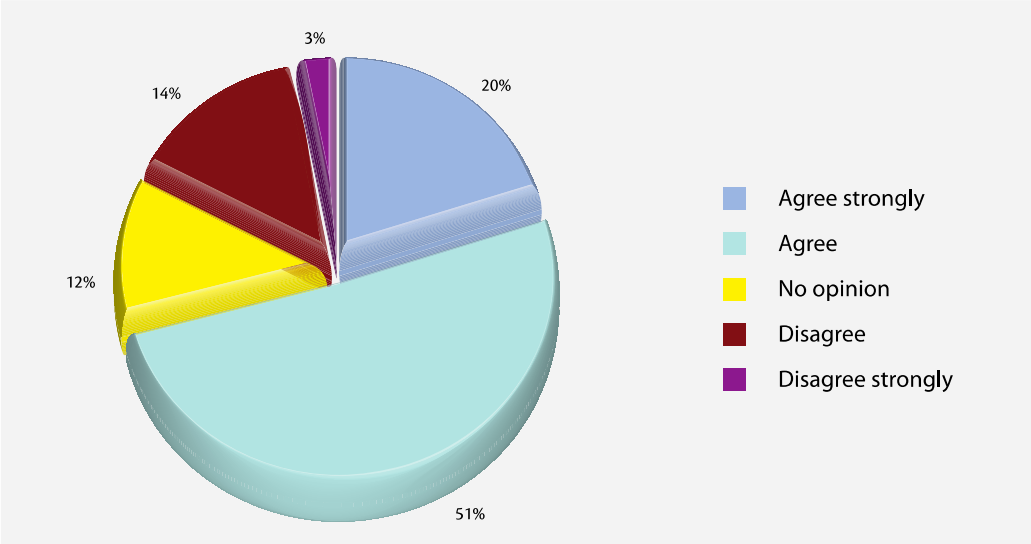


Table 7.1.9: Potential indicator for OA, no. 5

Information Item	OA5
Name of indicator	FP7 project coordinators' perception of self-archiving
Brief description	The indicator represents a set of questions on the issue of self-archiving.
Analytical level (logic model)	Output-related
Analytical level (aggregation)	EU-level
Qual / Quant	Quantitative
Source of data	European Commission. Directorate-General for Research and Innovation. Survey on open access in FP7.
Date	2012
Time-series	No
Measurement level	Ordinal

Unit of analysis	Individual level primary data
Coverage	States participating in FP7

Attributes	• Opinion on the implementation of Special Clause 39
	  o Having time/manpower to self-archive
	  o Getting enough external support (e.g. toolkits)
	  o Identifying a new, satisfactory publisher (journal)
	  o Changing publishers/journals
	  o Negotiating with the publishers/journals
	• Articles deposited in a repository
	• Open AIRE

Table 7.1.10: Data presentation, OA5

Having time/manpower to self-archive (i.e. deposit in a repository)		
	Number of requested answers	**% of total answers (194)**
1 (very difficult)	11	5.67
2 (difficult)	40	20.62
3 (easy)	63	32.47
4 (very easy)	19	9.79
N/A (no opinion)	61	31.44

Getting enough external support (e.g. toolkits)		
	Number of requested answers	**% of total answers (194)**
1 (very difficult)	7	3.61
2 (difficult)	32	16.49
3 (easy)	39	20.10
4 (very easy)	9	4.64
N/A (no opinion)	107	55.15

Identifying a new, satisfactory publisher (journal)		
	Number of requested answers	**% of total answers (194)**
1 (very difficult)	15	7.73
2 (difficult)	52	26.80
3 (easy)	37	19.07
4 (very easy)	8	4.12
N/A (no opinion)	82	42.27

Changing publishers/journals		
	Number of requested answers	**% Requested answers (194)**
1 (very difficult)	19	9.79
2 (difficult)	46	23.71
3 (easy)	32	16.49
4 (very easy)	3	1.55
N/A (no opinion)	94	48.45

Negotiating with the publishers/journals		
	Number of requested answers	**% of total answers (194)**
1 (very difficult)	24	12.37
2 (difficult)	48	24.74
3 (easy)	25	12.89
4 (very easy)	1	0.52
N/A (no opinion)	96	49.48

Among the articles resulting from your project (already published or accepted for publication), how many of these are deposited or will be deposited in a repository (online archive), regardless of whether or not they are Open Access?		
	Number of requested answers	**% of total number answers (194)**
No article published/accepted yet	118	60.82
0 deposited	23	11.86
1 deposited	11	5.67
2 deposited	7	3.61
3 deposited	5	2.58
4 deposited	4	2.06
5 deposited	4	2.06
More	22	11.34

How did you discover the project OpenAIRE?		
	Number of requested answers	**% of total number answers (194)**
I had never heard about it before it was mentioned in this questionnaire!	115	59.28
CORDIS website	22	11.34
EU-related event	21	10.82
Word of mouth/colleague/friend etc.	17	8.76
European Commission project officer	14	7.22
European Commission leaflets/publications	7	3.61
Conferences	7	3.61
General research ('Google') on the Internet	7	3.61
Science in Society website	5	2.58
National contact point (NCP)	5	2.58
Other	5	2.58
Participants' portal	3	1.55
Articles/journals	3	1.55

Table 7.1.11: Potential indicator for OA, no. 6

Information Item	**OA6**
Name of indicator	FP7 project coordinators' perception of open-access publishing
Brief description	The indicator represents a set of questions on the issue of open-access publishing.
Analytical level (logic model)	Output-related
Analytical level (aggregation)	EU-level

Qual / Quant	Quantitative
Source of data	European Commission. Directorate-General for Research and Innovation. Survey on open access in FP7.
Date	2012
Time-series	No
Measurement level	Ordinal
Unit of analysis	Individual level primary data
Coverage	States participating in FP7
Attributes	• Knowledge of the possibility of reimbursement • Use of reimbursement of open access publishing • Future use of reimbursement of open access publishing • Views on open access publishing

Table 7.1.12: Data presentation, OA6

Article II.16.4 of FP7 Model Grant Agreement allows for 100% reimbursement of publication costs (including Open Access publishing) during the lifespan of all FP7 projects. Did you know this before replying to the survey?

	Number of requested answers	% Requested answers (194)
Yes	93	47.94
No	101	52.06

Have you used this possibility in your project so far?

	Number of requested answers	% Requested answers (93)
Yes	8	8.60
No	85	91.40

Do you intend to make use of the possibility of Article II.16.4 of the FP7 model grant agreement in the future?

	Number of requested answers	% Requested answers (194)
Yes	85	43.81
Maybe/not sure	106	54.64
No	3	1.55

It is better to use self-archiving to satisfy the requirements of Special Clause 39.

	Number of requested answers	% Requested answers (194)
1 (strongly disagree)	8	4.12
2 (disagree)	33	17.01
3 (agree)	59	30.41
4 (strongly agree)	22	11.34
N/A (no opinion)	72	37.11

Table 7.1.13: Potential indicator for OA, no. 7

Information Item	**OA7**
Name of indicator	Open Data Barometer
Brief description	The indicator is a composite index indicator showing to what extent countries make date open to various socioeconomic groups. It combines peer-reviewed expert survey data and secondary indicators to look at open data readiness, implementation and emerging impacts.
Analytical level (logic model)	Context-related
Analytical level (aggregation)	Country-level
Qual / Quant	Quantitative

Source of data	Tim Davies et al. Open Data Barometer. 2013 Global Report. World Wide Web Foundation
Date	2013
Time-series	No
Measurement level	Ordinal / Ratio
Unit of analysis	Country
Coverage	Global (77 countries)
Attributes	• Readiness sub-index • Implementation sub-index • Impact sub-index • Open Data Barometer Overall

Table 7.1.14: Data presentation, OA7

Country	Rank	Readiness Sub-Index	Implementation Sub-Index	Impact Sub-Index	ODB Overall
United Kingdom	1	100.00	100.00	79.91	**100.00**
United States	2	95.26	86.67	100.00	**93.38**
Sweden	3	95.20	83.14	71.95	**85.75**
New Zealand	4	81.88	65.49	89.81	**74.34**
Norway	5	91.88	70.98	46.15	**71.86**
Denmark	5	83.54	70.20	55.73	**71.78**
Australia	7	87.88	64.71	51.19	**67.68**
Canada	8	79.11	63.92	51.59	**65.87**
Germany	9	74.50	63.14	53.81	**65.01**
France	10	79.39	64.31	39.07	**63.92**
Netherlands	10	85.92	67.06	21.42	**63.66**
Korea (Rep. of)	12	77.19	54.90	24.56	**54.21**
Iceland	13	62.99	52.94	26.45	**51.01**
Estonia	14	72.38	49.41	24.00	**49.45**
Finland	14	91.19	41.18	40.87	**49.44**
Japan	14	76.99	47.06	27.94	**49.17**
Spain	17	67.48	49.41	21.13	**48.19**
Austria	18	68.56	39.22	48.62	**46.03**
Israel	18	61.82	45.88	25.36	**45.58**
Italy	20	50.39	42.75	45.69	**45.30**
Russia	20	54.43	40.39	48.86	**44.79**
Switzerland	22	65.11	41.57	26.80	**43.24**
Czech Republic	22	61.83	40.00	35.36	**43.18**
Kenya	22	49.70	45.88	21.55	**43.06**
Mexico	25	49.10	45.49	8.37	**40.30**
Chile	25	65.79	39.22	18.27	**40.11**
Portugal	27	60.38	38.04	19.25	**38.63**
Brazil	28	66.03	32.16	27.87	**36.83**
Singapore	29	70.28	35.29	8.97	**36.29**
Ireland	29	61.81	32.55	23.92	**35.76**
Thailand	31	38.09	39.22	14.88	**35.33**
Argentina	31	46.08	36.47	17.29	**35.00**
Belgium	31	72.01	28.63	25.64	**34.80**
India	34	57.36	33.73	9.87	**33.38**
Uruguay	34	54.66	32.94	13.31	**33.04**
Costa Rica	36	47.34	35.29	0.00	**31.21**
Kazakhstan	37	34.96	32.16	2.84	**27.61**
Greece	37	43.95	27.84	12.30	**27.59**
Turkey	37	41.92	31.37	0.00	**27.58**
Morocco	40	36.46	27.84	16.59	**27.24**
Colombia	40	44.33	29.02	2.49	**26.71**
Hungary	42	32.42	28.63	10.51	**26.09**
Mauritius	42	35.71	30.59	0.00	**26.08**
United Arab Emirates	44	53.88	21.57	12.30	**24.59**
Rwanda	45	36.71	27.84	0.00	**24.27**
Jamaica	46	32.56	25.88	2.49	**22.69**
Philippines	47	40.33	21.18	10.31	**21.91**
Peru	47	36.36	23.14	4.95	**21.74**
Ghana	47	39.51	23.53	0.00	**21.60**
Ecuador	50	38.51	22.35	2.83	**21.12**
Tunisia	50	63.52	10.98	26.46	**21.02**
South Africa	52	35.39	18.43	10.31	**19.20**
Indonesia	52	34.91	20.39	0.00	**18.66**
Bahrain	54	42.94	18.04	0.00	**18.18**
Uganda	55	23.99	13.33	23.07	**16.15**
Botswana	55	12.16	21.57	0.00	**16.08**
Nepal	55	21.15	18.43	2.84	**15.70**
Tanzania	58	20.43	17.65	0.00	**14.51**
Malawi	59	28.24	11.76	16.52	**14.47**
Qatar	60	39.01	11.76	0.00	**13.09**
China	61	41.72	9.41	0.00	**11.82**
Venezuela	62	9.59	14.90	0.00	**10.91**
Pakistan	63	14.59	12.16	0.00	**9.70**
Jordan	63	30.95	8.63	0.00	**9.63**
Bangladesh	63	23.00	10.20	0.00	**9.56**
Ethiopia	66	15.45	10.59	0.00	**8.70**
Burkina Faso	67	17.63	8.24	0.00	**7.35**
Benin	67	11.60	9.41	0.00	**7.28**
Saudi Arabia	67	40.82	1.57	5.69	**7.09**
Namibia	67	11.57	9.02	0.00	**7.00**
Senegal	71	28.57	4.71	0.00	**6.46**
Cameroon	71	7.11	6.67	5.56	**5.65**
Zimbabwe	73	15.20	5.88	0.00	**5.30**
Yemen	73	0.00	8.24	0.00	**4.69**
Nigeria	75	36.90	0.00	0.00	**4.35**
Zambia	75	11.84	5.10	0.00	**4.23**
Mali	77	6.15	0.39	0.00	**0.00**
Grand Total		**47.99**	**32.20**	**18.62**	**32.47**

Table 7.1.15: Potential indicator for OA, no. 8

Information Item	OA8
Name of indicator	Existing funder mandates for open access publishing
Brief description	The indicator presents if and how many funder mandates for open access publishing there are in the EU Member States.
Analytical level (logic model)	Input-related
Analytical level (aggregation)	Country-level
Qual / Quant	Quantitative
Source of data	Commission Staff Working Document Impact Assessment Accompanying the document Commission Recommendation on access to and preservation of scientific information in the digital age {C(2012) 4890 final} {SWD(2012) 221 final} based on openaire.eu
Date	2011
Time-series	No
Measurement level	Ratio
Unit of analysis	Country
Coverage	Member States (excluding Croatia)
Attributes	• Number of funder mandates (if applicable)

Table 7.1.16: Potential indicator for OA, no. 9

Information Item	OA9
Name of indicator	Number of open access journals in 2011
Brief description	The indicator how many open access journals there are in the EU Member States (as of 2011).
Analytical level (logic model)	Output-related
Analytical level (aggregation)	Country-level
Qual / Quant	Quantitative
Source of data	Commission Staff Working Document Impact Assessment Accompanying the document Commission Recommendation on access to and preservation of scientific information in the digital age {C(2012) 4890 final} {SWD(2012) 221 final} based on www.doaj.org
Date	2011
Time-series	No
Measurement level	Ratio
Unit of analysis	Country
Coverage	Member States (excluding Croatia)
Attributes	• Number of open access journals as of 2011

Table 7.1.17: Potential indicator for OA, no. 10

Information Item	OA10
Name of indicator	Number of open access repositories
Brief description	The indicator presents how many open access repositories there are in the EU Member States.
Analytical level (logic model)	Output-related
Analytical level (aggregation)	Country-level
Qual / Quant	Quantitative
Source of data	Commission Staff Working Document Impact Assessment Accompanying the document Commission Recommendation on access to and preservation of scientific information

	in the digital age {C(2012) 4890 final} {SWD(2012) 221 final} based on www.opendoar.org
Date	2011
Time-series	No
Measurement level	Ratio
Unit of analysis	Country
Coverage	Member States (excluding Croatia)
Attributes	• Number of open access repositories

Table 7.1.18: Data presentation, OA8, OA9, OA10

Member State	**Existing funder mandates for OA publishing**[198]	**No. of OA journals in 2011**[199]	**No. of OA repositories**[200]
Austria	1	38	9
Belgium	1	23	29
Bulgaria	None	31	5
Cyprus	None	4	1
Czech Republic	None	54	6
Denmark	1	31	10
Estonia	None	20	5
Finland	None[203]	38	15
France	1	138	62
Germany	4	242	151
Greece	none	37	14
Hungary	1	20	11
Ireland	4	9	14
Italy	1	195	69
Latvia	1[204]	2	3

Lithuania	1[205]	24	3
Luxembourg	None	1	--
Malta	None	2	--
Netherlands	None	51	24
Poland	None	133	75
Portugal	None	57	40
Romania	None	217	1
Slovak Republic	None	26	--
Slovenia	1	33	3
Spain	4	400	83
Sweden	5	51	45
UK	15	530	203

Table footnotes:

198. According to openaire.eu (retrieved on 20 Dec 2011).
199. According to www.doaj.org (retrieved on 20 Dec 2011).
200. According to www.opendoar.org (retrieved on 20 Dec 2011).
201. Based on the total federal or central government budget appropriations or outlays on R&D, excluding EU funding, as per 2008 (2007 for Greece and Portugal), as reported in http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-32-10-225/EN/KS-32-10-225-EN.PDF.

Table 7.1.19: Potential indicator for OA, no 11,

Information Item	**OA11**
Name of indicator	Metric model of data publishing
Brief description	Indicators based on the data publication models aimed to measure the presence of data publications. Two dimensions of metrics are proposed: size dependent (these are metrics that capture the raw performance, in terms of data outputs) and size independent indicators(metrics capturing relative performance of outputs and units of analysis, e.g. through means or medians; and they can also capture an indicator on the publication venues, e.g. data journals or data repositories)
Analytical level (logic model)	Output-related
Analytical level (aggregation)	Multi-level (e.g. data publication level, data creator, institutional indicators, countries, publication venues, etc.)
Qual / Quant	Quantitative
Source of data	Costas et al (2013)
Date	2013
Time-series	Possible
Measurement level	Total counts, ratios, medians, and other statistics (e.g. percentiles,

	correlations, etc.)
Unit of analysis	Multi-level (data publications, individuals, groups, institutions, geographical entities, data venues, etc.)
Coverage	Global (but dependent on the coverage of the different repositories and sources providing the metrics)
Attributes	• Number of open data repositories

Table 7.1.20: Potential indicator for OA, no 12,

Information Item	**OA12**
Name of indicator	Metric model of data usage
Brief description	Indicators based on the data publication models aimed to measure the usage of data publications (e.g. citations). Two dimensions of metrics are proposed: size dependent (these are metrics that capture the raw performance, in terms of data citations) and size independent indicators (metrics capturing relative performance of citations and units of analysis, e.g. through means or medians;
Analytical level (logic model)	Outcome related
Analytical level (aggregation)	Multi-level (e.g. data publication level, data creator, institutional indicators, countries, publication venues, etc.)
Qual / Quant	Quantitative
Source of data	Costas et al (2013)
Date	2013
Time-series	Possible
Measurement level	Total counts, ratios, medians, and other statistics (e.g. percentiles, correlations, etc.)
Unit of analysis	Multi-level (data citations, individuals, groups, institutions, geographical entities, data venues, etc.)
Coverage	Global (but dependent on the coverage of the different repositories and sources providing the metrics)
Attributes	• Number of open data repositories

8. Ideas bank

The purpose of this chapter is to provide an outlet for collecting ideas, notes and thoughts on the design of indicators and in particular with regard to the subsequent analysis of RRI benefits. This chapter will not form part of the final deliverable but can be considering a working document or ideas bank for the ensuing analytical work.

As maybe lost in the length of a dimension report, two issues are of relevance for OA indicators:

1. Difficulties in aligning production of and citations to OA publications with the classical bibliometric analysis methods. This is particularly relevant in the transition phase to more open access.
2. Getting a better idea of Research data management plan requirements at the Member Stae level. Policies to this end could serve as a proxy to measure more open data activities (publishing and citations).

9. References

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., & Moorman, D. (2004). Promoting Access to Public Research Data for Scientific, Economic, and Social Development. Data Science Journal, 3(November), 135–152.3(November), 135–152.

Borgman, C. L. (2012). The Conundrum of Sharing Research Data. Journal of the American Society for Information Science and Technology, 63(6), 1059–1078. doi:10.1002/asi

Brase, J., Farquhar, A., Gastl, A., Gruttemeier, H., & Heijne, M. (2009). Approach for a joint global registration agency for research data. Inforamtion Services & Use, 29, 13–27. doi:10.3233/ISU-2009-0595

Costas, R., Meijer, I., Zahedi, Z., & Wouters, P. (2013). The value of research data - Metrics for datasets from a cultural and technical point of view. A Knowledge Exchange Report (pp. 1–48).

Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 368(1926), 4023–38. doi:10.1098/rsta.2010.0165

Dallmeier-Tiessen, S., Darby, R., Gitmans, K., Lambert, S., Suhonen, J., Wilson, M., … Coordination, A. (2012). Compilation of results on drivers and barriers and new opportunities.

Fienberg, S. E., Martin, M. E., & Straf, M. L. (1985). Sharing Research Data. Washington: National Academy Press.

Groves, T. (2010). The wider concept of data sharing : view from the BMJ. Biostatistics, 11(3), 391–392. doi:10.1136/bmj.b3928.G

Kaye, J., Heeney, C., Hawkins, N., & Vries, J. De. (2009). Data sharing in genomics — re-shaping scientific practice. Nature Reviews, 10(May), 331–335.

Knoppers, B. M., Harris, J. R., Tassé, A. M., Budin-ljøsne, I., Kaye, J., & Deschênes, M. (2011). Towards a data sharing Code of Conduct for international genomic research. GenomeMedicine, 3, 46.

Krumholz, H. M. (2012). Open science and data sharing in clinical research: basing informed decisions on the totality of the evidence. Circulation. Cardiovascular quality and outcomes, 5(2), 141–2. doi:10.1161/CIRCOUTCOMES.112.965848

Piwowar, H. A. (2011). Who shares? Who doesn't? Factors associated with openly archiving raw research data. PloS one, 6(7), e18657. doi:10.1371/journal.pone.0018657

Piwowar, H. A., Becich, M. J., Bilofsky, H., & Crowley, R. S. (2008). Towards a data sharing culture: recommendations for leadership from academic health centers. PLoS medicine, 5(9), e183. doi:10.1371/journal.pmed.0050183

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. PloS one, 2(3), e308. doi:10.1371/journal.pone.0000308

Robinson-García, N., Jiménez-Contreras, E., & Torres-Salinas, D. (2015). Analyzing data citation practices according to the Data Citation Index. Journal of the Associatoin for Information Science and Technology, 1–20.

Ross, J. S., & Krumholz, H. M. (2013). Ushering in a new era of open science through data sharing: the wall must come down. JAMA, 309(13), 1355–6. doi:10.1001/jama.2013.1299

Ross, J. S., Lehman, R., & Gross, C. P. (2012). The importance of clinical trial data sharing: toward more open science. Circulation. Cardiovascular quality and outcomes, 5(2), 238–40. doi:10.1161/CIRCOUTCOMES.112.965798

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., … Frame, M. (2011). Data sharing by scientists: practices and perceptions. PloS one, 6(6), e21101. doi:10.1371/journal.pone.0021101
Torres-Salinas, D., Martín-Martín, A., & Fuente-Gutiérrez, E. (2014). Analysis of the coverage of the Data Citation Index – Thomson Reuters : disciplines , document types and repositories. Revista Española de Documentación Científica, 37(1), 1–6.
Van den Eynden, V., & Bishop, L. (2014). Sowing the seed: Incentives and motivations for sharing research data, a researcher's perspective (p. 48).
Archambault, Eric, Didier Amyot, Philippe Deschamps, Aurore Nicol, Françoise Provencher, Lise Rebout, and Guillaume Roberge. 2014. "Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels—1996–2013*." Rapport, Commission Européenne DG Recherche & Innovation; RTD-B6-PP-2011-2: Study to Develop a Set of Indicators to Measure Open Access.*

Björk, Bo-Christer, Patrik Welling, Mikael Laakso, Peter Majlender, Turid Hedlund, and Guðni Guðnason. 2010. "Open Access to the Scientific Journal Literature: Situation 2009." *PLoS ONE* 5 (6): e11273. doi:10.1371/journal.pone.0011273.

Björk, Bo-Christer. 2012. "The Hybrid Model for Open Access Publication of Scholarly Articles: A Failed Experiment?" *Journal of the American Society for Information Science and Technology* 63 (8): 1496–1504. doi:10.1002/asi.22709.

Björk, Bo-Christer, Mikael Laakso, Patrik Welling, and Patrik Paetau. 2014. "Anatomy of Green Open Access." *Journal of the Association for Information Science and Technology* 65 (2): 237–50. doi:10.1002/asi.22963.

BOAI. 2002. "Budapest Open Access Initiative." *The Open Society Foundations*. http://www.opensocietyfoundations.org/openaccess.

Carpenter, Todd. 2013. "Progress Toward Open Access Metadata." *Serials Review* 39 (1): 1–2. doi:10.1016/j.serrev.2013.02.001.

Craig, Iain D., Andrew M. Plume, Marie E. McVeigh, James Pringle, and Mayur Amin. 2007. "Do Open Access Articles Have Greater Citation Impact?: A Critical Review of the Literature." *Journal of Informetrics*, The Hirsch Index, 1 (3): 239–48. doi:10.1016/j.joi.2007.04.001.

Davis, Philip M., and Matthew J. L. Connolly. 2007. "Institutional Repositories: Evaluating the Reasons for Non-use of Cornell University's Installation of DSpace", March. http://hdl.handle.net/1813/5195.

Davis, P. M, B. V Lewenstein, D. H Simon, J. G Booth, and M. J L Connolly. 2008. "Open Access Publishing, Article Downloads, and Citations: Randomised Controlled Trial." *BMJ* 337 (jul31 1): a568–a568. doi:10.1136/bmj.a568.

Eysenbach, Gunther. 2006. "Citation Advantage of Open Access Articles." *PLoS Biol* 4 (5): e157. doi:10.1371/journal.pbio.0040157.

Gargouri, Yassine, Vincent Larivière, Yves Gingras, and Stevan Harnad. 2012. "Green and Gold Open Access Percentages and Growth, by Discipline." In *Proceedings of 17th International Conference on Science and Technology Indicators*, edited by Eric Archambault, Yves Gingras, and Vincent Larivière. Montréal: Science-Metrix and OST.

Kurtz, Michael J., Guenther Eichhorn, Alberto Accomazzi, Carolyn Grant, Markus Demleitner, Edwin Henneken, and Stephen S. Murray. 2005. "The Effect of Use and Access on Citations." *Information Processing & Management*, Special Issue on Infometrics, 41 (6): 1395–1402. doi:10.1016/j.ipm.2005.03.010.

Laakso, Mikael, and Bo-Christer Björk. 2013. "Delayed Open Access: An Overlooked High-impact Category of Openly Available Scientific Literature." *Journal of the*

American Society for Information Science and Technology 64 (7): 1323–29. doi:10.1002/asi.22856.

Lee, Kyungjoon, John S. Brownstein, Richard G. Mills, and Isaac S. Kohane. 2010. "Does Collocation Inform the Impact of Collaboration?" *PLoS ONE* 5 (12): e14279. doi:10.1371/journal.pone.0014279.

Moed, Henk F. 2007. "The Effect of 'open Access' on Citation Impact: An Analysis of ArXiv's Condensed Matter Section." *Journal of the American Society for Information Science and Technology* 58 (13): 2047–54. doi:10.1002/asi.20663.

NISO. 2015. *Access License and Indicators NISO RP-22-2015*. National Information Standards Organization. ISBN: 978-1-937522-49-0

Swan, Alma. 2010. "The Open Access Citation Advantage: Studies and Results to Date". Technical Report. http://eprints.ecs.soton.ac.uk/18516/.

Xia, Jingfeng, Rebekah Lynette Myers, and Sara Kay Wilhoite. 2011. "Multiple Open Access Availability and Citation Impact." *Journal of Information Science* 37 (1):

van Leeuwen, Thed, Clifford Tatum, and Paul Wouters. 2015. "Open Access publishing and citation impact - an international study". 5th INTERNATIONAL CONFERENCE ON SCIENTOMETRICS and INFORMETRICS. Istanbul, July 2015

10. Appendix – literature review

10.1 Review guidelines

MoRRI

Final version / 17.11.2014 (rl)

Task 1: Literature review | Review template

Background and objectives

The purpose of this template is to provide each member of the review team with a common framework and reference point to conduct the literature review and, one the reviews are conducted, to facilitate a systematic and structured analysis of the literature.

According to the TOR, the main objective of this first task in the MoRRI project is to

- review of the state of knowledge regarding RRI
- define the policy context of RRI in Europe and elsewhere
- give a comparative assessment of RRI dimensions, weighing-up advantages, disadvantages and available options

- conduct a preliminary assessment of the availability of empirical evidence on the dimensions
- finalise the definitions and properties of the RRI key dimensions
- finalise the definition and properties of additional factors that may be relevant for the monitoring tasks.

How to use this document

- Due to the standardized nature of this template, you may feel that the content of the literature cannot be adequately represented. In these cases, please use the comment spaces provided for most questions.
- The literature review takes into account a selection of relevant publications in the 5 key dimensions of RRI (as defined by the EC: citizen engagement, science literacy, gender equality, open access, governance and ethics) and a selection of key publications dealing explicitly with RRI. Some of the questions in this template only relate to the 5 key dimensions, others only to the explicit RRI literature. Please make sure to fill in the template accordingly.
- Try to briefly summarise the relevant statements of the review document in your own words, perhaps using bullet points; please always refer to the page number of the document.
- If a question in the template does not apply to the publication at hand, please leave the entry blank.
- Important definitions or other central statements may be copied into the template; please always make reference to the page number of the review document
- Given the diversity of literature covered in this review, it is difficult to provide guidance on how extensive each review should be. For a normal journal article we expect the filled-in template to count roughly about 8-10 pages.

If you have any questions, please get in touch:

Ralf Lindner, ph.: +49 (0) 721 / 6809-292

ralf.lindner@isi.fraunhofer.de

10.2 Review reports

Policy Reports

Basic information		**Document no.:** (citavi #)	**000**
Reviewer's name	XP		
1. Bibliographical information (author/s, year, title, editor/s, journal/book, volume, publisher, place of publication, pages, DOI)	Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Towards better access to scientific information: Boosting the benefits of public investments in research. COM(2012) 401 final. European Commission, 7.7.2012		
2. Abstract	The Europe 2020 strategy for a smart, sustainable and inclusive economy underlines the		

(copy and paste)	central role of knowledge and innovation in generating growth. Research results, including both publications and data collections, need to be circulated rapidly and widely, using digital media. This accelerates scientific discovery, enables new forms of data-intensive research and allows research findings to be systematically taken up by European business and industry. To spur scientific and technological progress, the European Union (EU) should review its policies and practices on disseminating scientific information, and take the necessary steps to improve access to the results of publicly-funded scientific research. This Communication sets out the action that the Commission intends to take to improve access to scientific information and to boost the benefits of public investment in research. It also explains how open access policies will be implemented under 'Horizon 2020', the EU's Framework Programme for Research and Innovation (2014-2020). The Communication is accompanied by a Recommendation to the Member States, calling for improved policies and practices on access and preservation in the Member States. This initiative springs from two mutually-reinforcing policy strands. One is the Digital Agenda for Europe, which sets out an 'open data' policy covering the full range of information that public bodies across the European Union produce, collect or pay for. The other is the Innovation Union Communication, which outlines the EU's research and innovation policies and programmes. The proposed measures build on earlier work, in particular the 2007 Communication on scientific information in the digital age and the related Council Conclusions, the 2009 Communication on ICT infrastructures for e-Science and the strategic policy developed for the European Research Area (ERA). To improve access to scientific information, Member States, research funding bodies, researchers, scientific publishers, universities and their libraries, innovative industries, and society at large need to work together. Europe's scientific information system must be made fit for the digital age so that the 'fifth freedom' of the EU — the free circulation of knowledge — can become a reality.							
3. Main focus (key dimensions according to MoRRI)	RRI / RI	☐	Citizen participation	☐	Science literacy	☐	Gender equality	☐
	Open access	☒	R&I governance and ethics	☐	Other	☐		
Comment on 3:	Also focuses on Open Data and, more generally, on the rationale and benefits of free circulation of knowledge							
4. Main perspective (multiple entries possible)	Theoretical, conceptual	☐	Methodological	☐	Policy oriented	☒	Evaluative	☐
	Other	☐	Comment on 4:					
5. Type of document	Scientific article	☐	Book chapter	☐	Book	☐	Report	☐
	Project deliverable	☐	Policy/ strategy document	☒	Other	☐		
Comment on 5:								
6. System level (if applicable)	Global	☐	European	☒	National	☐	Sub-national	☐
Comment on 6:								
7.1 Country focus (if applicable, please specify)	European Union							
7.2 Country/ies of origin indicated by					Comments on 7:			

institutional affiliation of editor(s)/ author(s) (if applicable, please specify)		

Data and indicator availability				
8.1 Data, indicators, measurements	Document contains data	☒	If yes, please specify (including page numbers in document)	Section 4, p 4-5 Conclusion and targets, p 10-12
Comment on 8.1	Some figures and data backing general claims, mostly at EU-level but not disaggregated per country.			
8.2 Reference made to data, indicators measurements in other sources	Document refers to relevant sources	☐	If yes, please list source(s): (URLs, data banks, reports, statistics, etc.)	
Comment on 8.2:				
Guiding questions for review *- please add page numbers where appropriate -*				
9. How is RRI characterized? (For literature dealing explicitly with responsible (research) and innovation. If the publication deals with one of the 5 key dimensions, please proceed to 11.)				
9.1 Which definition of RRI is being used? (author's definition or reference to other source)				
9.2 Which aspects of RRI receive special emphasis? (e.g., certain normative goals, procedural approaches, reference to one or more of the 5 key dimensions, ...)				
9.2 Which arguments are presented in support or rejection/criticism of RRI?				

Basic information		**Document no.:** (citavi #)	**000**
Reviewer's name	XP		
1. Bibliographical information (author/s, year, title, editor/s, journal/book, volume, publisher, place of publication, pages, DOI)	Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Towards a thriving data-driven economy. COM(2014) 442 final. European Commission, 2.7.2014.		
2. Abstract (copy and paste)	The European Council's conclusions of October 2013 focused on the digital economy, innovation and services as drivers for growth and jobs. They called for EU action to provide the right framework conditions for a single market for big data and cloud computing. This Communication responds by sketching the features of the data-driven economy of the future and setting out some operational conclusions to support and accelerate the transition towards it. It also sets out current and future activities in the field of cloud computing.		

	This Communication builds on the results of various consultations and on relevant legislative proposals already tabled, such as on reform of the EU rules on the protection of personal data and on network and information security.							
3. Main focus (key dimensions according to MoRRI)	RRI / RI	☐	Citizen participation	☐	Science literacy	☐	Gender equality	☐
	Open access	☒	R&I governance and ethics	☐	Other	☐		
Comment on 3:	Mostly related to the framework conditions needed to exploit the full potential of data, and the policy challenges and next-steps at EU-level.							
4. Main perspective (multiple entries possible)	Theoretical, conceptual	☐	Methodological	☐	Policy oriented	☒	Evaluative	☐
	Other	☐	Comment on 4:					
5. Type of document	Scientific article	☐	Book chapter	☐	Book	☐	Report	☐
	Project deliverable	☐	Policy/ strategy document	☒	Other	☐		
Comment on 5:								
6. System level (if applicable)	Global	☐	European	☒	National	☐	Sub-national	☐
Comment on 6:								
7.1 Country focus (if applicable, please specify)	European Union							
7.2 Country/ies of origin indicated by institutional affiliation of editor(s)/ author(s) (if applicable, please specify)					Comments on 7:			

Data and indicator availability				
8.1 Data, indicators, measurements	Document contains data	☐	If yes, please specify (including page numbers in document)	
Comment on 8.1				
8.2 Reference made to data, indicators measurements in other sources	Document refers to relevant sources	☐	If yes, please list source(s): (URLs, data banks, reports, statistics, etc.)	
Comment on 8.2:				
Guiding questions for review *- please add page numbers where appropriate -*				
9. How is RRI characterized? (For literature dealing explicitly with responsible (research) and innovation. If the publication deals with one of the 5 key dimensions, please proceed to 11.)				
9.1 Which definition of RRI is being used? (author's definition or reference to other source)				
9.2 Which aspects of RRI receive				

special emphasis? (e.g., certain normative goals, procedural approaches, reference to one or more of the 5 key dimensions, ...)	
9.2 Which arguments are presented in support or rejection/criticism of RRI?	

Basic information								**Document no.:** (citavi #)	**000**
Reviewer's name	BB								
1. Bibliographical information (author/s, year, title, editor/s, journal/book, volume, publisher, place of publication, pages, DOI)	The Finch Report (2012): *Accessibility, sustainability, excellence: how to expand access to research publications June. Online at:* www.researchinfonet.org/publish/finch/								
2. Abstract (copy and paste)	The Report was initiated by the Working Group on Expanding Access to Published Research Findings, an independent group chaired by Dame Janet Finch. The group emerged from transparency discussions initiated at a high level round table meeting, in March 2011, by the David Willetts, UK-Minister for Universities and Science. It was made up of representatives of the University sector, research funders, the research community, scholarly publishers, and libraries. Dame Janet Finch is a Professor of Sociology at Manchester University and independent co-Chair of the Council for Science and Technology. The draft of the report was compiled by The Research Information Network (RIN), a London-based community interest company and in July 2012, the UK government accepted the recommendations of the Working Group. The report recommends a balanced programme of action to enable more people to read and use the publications arising from research, and to accelerate the progress towards a fully open access environment. Following the description of RIN at www.researchinfonet.org/publish/finch/, the report made clear that several different channels for communicating research results will remain important over the next few years. But it recommended a clear policy direction in the UK towards support for 'Gold' open access publishing, where publishers receive their revenues from authors rather than readers, and so research articles become freely accessible to everyone immediately upon publication. At the same time, the report recommended extensions to current licensing arrangements in the higher education, health and other sectors; improvements to the infrastructure of repositories, and support for the moves by publishers to provide access to the great majority of journals in public libraries.								
3. Main focus (key dimensions according to MoRRI)	RRI / RI	☐	Citizen participation	☐	Science literacy	☐	Gender equality	☐	
	Open access	☒	R&I governance	☐	Other	☐			

Comment on 3:			and ethics					
4. Main perspective (multiple entries possible)	Theoretical, conceptual	☐	Methodological	☐	Policy oriented	☒	Evaluative	☐
	Other	☐	Comment on 4:					
5. Type of document	Scientific article	☐	Book chapter	☐	Book	☐	Report	☐
	Project deliverable	☐	Policy/ strategy document	☒	Other	☐		
Comment on 5:								
6. System level (if applicable)	Global	☐	European	☒	National	☐	Sub-national	☐
Comment on 6:								
7.1 Country focus (if applicable, please specify)	UK							
7.2 Country/ies of origin indicated by institutional affiliation of editor(s)/ author(s) (if applicable, please specify)	UK				Comments on 7:			

Data and indicator availability				
8.1 Data, indicators, measurements	Document contains data	☐	If yes, please specify (including page numbers in document)	
Comment on 8.1				
8.2 Reference made to data, indicators measurements in other sources	Document refers to relevant sources	☐	If yes, please list source(s): (URLs, data banks, reports, statistics, etc.)	
Comment on 8.2:				
Guiding questions for review *- please add page numbers where appropriate -*				
9. How is RRI characterized? (For literature dealing explicitly with responsible (research) and innovation. If the publication deals with one of the 5 key dimensions, please proceed to 11.)				
9.1 Which definition of RRI is being				

used? (author's definition or reference to other source)	
9.2 Which aspects of RRI receive special emphasis? (e.g., certain normative goals, procedural approaches, reference to one or more of the 5 key dimensions, …)	
9.2 Which arguments are presented in support or rejection/criticism of RRI?	

Basic information								**Document no.:** (citavi #)	**000**
Reviewer's name	Annette Braun								
1. Bibliographical information (author/s, year, title, editor/s, journal/book, volume, publisher, place of publication, pages, DOI)	Tim Davies, 2013, Open Data – Barometer, 3013 Global Report, World Wide Web Foundation and Open Data Institute, 35 p,								
2. Abstract (copy and paste)	Open data is still in its infancy. Less than five years after the first major Open Government Data (OGD) portal went live, hundreds of national and local governments have established OGD portals, joined by international institutions, NGOs and businesses. All are exploring, in different ways, how opening data can unlock latent value, stimulate innovation and increase transparency and accountability. Against this backdrop of rapid growth of the open data field, this Open Data Barometer global report provides a snapshot of OGD practices at national level. It also outlines a country-bycountry ranking. Covering a broad sample of 77 countries, it combines peer-reviewed expert survey data and secondary indicators to look at open data readiness, implementation and emerging impacts.								
3. Main focus (key dimensions according to MoRRI)	RRI / RI	☐	Citizen participation	☐	Science literacy	☐	Gender equality	☐	
	Open access	☒	R&I governance and ethics	☐	Other	☐			
Comment on 3:									

4. Main perspective (multiple entries possible)	Theoretical, conceptual	☐	Methodological	☐	Policy oriented	☐	Evaluative	☒
	Other	☐	Comment on 4:					
5. Type of document	Scientific article	☐	Book chapter	☐	Book	☐	Report	☒
	Project deliverable	☐	Policy/ strategy document	☐	Other	☐		
Comment on 5:								
6. System level (if applicable)	Global	☒	European	☐	National	☐	Sub-national	☐
Comment on 6:								
7.1 Country focus (if applicable, please specify)								
7.2 Country/ies of origin indicated by institutional affiliation of editor(s)/ author(s) (if applicable, please specify)	UK				Comments on 7:			

Data and indicator availability				
8.1 Data, indicators, measurements	Document contains data	☒	If yes, please specify (including page numbers in document)	Source of data: www.opendatabarometer.org 6: 55% of the analyzed countries have OGD policies 11: 25% have dedicated resources and senior level political backing of OGD 6: Just 7% of the researched **datasets** come up to really advanced OGD standards. 12: Table: Country development level related to OGD implementation levels. 18: Comprehensive assessment of different countries regarding their OGD performance, in relation to different kinds of datasets
Comment on 8.1				
8.2 Reference made to data, indicators measurements in other sources	Document refers to relevant sources	☐	If yes, please list source(s): (URLs, data banks, reports, statistics, etc.)	
Comment on 8.2:				

Guiding questions for review *- please add page numbers where appropriate -*	
11. Claims regarding the effects of RRI and / or the key dimension (benefits, costs, disadvantages, trade-offs)	
11.1 What claims are being made?	The Open Data – Barometer is first of all a comprehensive study about the state of the art regarding Open Governmental Data (OGD) in 77 countries. Thus, explicit claims are raised only very seldom. But, the performance assessment allows deducing the implicit best practice claims. 6, 14: As truly Open Data (OD) is not common yet, potential entrepreneurs reject from raising their business based on this data due to legal uncertainty. 7: The open data vision is a bold one: but one that will take considerable work to make a reality. It cannot just be a case of ad-hoc dataset publication, but needs attention paid to legal, social, economic, technical, organization and political dimensions of open data publication and re-use. 6: Expectation concerning Open Government Data (OGD) are: - to establish a latent value - to stimulate innovation - to increase transparency 6: OGD policies have quite different levels of implementation quality. Some are just insolated portals and some are already quite comprehensive. Most comprehensive concepts comprise of machine-readable data sets and use open licenses. UK ranks highest regarding OGD implementation. 8: Beyond pure dataset publication, legal, societal, economical, technical, organizational and political dimensions as well as re-use questions have to be considered. 9: The democratic, social and economic potential of OGD is still far from being exploited. Actually, there is the risk of a falling backwards, due to missing political support. 9: There is no one-size-fits-all approach for OGD concepts. 9: The Open Data Barometer – from a meta perspective – follows the policy of: - transparent methodology - allowing others to use the data - making re-mix and re-interpretation possible 9/ 12: Target group of OGD is first of all the governments themselves that they can use their own data in a better way. Public services should be improved and the needs of the citizens should be met. Also, the operational efficiency can be increased. 9: OGD can be seen as a kind of an economic raw material in the information economy. New products and services can be developed upon it.

	9: OGD can be used for democratic control of the government and increase transparency. 10: Marginalized groups can be empowered by OGD due to better information equilibrium. 10: OGD concerns all kind of non-personal data governments collect. 11: The assessment criteria for OGD-implementation are: • Is there a Right to Information (RTI) law established? • Do central government/ regions/ cities have an OGD initiative? • Is there a demand from civil society for OGD? • Is a technology community for OGD established? • Are there governmental support measures for OGD re-use (training, innovation events, grants and voucher schemes)? 12: There is a strong relation between levels of development and the diffusion of OGD policies and practice in the analyzed countries. 12: A really advanced OGD policy provides intermediaries to exploit the economic potential of the data. It is crucial to have not only the data available but to create a whole eco-system for data exploitation. A broad range of actors are necessary for a successful implementation. 12: Beside Right to Information, Data Protection regimes are a crucial precondition for OGD policy. The citizens have to be protected from data abuse. 13: In Africa the access to the Internet is still under-developed. Data should be provided also via print media, community radio and mobile phones. 13: Even though in Middle East and Central Asia, the technical capacity for OGD is given, the civil rights are still lagging behind. 14: The quality of OGD show different levels: 1. Provision of online data 2. Data in machine readable, open formats 3. Accessible, standardized and linked datasets 14: Even though, many countries have developed OGD policy, the implementation still lags behind. 14: OGD should be published under a clear open license statement, in general clear license terms, and should be available for bulk download. 15: By far the most analyzed OGD sets were still in a very early, user-unfriendly stage. 15: Provided data does not necessarily meet the demand of the citizens and contribute to more transparency, accountability, innovation and greater inclusion. 15: OGD can be e.g. :

	• Census data • Trade data • National budgets • Public spending (with the opportunity to track detailed transactions) • Land registry • Company registry 15: Until today, standardization of data formats is almost limited to public transport data. 15: It would be good to use format standards in order to compare statistics, financial information, and company information across borders. 16: Open standards should be aimed at in the future. 17: Quality of data and their trustworthiness is essential. Insufficient data should be withdrawn. 17: If companies use OGD for their business they require frequent updates and timely access. 17: Beside the pure provision of data, the scope of the dataset is relevant. Partially missing or very generic data might be useless. 17: Comprehensiveness of data is crucial. 17: Discoverability of data is an issue. 17: The implementation of OGD concepts depends not only on top-down decisions, but particularly on contextually aware manpower. 24: The political will and the technical capacity are decisive for the implementation of OGD. 18: Ranking of OGD performance: 1. UK 2. USA 3. Sweden 4. Norway 5. Denmark 6. Netherlands 7. New Zeland 8. Australia 9. France 10. Canada 11. Germany 12. South Korea 13. …. 21: Dimensions of Open Data impact, with ranking: 1. Transparency & accountability 2. Entrepreneurial open data use 3. Government efficiency 4. Economic growth 5. Environmental sustainability

	6. Inclusion of marginalized groups
11.2 Which arguments are used to support the claim(s)?	
11.3 What evidence is presented to support the claims? (e.g., data, indicators, research results, case studies, anecdotal evidence)	
11.4 According to the author(s), which type of evidence/data is missing to better support the claim? (e.g. data gaps, limitations with regard to analytical levels, lack of indicator specifications etc.)	15: It should be analyzed which kind of dataset contribute most to certain kinds of impacts in different contexts, and how the technical features of those datasets affect their use. 21: The benefit of OGD should be quantitatively analyzed. 21: Sectoral capacity building around OGD should be researched.
Comments on 11.	
12. Key dimensions of RRI (For literature dealing with one or more of the 5 key dimensions.)	
12.1 How is the key dimension defined? (terminology applied, central features/characteristics)	
12.2 Does the document reach beyond one single dimension / are more than one of the key dimensions discussed? If yes, what is the proposed relationship between different dimensions (complementary, contradictory…)?	
12.3 To which concepts, theories, approaches, schools of thought, communities (scientific or practice) in the area of research and innovation does the literature relate or make reference to? (e.g., STS, constructive TA, anticipatory governance, foresight, deliberative democracy, …)	
Comments on 12.	
13. Are other important dimensions / aspects of RRI discussed, presented which are so far not covered by MoRRI?	

14. Anything else deemed relevant?	
15. General comments and remarks	
16. Relevant sources cited (Please list references to other sources cited in the literature which seem to be highly relevant for MoRRI and/or represent important contributions in the field)	

Basic information		**Document no.:** (citavi #)	**000**
Reviewer's name	BB		
1. Bibliographical information (author/s, year, title, editor/s, journal/book, volume, publisher, place of publication, pages, DOI)	Van den Eynden, V. and Bishop, L. (2014). Sowing the Seed: Incentives and motivations for sharing research data, a researcher's perspective. A Knowledge Exchange Report. (http://www.knowledge-exchange.info/Default.aspx?ID=733)		
2. Abstract (copy and paste)	This 38 pages paper presents the results of a study in which 22 selected researchers from different academic disciplines were interviewed concerning data sharing practices and attitudes towards open data. From the interviews, five case studies were compiled. The study was commissioned by Knowledge Exchange (KE) which is – according to its website a co-operative effort that supports the use and development of Information and Communications Technologies (ICT) infrastructure for higher education and research. The Knowledge Exchange partners are: - CSC - IT Center for Science in Finland - Denmark's Electronic Research Library (DEFF) in Denmark - German Research Foundation (DFG) in Germany - Jisc in the United Kingdom and - SURF in the Netherlands. The partners work together on activities in the field of Open Access, Research Data, Research Tools and Technologies and Interoperability Standards. See (http://www.knowledge-exchange.info). The overall objective of the study was to provide evidence and examples of useful incentives for data sharing from the researchers' point of view to inform scientists and policy makers. The five case studies span various academic disciplines: arts and humanities, social sciences, biomedicine, chemistry and biology. A major output was to find out, what motivates researchers to share data. The study says that the incentives that currently motivate interviewed researchers to share their research data, fall within three main categories: 1. Direct benefits (for the research itself (more robust), for the career of the researcher (recognition), for discipline (get wiser), and for science (better science). 2. Norms of the project, research group, and/or discipline 3. External drivers: policies and expectations from research funders and publishers The report analyses current practices in the covered disciplines and concludes with		

	recommendations for research funders, learned societies, research institutions, publishers, data centres and repositories and finally for the Knowledge Exchange-organisations, the organization which commissioned and funded the report. The report concludes that leadership is needed from various actors for research data sharing to become the norm in a wider range of disciplines, institutions and research groups. (p. 35)							
3. Main focus (key dimensions according to MoRRI)	RRI / RI	☐	Citizen participation	☐	Science literacy	☐	Gender equality	☐
	Open access	☒	R&I governance and ethics	☐	Other	☒	Open Data	
Comment on 3:								
4. Main perspective (multiple entries possible)	Theoretical, conceptual	☐	Methodological	☐	Policy oriented	☒	Evaluative	☐
	Other	☐	Comment on 4:	For the report, interviews were carried out in order to identify incentives for researchers to share their research data. The addressees of the paper are research funders, publishers, data centers and repositories and educational institutes.				
5. Type of document	Scientific article	☐	Book chapter	☐	Book	☐	Report	☒
	Project deliverable	☐	Policy/ strategy document	☐	Other	☐		
Comment on 5:								
6. System level (if applicable)	Global	☐	European	☒	National	☐	Sub-national	☐
Comment on 6:	Since the open data movement is global, the results of the survey and the report in general are relevant for all countries. However, the focus is on Europe.							
7.1 Country focus (if applicable, please specify)								
7.2 Country/ies of origin indicated by institutional affiliation of editor(s)/ author(s) (if applicable, please specify)	See above. Countries in which researchers were interviewed are: Netherlands, Germany, Denmark, Finland and the UK.				Comments on 7:			
Data and indicator availability								
8.1 Data, indicators, measurements	Document contains data	☐	If yes, please specify (including page numbers in document)					
Comment on 8.1								
8.2 Reference made to data, indicators	Document refers to relevant	☒	If yes, please list source(s): (URLs, data banks, reports, statistics, etc.)	The report lists relevant reports and studies covering open data in chapter 1 Background from page 10.				

measurements in other sources	sources			
Comment on 8.2:				

Guiding questions for review *- please add page numbers where appropriate -*	
9. How is RRI characterized? (For literature dealing explicitly with responsible (research) and innovation. If the publication deals with one of the 5 key dimensions, please proceed to 11.)	
9.1 Which definition of RRI is being used? (author's definition or reference to other source)	
9.2 Which aspects of RRI receive special emphasis? (e.g., certain normative goals, procedural approaches, reference to one or more of the 5 key dimensions, …)	
9.2 Which arguments are presented in support or rejection/criticism of RRI?	

Basic information		**Document no.:** (citavi #)	**000**
Reviewer's name	BB		
1. Bibliographical information (author/s, year, title, editor/s, journal/book, volume, publisher, place of publication, pages, DOI)	**Genova, F. et al. (2014) The Data Harvest. How sharing research data can yield knowledge, jobs and growth. A Special Report by RDA Europe.** (https://europe.rd-alliance.org/documents/publications-reports/data-harvest-how-sharing-research-data-can-yield-knowledge-jobs-and)		
2. Abstract (copy and paste)	RDA means Research Data Alliance, which is a EU-funded organization that, with counterparts in USA and Australia, is working to speed the path for scientific data sharing around the world. According to its Website, the RDA was built on the collaboration between data practitioners with a vision of openly sharing data across technologies, disciplines, and countries to address the grand challenges of society. Scientists, researchers and innovators join forces with technical experts in focused Working Groups and exploratory Interest Groups. Membership is free and open to all (http://europe.rd-alliance.org/). In this 38-page paper, the authors mainly address policy makers in Europe to support the idea of open data. The paper is more like a brochure in which Open Data is being explained. Examples of European data sharing projects are listed and described in detail (COPERNICUS, CLARIN, The Human Brain project and others). It is claimed that open data will repreat what the Internet did in the last decades to enhance science, transparency, accessibility to		

<table>
<tr><td></td><td colspan="8">information and individual information. Open data will fundamentally change the way science is conducted, who does it, who pays for it and who benefits from it. And most importantly, the rising capacity to share all this data –electronically, efficiently, across borders and disciplines – magnifies the impact (p. 2).

In the main chapter Why bother? the perceived benefits of open data are described in detail. These are:
Benefit 1: Creating jobs, spurring growth
Benefit 2: Boosting research productivity and creativity
Benefit 3: Helping people, engaging citizens
In the final chapter The way ahead the authors propose five dos and two don´ts for politicians in order to support the open data movement:
1. DO require a data plan, and show it is being implemented.
2. DO promote data literacy across society, from researcher to citizen.
3. DO develop incentives and grants for data sharing (and don't forget Horizon 2020).
4. DO develop tools and policies to build trust and data-sharing.
5. DO support international collaboration.
6. DON'T regulate what we don't yet understand.
7. DON'T stop what has begun well.</td></tr>
<tr><td rowspan="3">3. Main focus
(key dimensions according to MoRRI)

Comment on 3:</td><td>RRI / RI</td><td>☐</td><td>Citizen participation</td><td>☐</td><td>Science literacy</td><td>☐</td><td>Gender equality</td><td>☐</td></tr>
<tr><td>Open access</td><td>☒</td><td>R&I governance and ethics</td><td>☐</td><td>Other</td><td>☒</td><td colspan="2">Open Data</td></tr>
<tr><td colspan="8">Open Data is seen as the next step in the open access development</td></tr>
<tr><td rowspan="2">4. Main perspective
(multiple entries possible)</td><td>Theoretical, conceptual</td><td>☐</td><td>Methodological</td><td>☐</td><td>Policy oriented</td><td>☒</td><td>Evaluative</td><td>☐</td></tr>
<tr><td>Other</td><td>☐</td><td>Comment on 4:</td><td colspan="5">Arguments are being displayed on why it is important to support open data and a policy agenda is being developed for European leaders.</td></tr>
<tr><td rowspan="3">5. Type of document

Comment on 5:</td><td>Scientific article</td><td>☐</td><td>Book chapter</td><td>☐</td><td>Book</td><td>☐</td><td>Report</td><td>☐</td></tr>
<tr><td>Project deliverable</td><td>☐</td><td>Policy/ strategy document</td><td>☒</td><td>Other</td><td>☐</td><td colspan="2"></td></tr>
<tr><td colspan="8">It is a policy paper which was written by the EU-financed Research Data Alliance-Europe in order to inform about benefits and requirements of open data.</td></tr>
<tr><td rowspan="2">6. System level
(if applicable)

Comment on 6:</td><td>Global</td><td>☒</td><td>European</td><td>☒</td><td>National</td><td>☐</td><td>Sub-national</td><td>☐</td></tr>
<tr><td colspan="8">Focus is on Europe and it is being stated that Europe has a good starting point in the movement for open data. However, if not backed actively by politicians, open data will benefit other nations. In principle, however, as the open data movement is global, it also has to be addressed globally.</td></tr>
<tr><td>7.1 Country focus
(if applicable, please specify)</td><td colspan="8">See above</td></tr>
<tr><td>7.2 Country/ies of origin indicated by</td><td colspan="4">The RDA initiative is a European initiative. In fact, it is a Coordinated and Support Action (CSA) in FP7, consisting of 12 academic</td><td colspan="4">Comments on 7:</td></tr>
</table>

institutional affiliation of editor(s)/ author(s) (if applicable, please specify)	partners organizing the platform and the exchange activities. The partners are: CSC, Finnish IT Center for Science (Fi), Association of Commonwealth Universities (UK), Athena Research & Innovation Centre (Gr), BSC (Es), CINECA (It), CNR, National Research Council of Italy (It), Centre National de la Recerche Scientifique (CNRS) (Fr), EPCC, University of Edinburgh (UK), Maastricht University (NL), Max Planck Society for the Advancement of Science (D), Trust-IT Services Ltd (UK), Science and Technology Facilities Council (UK).	

Data and indicator availability				
8.1 Data, indicators, measurements	Document contains data	☐	If yes, please specify (including page numbers in document)	
Comment on 8.1				
8.2 Reference made to data, indicators measurements in other sources	Document refers to relevant sources	☒	If yes, please list source(s): (URLs, data banks, reports, statistics, etc.)	The report lists the scientific and economic benefits of open data and reports of studies which calculate the benefit of - the ease of access and of finding the right data. This means users make efficiency gains of more than £100 million per annum. For every £1 spent on the Economic and Social Data Service, there is a value to the economy of £5.40 (p. 21). - of the US Human Genome Project. The project included an international agreement to put sequence data in the public domain. A study showed that the $3.8 billion that the US government invested in the project yielded $796 billion in economic output (p. 22). A study by McKinsey consultants tried to summarise the impact of open, shared data – not just in science but across the economy. It suggested that seven sectors – education, transportation, consumer products, electricity, oil and gas, health care, and consumer finance – could generate more than $3 trillion a year in additional value as a result of open data. (p. 24)
Comment on 8.2:	In order to illustrate the societal benefits of open data, the report describes the Everyaware-project in detail: Citizen-science projects are spreading across Europe, based on shared data. One EU-funded project, Everyaware, is building on the fact that lowcost sensing technologies now allow citizens to collect environmental data, and			

	blend these with their personal, subjective perceptions. For example, all smartphones have microphones that can record noise pollution; many also have thermometers for recording temperatures. Social networking tools can help collect and distribute data from thousands of smartphones, to be analysed, interpreted and visualised. Everyaware aims to pull these elements together in a single technology platform for conducting environmental surveys and analysing the results. Can this strengthen democracy? Hopefully: Citizens not only gain greater insight into what is being done in their name, but they can also look at the data themselves and suggest policy improvements. Politicians, take note: It will change your business forever. (p. 26)

Guiding questions for review
- please add page numbers where appropriate -

9. How is RRI characterized? (For literature dealing explicitly with responsible (research) and innovation. If the publication deals with one of the 5 key dimensions, please proceed to 11.)	
9.1 Which definition of RRI is being used? (author's definition or reference to other source)	
9.2 Which aspects of RRI receive special emphasis? (e.g., certain normative goals, procedural approaches, reference to one or more of the 5 key dimensions, ...)	
9.2 Which arguments are presented in support or rejection/criticism of RRI?	

Basic information		**Document no.:** (citavi #)	**000**
Reviewer's name	BB		
1. Bibliographical information (author/s, year, title, editor/s, journal/book, volume, publisher, place of publication, pages, DOI)	**Archambault, E. et al. (2014) Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels—1996–2013, Rapport, Commission Européenne DG Recherche & Innovation.** (http://science-metrix.com/files/science-metrix/publications/d_1.8_sm_ec_dg-rtd_proportion_oa_1996-2013_v11p.pdf		

<table>
<tr><td>2. Abstract
(copy and paste)</td><td colspan="8">This study measures the open access publications and their citation rates for the time period 1996 until 2013. According to the authors, the report is the “largest scale measurement of open access availability performed to date: a sample of one-quarter of a million records was used to study the historical evolution of open access (OA) between 1996 and 2013 and a larger, one million records sample was
used to perform an in-depth assessment of the proportion and scientific impact of OA in different types of OA, for different scientific fields of knowledge, and for 44 countries, the EU28, ERA, and the world” (p. i)
The 40-pages-study was carried out by Science-Metrix, a research evaluation firm based in Montreal, Canada with offices in the USA and in Brussels. The study was commissioned by the European Commission and is the latest in a series of reports produced by Science-Metrix under a research contract with the aim to develop a set of indicators to measure open access activities and effects.
The main results of the study are:
- OA is on the rise in all scientific disciplines
- All the fields derive an OA citation advantage.
In the policy recommendations, the authors clearly advocate models, in which authors or research institutions pay for their own publication (Front End Paid Access) and not subscribers (Back End Paid Access).</td></tr>
<tr><td rowspan="3">3. Main focus
(key dimensions according to MoRRI)

Comment on 3:</td><td>RRI / RI</td><td>☐</td><td>Citizen participation</td><td>☐</td><td>Science literacy</td><td>☐</td><td>Gender equality</td><td>☐</td></tr>
<tr><td>Open access</td><td>☒</td><td>R&I governance and ethics</td><td>☐</td><td>Other</td><td>☐</td><td colspan="2"></td></tr>
<tr><td colspan="8">The topic is measuring the use of open access in the different disciplines and tracking the citations from OA publications as compared to traditional publications.</td></tr>
<tr><td rowspan="2">4. Main perspective

(multiple entries possible)</td><td>Theoretical, conceptual</td><td>☐</td><td>Methodological</td><td>☐</td><td>Policy oriented</td><td>☐</td><td>Evaluative</td><td>☒</td></tr>
<tr><td>Other</td><td>☐</td><td colspan="6">Comment on 4:</td></tr>
<tr><td rowspan="3">5. Type of document

Comment on 5:</td><td>Scientific article</td><td>☐</td><td>Book chapter</td><td>☐</td><td>Book</td><td>☐</td><td>Report</td><td>☒</td></tr>
<tr><td>Project deliverable</td><td>☒</td><td>Policy/ strategy document</td><td>☐</td><td>Other</td><td>☐</td><td colspan="2"></td></tr>
<tr><td colspan="8"></td></tr>
<tr><td rowspan="2">6. System level
(if applicable)
Comment on 6:</td><td>Global</td><td>☒</td><td>European</td><td>☐</td><td>National</td><td>☐</td><td>Sub-national</td><td>☐</td></tr>
<tr><td colspan="8"></td></tr>
<tr><td>7.1 Country focus
(if applicable, please specify)</td><td colspan="8"></td></tr>
<tr><td>7.2 Country/ies of origin indicated by institutional affiliation of editor(s)/ author(s)
(if applicable, please specify)</td><td colspan="4">The authors are Canadian and the study was commissioned by the EC. The data analysed for the report comprises all publications listed in the Scopus database in which researchers from all over the world publish their scientific results. However, the analysis was concentrated on EU28 countries, ERA associated countries and others, whereas “others” comprise of the United States, Canada, Japan and Brasil.</td><td colspan="4">Comments on 7:</td></tr>
</table>

Data and indicator availability				
8.1 Data, indicators, measurements	Document contains data	☒	If yes, please specify (including page numbers in document)	The appendix contains tables with detailed numbers of papers published under the Green model, the Gold model, "Other" OA models and "total OA" for the time period 2008-2013 for all EU28 countries, in ERA associated countries and others (United States, Canada, Japan and Brasil) in the following disciplines: Agriculture, Fisheries & Forestry Biology Biomedical Research Built Environment & Design Chemistry Clinical Medicine Communication & Textual Studies Earth & Environmental Sciences Economics & Business Enabling & Strategic Technologies Engineering Gen. Arts, Humanities & Social Sciences Gen. Science & Technology Historical Studies Information & Communication Tech Mathematics & Statistics Philosophy & Theology Physics & Astronomy Psychology & Cognitive Sciences Public Health & Health Services Social Sciences Visual & Performing Arts
Comment on 8.1				
8.2 Reference made to data, indicators measurements in other sources	Document refers to relevant sources	☒	If yes, please list source(s): (URLs, data banks, reports, statistics, etc.)	
Comment on 8.2:	Especially in the discussion of the citation advantage section (p. 19ff) there are many sources given which are relevant in this field.			

Guiding questions for review *- please add page numbers where appropriate -*	
9. How is RRI characterized? (For literature dealing explicitly with responsible (research) and innovation. If the publication deals with one of the 5 key dimensions, please proceed to 11.)	
9.1 Which definition of RRI is being used? (author's definition or reference to other source)	

9.2 Which aspects of RRI receive special emphasis? (e.g., certain normative goals, procedural approaches, reference to one or more of the 5 key dimensions, ...)	
9.2 Which arguments are presented in support or rejection/criticism of RRI?	

Basic information							**Document no.:** (citavi #)	**000**
Reviewer's name	BB							
1. Bibliographical information (author/s, year, title, editor/s, journal/book, volume, publisher, place of publication, pages, DOI)	**Dallmeier-Tiessen, S. et al. (2011) Highlights from the SOAP Project Survey. What Scientists Think About Open Access Publishing.** (http://arxiv.org/abs/1101.5260)							
2. Abstract (copy and paste)	This 14-pages-paper contains the summary of a research project about attitudes towards Open Access in the science community. The project was called SOAP: Study of Open Access Publishing. In the center of the project was a large-scale survey on the attitudes of researchers on, and the experiences with, open access publishing. Around 40.000 answers were collected across disciplines and around the world, showing an overwhelming support for the idea of open access, while highlighting funding and (perceived) quality as the main barriers to publishing in open access journals. The summary is seen as an introduction to the survey and presents highlights from a preliminary analysis of the survey responses. To allow a maximal re-use of the information collected by the survey, the data were released under a CC0 waiver, so to allow libraries, publishers, funding agencies and academics to further analyse risks and opportunities, drivers and barriers, in the transition to open access publishing (from the abstract). The project was financed by the European Commission and ran from 2009 to 2011. The following parties were involved in the project: CERN (coordinator),the publishers Springer, Sage and BioMed Central, the Max Planck Digital Library of the Max Planck Society and the funding agency UK Science and Technology Facilities Council.							
3. Main focus (key dimensions according to MoRRI)	RRI / RI	☐	Citizen participation	☐	Science literacy	☐	Gender equality	☐
	Open access	☒	R&I governance and ethics	☐	Other	☐		
Comment on 3:	Similar to the study by Knowledge Experts, here researchers attitudes and experiences							

	are at the center. Whereas the Knowledge Expert study looked at sharing research data, hiere the subject is open access.							
4. Main perspective (multiple entries possible)	Theoretical, conceptual	☐	Methodological	☐	Policy oriented	☒	Evaluative	☒
	Other	☐	Comment on 4:					
5. Type of document	Scientific article	☒	Book chapter	☐	Book	☐	Report	☒
	Project deliverable	☐	Policy/ strategy document	☐	Other	☐		
Comment on 5:								
6. System level (if applicable)	Global	☒	European	☐	National	☐	Sub-national	☐
Comment on 6:								
7.1 Country focus (if applicable, please specify)	Global							
7.2 Country/ies of origin indicated by institutional affiliation of editor(s)/ author(s) (if applicable, please specify)	The paper has 17 authors, coming from the following countries: CH, UK, G, NL, Norway and Quatar				Comments on 7:			

Data and indicator availability				
8.1 Data, indicators, measurements	Document contains data	☐	If yes, please specify (including page numbers in document)	
Comment on 8.1				
8.2 Reference made to data, indicators measurements in other sources	Document refers to relevant sources	☐	If yes, please list source(s): (URLs, data banks, reports, statistics, etc.)	
Comment on 8.2:				

Guiding questions for review *- please add page numbers where appropriate -*	
9. How is RRI characterized? (For literature dealing explicitly with responsible (research) and innovation. If the publication deals with one of the 5 key dimensions, please proceed to 11.)	
9.1 Which definition of RRI is being used? (author's definition or reference to other source)	
9.2 Which aspects of RRI receive special emphasis? (e.g., certain normative goals, procedural approaches, reference to one or more of the 5 key dimensions, …)	
9.2 Which arguments are presented in support or rejection/criticism of RRI?	

9.3 To which concepts, theories, approaches, schools of thought, communities (scientific or practice) in the area of research and innovation does the literature relate or make reference to? (e.g., STS, constructive TA, anticipatory governance, foresight, deliberative democracy, …)	
Comments on 9.	
10. Policy context of RRI (For literature dealing explicitly with responsible (research) and innovation. If the publication deals with one of the 5 key dimensions, please proceed to 11.)	
10.1 Which RRI-related developments (international, EU, national, sub-national) are mentioned, how are they characterized and what are they aiming at (strategies, funding initiatives, regulation etc.)?	
10.2 Which approaches, instruments are discussed to facilitate the uptake of RRI?	
10.3 Which problems, barriers, potential drawbacks for RRI are brining discussed, how could they be addressed?	
Comments on 10.	
11. Claims regarding the effects of RRI and / or the key dimension (benefits, costs, disadvantages, trade-offs)	
11.1 What claims are being made?	In the SOAP-online survey, about 40.000 active researchers around the world in all scientific disciplines gave their assessments on Open Access. According to these researchers, the benefit of open access publishing lies in fast sharing of research results, (36%), followed by financial issues and the relevance for the public good, with around 20%. The benefit for the individual, to gain more visibility as an author, was fourth at 10%." (p. 5) Also, the survey asked for perceived barriers of open access publishing. "Funding" was mentioned most often with 39%, followed by "journal quality", "accessibility" and "unawareness" (p. 7)

	In the conclusion, the authors claim that their survey presents a representative cross-section of attitudes on open access publishing which was previously not analysed. (p. 10) "In addition, a "survey within the survey" of scholars with experience in open access publishing presents novel data on their experience with the process of paying publication fees. The most relevant findings of the survey are that around 90% of researchers (…) are convinced that open access is beneficial for (…) research field, directly improving the way the scientific community work." (p. 10) At the same time, it is stated that only 10% of all articles published annually are published in open access journals. "The origin of this gap", the authors conclude, "is apparently mostly due to funding and to the (perceived) lack of high-quality open access journals in particular fields. At the same time, many scientists publish open access articles, without directly incurring costs. Those who do pay fees, however, have a wide varying level of experience on the ease of accessing funds." (p. 10)
11.2 Which arguments are used to support the claim(s)?	
11.3 What evidence is presented to support the claims? (e.g., data, indicators, research results, case studies, anecdotal evidence)	
11.4 According to the author(s), which type of evidence/data is missing to better support the claim? (e.g. data gaps, limitations with regard to analytical levels, lack of indicator specifications etc.)	
Comments on 11.	
12. Key dimensions of RRI (For literature dealing with one or more of the 5 key dimensions.)	
12.1 How is the key dimension defined? (terminology applied, central features/characteristics)	
12.2 Does the document reach beyond one single dimension / are more than one of the key dimensions discussed? If yes, what is the proposed relationship between different dimensions (complementary, contradictory…)?	
12.3 To which concepts, theories, approaches, schools of thought, communities (scientific or practice) in the area of research and innovation does the literature	

relate or make reference to? (e.g., STS, constructive TA, anticipatory governance, foresight, deliberative democracy, ...)	
Comments on 12.	
13. Are other important "dimensions" / aspects of RRI discussed, presented which are so far not covered by MoRRI?	
14. Anything else deemed relevant?	
15. General comments and remarks	Although only 4 years old, this study seems to be already outdated. Especially in the light of the most current study on Open access, which is from Archambault (2014), already in the sample of texts for this project, the results seem to be not valifd anymore. Especially the claim that there are no high-quality open access journals has proved to be wrong in the meantime. Also, the level of detail given in the study is poor. It is recommended to turn to more current papers in order to get an up-to-date-picture of open access.
16. Relevant sources cited (Please list references to other sources cited in the literature which seem to be highly relevant for MoRRI and/or represent important contributions in the field)	

Basic information		**Document no.:** (citavi #)	**000**
Reviewer's name	RC		
1. Bibliographical information (author/s, year, title, editor/s, journal/book, volume, publisher, place of publication, pages, DOI)	Fienberg, S. E., Martin, M. E., & Straf, M. L. (1985). Sharing Research Data. Washington: National Academy Press.		
2. Abstract (copy and paste)	This report originated from a letter sent in May 1979 by Professor Melvin Reder of the University of Chicago School of Business to the executive director of the Committee on National Statistics (CNSTAT). Professor Reder proposed a conference on the sharing of social science research data to examine and discuss the conflicting pressures affecting researchers regarding the disclosure to others of data and preliminary analyses. Such a conference, chaired by Clifford Hildreth, was held in October 1979. The participants raised many points and recommended further work by CNSTAT. The committee expresses its thanks and appreciation to the participants, who are listed in the appendix to this volume. In response to the conference recommendation, the Sloan Foundation provided the committee with a grant to work toward the development and dissemination of guidelines for the sharing of scientific data, and the System Development Foundation provided a further grant for work on this report. The study was also supported by a consortium of federal agencies that provide funding for the general activities of CNSTAT. A subcommittee of CNSTAT members was appointed to oversee the project; it was responsible for obtaining and reviewing commissioned papers, developing a set of guidelines for sharing data, and preparing this report for the committee. Although some of their terms of appointment on the full		

	committee expired, all subcommittee members continued to serve throughout the study. We were fortunate to obtain the services and cooperation of several scholars who prepared papers following a general outline developed by the sub- committee. The commissioned papers are Part II of this volume and represent different vantage points on the issues of data sharing. The sub committee is especially appreciative of the detailed materials and suggestions contained in these papers and has relied heavily on them in formulating and structuring the discussion of the costs and benefits of data sharing as well as in developing its recommendations. The first paper, prepared at the Inter-university Consortium for Political and Social Research at the University of Michigan by Jerome M. Clubb with coauthors Erik W. Austin, Carolyn L. Geda, and Michael W. Traugott, deals primarily with large social science data sets. The other four papers deal with the advantages and disadvantages of data sharing more broadly. The paper by Robert F. Boruch of the Department of Psychology at Northwestern University describes products of data sharing. The paper by Terry E. Hedrick of the Institute for Program Evaluation of the U.S. General Accounting Office discusses justifications for and obstacles to data sharing. The paper by Joe Shelby Cecil of the Federal Judicial Center and Eugene Griffin of Northwestern University discusses legal issues relevant to data sharing and provides an important analysis of current pertinent law. And the paper by Robert F. Boruch and David S. Cordray of the Department of Psychology at Northwestern University suggests professional codes and guidelines for data sharing. Margaret E. Martin and Miron L. Straf served as staff of the subcommittee and coeditors of this report. Lenore Bixby prepared a report of the early conference that led to the development of this study. Eugenia Grohman contributed greatly in editing our manuscript and guiding it toward publication. Valuable assistance was provided by Roberta Pirosko in bibliographic work and in typing and by Diane Goldman in proofreading and manuscript preparation. Using the computer for word processing, telecommunications, and typesetting, Lee R. Paulson prepared many versions of our manuscript; she also provided bibliographic and other research assistance. Reviewers and many others offered valuable comments and suggestions for our report. To all who have worked with us or otherwise contributed, we are very grateful. The committee views this report as an initial examination of some of the issues of data sharing, on which readers are invited to comment.

3. Main focus (key dimensions according to MoRRI)	RRI / RI	☐	Citizen participation	☐	Science literacy	☐	Gender equality	☐
	Open access	**X**	R&I governance and ethics	☐	Other	☐		
Comment on 3:								
4. Main perspective (multiple entries possible)	Theoretical, conceptual	**X**	Methodological	☐	Policy oriented	**X**	Evaluative	☐
	Other	☐	Comment on 4:					
5. Type of document	Scientific article	☐	Book chapter	☐	Book	**X**	Report	☐
	Project deliverable	☐	Policy/ strategy document	☐	Other	☐		
Comment on 5:								
6. System level (if applicable)	Global	**X**	European	☐	National	☐	Sub-national	☐
Comment on 6:								
7.1 Country focus (if applicable, please specify)								
7.2 Country/ies of origin indicated by institutional affiliation of editor(s)/ author(s) (if applicable, please specify)					Comments on 7:			
Data and indicator availability								
8.1 Data, indicators, measurements	Document contains data	☐	If yes, please specify (including page numbers in document)					
Comment on 8.1								

8.2 Reference made to data, indicators measurements in other sources	Document refers to relevant sources	☐	If yes, please list source(s): (URLs, data banks, reports, statistics, etc.)	
Comment on 8.2:				

Guiding questions for review

- please add page numbers where appropriate -

9. How is RRI characterized? (For literature dealing explicitly with responsible (research) and innovation. If the publication deals with one of the 5 key dimensions, please proceed to 11.)	
9.1 Which definition of RRI is being used? (author's definition or reference to other source)	
9.2 Which aspects of RRI receive special emphasis? (e.g., certain normative goals, procedural approaches, reference to one or more of the 5 key dimensions, …)	
9.2 Which arguments are presented in support or rejection/criticism of RRI?	
9.3 To which concepts, theories, approaches, schools of thought, communities (scientific or practice) in the area of research and innovation does the literature relate or make reference to? (e.g., STS, constructive TA, anticipatory governance, foresight, deliberative democracy, …)	

Comments on 9.	
10. Policy context of RRI (For literature dealing explicitly with responsible (research) and innovation. If the publication deals with one of the 5 key dimensions, please proceed to 11.)	
10.1 Which RRI-related developments (international, EU, national, sub-national) are mentioned, how are they characterized and what are they aiming at (strategies, funding initiatives, regulation etc.)?	
10.2 Which approaches, instruments are discussed to facilitate the uptake of RRI?	
10.3 Which problems, barriers, potential drawbacks for RRI are brining discussed, how could they be addressed?	
Comments on 10.	
11. Claims regarding the effects of RRI and / or the key dimension (benefits, costs, disadvantages, trade-offs)	
11.1 What claims are being made?	One of the founding papers on the discussion of Data sharing problems, limitations and possibilities. Present a thorough discussion on the benefits, problems, issues, controversies and other consequences of sharing data. Elements discussed are issues related with data sharing, parties involved, benefits, costs, changing environments for data sharing and recommendations. Many of the conclusions of the report are still valid and relevant, such as the need for developing guidelines on data sharing, the need of involving multiple institutions and stakeholders, the need for government policies and standards for accessing, classifying, documenting and archiving data.
11.2 Which arguments are used to support the claim(s)?	Issues: 1) need for guidelines and desirable practices. 2) need for institutions to promulgate these practices. 3) Importance of government policies. 4) Similar problems of data access in the natural sciences and the social sciences. 5) Need for standards - Parties related with data-sharing: investigators (who collect data for analysis); subsequent analysts (whoe analyize one or more datasets collected by others); scientific community (all scientists who engage in research); agencies and research funders (public and private groups that give grants or contracts for research to be performed by others); organizations that conduct research (universities, nonprofit institutions, commercial organizations, individuals and government agencies that conduct research); respondents to surveys and participants in experiments (those who agree to participate in a survey or experiment, might have an interest in the protection and confidentiality of information); the public (society generally). The parties may have different and sometimes conflicting interests when it comes to data sharing. - Benefits: reinforcement of open scientific inquiry; Verification, refutation, or refinement of original results; new results through existing data; encouraging more appropriate use of empirical data in policy formulation and evaluation;

	improvements of measurement and data collection methods; better theories and analytical techniques; encouragement of multiple perspectives; resources for training in research; protection agains faulty data; science would be more efficiently advanced and more effectively applied to making decisions; - Costs of data sharing: technical obstacles (incompatibilities in machine and software systems and data file structures); documentation; cost to the original researchers; costs for the subsequent analysts; - Changing environment for data sharing: developments in computers and software; changes in research practices; different rewards and incentives for research; new laws and regulations; privacy and confidencialit; freedom of information; - Recommendations: 1.Sharing data should be a regular practice; 2. Data should be shared by the time of publication of initial major results; 3. data relevant to public policy should be shared as quickly and widely as possible; 4. Data sharing plans should be part of the research plan; 5. Data should be preserved for a reasonable period after publication; 6. Subsequent analysts should bear the associated incremental costs; 7. Subsequent analysts should endeavor to keep the burdens of data sharing on initial investigators to a minimum and explicitly acknowledge the contribution of the initial investigators. 8. Funders should encourage data sharing, including plans in their applications for research funds. 9. Large-scale, general-purpose datasets should be alert to the need for data archives. 10. Journal editors should require authors to provide access to data during peer review process. 11. Journals should give more emphasis to reports of secondary analyses and to replications. 12. Journals should require full credit and appropriate citatinos to original data collections in reports based on secondary analyses. 13. Journals should strongly encourage authors to make detailed data accessible to other reserachers. 14. Opportunties to provide training on data sharing principles and practices should be pursued and explained. 15. A comprehensive reference service for computer-readable social science data should be developed. 16.Institutions and organizations through which scientists are rewarded should recognize the contributions of appropriate data-sharing practices. Pruposes and products of data sharing: 1.Verifying or examinnig the conclusions of earlier analyses; 2. Facilitating education and training through examples; 3.Testing new hypothesis; 4. Facilitating new methods of analysis; 5.Design other studies; 6.Facilitate syntheses of knowledge
11.3 What evidence is presented to support the claims? (e.g., data, indicators, research results, case studies, anecdotal evidence)	Discussion of a collection of other publications.
11.4 According to the author(s), which type of evidence/data is missing to better support the claim? (e.g. data gaps, limitations with regard to analytical levels, lack of indicator specifications etc.)	
Comments on 11.	
12. Key dimensions of RRI (For literature dealing with one or more of the 5 key dimensions.)	
12.1 How is the key dimension defined? (terminology applied, central features/characteristics)	Data sharing in the main term used. Open research, Open scientific inquiry, Open communication, open science, open access to data, etc. are terms used in the description of the report.
12.2 Does the document reach beyond one single dimension / are more than one of the key dimensions discussed? If yes, what is the proposed relationship between different dimensions (complementary, contradictory…)?	Only data sharing.
12.3 To which concepts, theories,	

approaches, schools of thought, communities (scientific or practice) in the area of research and innovation does the literature relate or make reference to? (e.g., STS, constructive TA, anticipatory governance, foresight, deliberative democracy, …)	
Comments on 12.	
13. Are other important "dimensions" / aspects of RRI discussed, presented which are so far not covered by MoRRI?	
14. Anything else deemed relevant?	
15. General comments and remarks	Foundational document in the data sharing research realm. Still many of the claims and main views there presented have some value in the current debate.
16. Relevant sources cited (Please list references to other sources cited in the literature which seem to be highly relevant for MoRRI and/or represent important contributions in the field)	

Basic information								**Document no.:** (citavi #)	**000**
Reviewer's name	Rodrigo Costas								
1. Bibliographical information (author/s, year, title, editor/s, journal/book, volume, publisher, place of publication, pages, DOI)	Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 368(1926), 4023–38. doi:10.1098/rsta.2010.0165								
2. Abstract (copy and paste)	Results are presented from the Data Curation Profiles project research, on who is willing to share what data with whom and when. Emerging from scientists' discussions on sharing are several dimensions suggestive of the variation in both what it means 'to share' and how these processes are carried out. This research indicates that data curation services will need to accommodate a wide range of subdisciplinary data characteristics and sharing practices. As part of a larger set of strategies emerging across academic institutions, institutional repositories (IRs) will contribute to the stewardship and mobilization of scientific research data for e-Research and learning. There will be particular types of data that can be managed well in an IR context when characteristics and practices are well understood. Findings from this study elucidate scientists' views on 'sharable' forms of data—the particular representation that they view as most valued for reuse by others within their own research areas—and the anticipated duration for such reuse. Reported sharing incidents that provide insights into barriers to sharing and related concerns on data misuse are included								
3. Main focus (key dimensions according to MoRRI)	RRI / RI	☐	Citizen participation	☐	Science literacy	☐	Gender equality	☐	
	Open access	**X**	R&I governance and ethics	☐	Other	☐			
Comment on 3:									

4. Main perspective (multiple entries possible)	Theoretical, conceptual	X	Methodological	☐	Policy oriented	☐	Evaluative	☐
	Other	☐	Comment on 4:					
5. Type of document	Scientific article	X	Book chapter	☐	Book	☐	Report	☐
	Project deliverable	☐	Policy/ strategy document	☐	Other	☐		
Comment on 5:								
6. System level (if applicable)	Global	X	European	☐	National	☐	Sub-national	☐
Comment on 6:								
7.1 Country focus (if applicable, please specify)								
7.2 Country/ies of origin indicated by institutional affiliation of editor(s)/ author(s) (if applicable, please specify)					Comments on 7:			
Data and indicator availability								
8.1 Data, indicators, measurements	Document contains data	☐	If yes, please specify (including page numbers in document)					
Comment on 8.1								
8.2 Reference made to data, indicators measurements in other sources	Document refers to relevant sources	☐	If yes, please list source(s): (URLs, data banks, reports, statistics, etc.)					
Comment on 8.2:								
Guiding questions for review *- please add page numbers where appropriate -*								
9. How is RRI characterized? (For literature dealing explicitly with responsible (research) and innovation. If the publication deals with one of the 5 key dimensions, please proceed to 11.)								
9.1 Which definition of RRI is being used? (author's definition or reference to other source)								
9.2 Which aspects of RRI receive								

special emphasis? (e.g., certain normative goals, procedural approaches, reference to one or more of the 5 key dimensions, …)	
9.2 Which arguments are presented in support or rejection/criticism of RRI?	

9.3 To which concepts, theories, approaches, schools of thought, communities (scientific or practice) in the area of research and innovation does the literature relate or make reference to? (e.g., STS, constructive TA, anticipatory governance, foresight, deliberative democracy, …)	
Comments on 9.	
10. Policy context of RRI (For literature dealing explicitly with responsible (research) and innovation. If the publication deals with one of the 5 key dimensions, please proceed to 11.)	
10.1 Which RRI-related developments (international, EU, national, sub-national) are mentioned, how are they characterized and what are they aiming at (strategies, funding initiatives, regulation etc.)?	
10.2 Which approaches, instruments are discussed to facilitate the uptake of RRI?	
10.3 Which problems, barriers, potential drawbacks for RRI are brining discussed, how could they be addressed?	
Comments on 10.	
11. Claims regarding the effects of RRI and / or the key dimension (benefits, costs, disadvantages, trade-offs) Open data and data sharing	
11.1 What claims are being made?	Analysis on the perceptions about data sharing. - Data sharing in 'small science' is not common or expected. - Scholars seek for assistance with their data problems in their libraries and repositories. -In table 1 there are description of concepts such as 'types and value', 'format', 'size', 'accessibility', 'intellectual property', etc. - There are positive views of data sharing. Importance of the time involved for the willing scientists to make the data available (participants say they are willing to say, but only a few had done it) - Rules for sharing are not systematic - Co-authorships of data creators is emphasized.

	- Distinguishing private and public sharing. - Concerns regarding sharing with the wider public: misuse, use by others before creators have published about it; industry uses of the data (e.g. cherry-picked data to make wrong or false claims); wrong interpretation of data; improper citation - Scientists rarely have the skills for sharing, thus they need support of other professionals (e.g. data curators, curation services) - Unawareness among scholars of the standards about metadata and standards. - In small-science research areas there were no field-wide norms for sharing. - Need to untangle the concept of data sharing. - Openness in the sharing and publication of scientific data can also work as a protection against fraud and faulty data in scientific research. - Extra costs and effort to share data and data publication is perceived as an important barrier for sharing. - Perception of data misuse as an important problem for data sharing.
11.2 Which arguments are used to support the claim(s)?	
11.3 What evidence is presented to support the claims? (e.g., data, indicators, research results, case studies, anecdotal evidence)	Interviews and structured worksheets from 20 scientists who conduct small-science research and have an interest in data management or sharing. The sample included researchers whose work is data intensive and who generate large digital datasets.
11.4 According to the author(s), which type of evidence/data is missing to better support the claim? (e.g. data gaps, limitations with regard to analytical levels, lack of indicator specifications etc.)	
Comments on 11.	
12. Key dimensions of RRI (For literature dealing with one or more of the 5 key dimensions.)	
12.1 How is the key dimension defined? (terminology applied, central features/characteristics)	
12.2 Does the document reach beyond one single dimension / are more than one of the key dimensions discussed? If yes, what is the proposed relationship between different dimensions (complementary, contradictory…)?	
12.3 To which concepts, theories, approaches, schools of thought, communities (scientific or practice) in the area of research and innovation does the literature relate or make reference to? (e.g., STS, constructive TA, anticipatory governance, foresight, deliberative democracy, …)	
Comments on 12.	
13. Are other important "dimensions" / aspects of RRI discussed, presented which are so far not covered by MoRRI?	

14. Anything else deemed relevant?	
15. General comments and remarks	
16. Relevant sources cited (Please list references to other sources cited in the literature which seem to be highly relevant for MoRRI and/or represent important contributions in the field)	

Basic information									**Document no.:** (citavi #)	**000**
Reviewer's name	Rodrigo Costas									
1. Bibliographical information (author/s, year, title, editor/s, journal/book, volume, publisher, place of publication, pages, DOI)	Costas, R., Meijer, I., Zahedi, Z., & Wouters, P. (2013). Metrics for datasets from a cultural and technical point of view. A Knowledge Exchange Report. Copenhagen. http://www.knowledge-exchange.info/datametrics									
2. Abstract (copy and paste)	Scientific research revolves around the production, analysis, storage, management, and re-use of data. Data sharing offers important benefits for scientific progress and advancement of knowledge. However, several limitations and barriers in the general adoption of data sharing are still in place. Probably the most important challenge is that data sharing is not yet very common among scholars and is not yet seen as a regular activity among scientists, although important efforts are being invested in promoting data sharing. In addition, there is a relatively low commitment of scholars to cite data. The most important problems and challenges regarding data metrics are closely tied to the more general problems related to data sharing. The development of data metrics is dependent on the growth of data sharing practices, after all it is nothing more than the registration of researchers' behaviour. At the same time, the availability of proper metrics can help researchers to make their data work more visible. This may subsequently act as an incentive for more data sharing and in this way a virtuous circle may be set in motion. This report seeks to further explore the possibilities of metrics for datasets (i.e. the creation of reliable data metrics) and an effective reward system that aligns the main interests of the main stakeholders involved in the process. The report reviews the current literature on data sharing and data metrics. It presents interviews with the main stakeholders on data sharing and data metrics. It also analyses the existing repositories and tools in the field of data sharing that have special relevance for the promotion and development of data metrics. On the basis of these three pillars, the report presents a number of solutions and necessary developments, as well as a set of recommendations regarding data metrics. The most important recommendations include the general adoption of data sharing and data publication among scholars; the development of a reward system for scientists that includes data metrics; reducing the costs of data publication; reducing existing negative cultural perceptions of researchers regarding data publication; developing standards for preservation, publication, identification and citation of datasets; more coordination of data repository initiatives; and further development of interoperability protocols across different actors.									
3. Main focus (key dimensions according to MoRRI)	RRI / RI	☐	Citizen participation	☐	Science literacy	☐	Gender equality	☐		
	Open access	**X**	R&I governance and ethics	☐	Other	☐				
Comment on 3:										
4. Main perspective (multiple entries possible)	Theoretical, conceptual	**X**	Methodological	**X**	Policy oriented	**X**	Evaluative	**X**		
	Other	☐	Comment on 4:							

5. Type of document	Scientific article	☐	Book chapter	☐	Book	☐	Report	**X**
	Project deliverable	☐	Policy/ strategy document	☐	Other	☐		
Comment on 5:								
6. System level (if applicable)	Global	**X**	European	☐	National	☐	Sub-national	☐
Comment on 6:								
7.1 Country focus (if applicable, please specify)								
7.2 Country/ies of origin indicated by institutional affiliation of editor(s)/ author(s) (if applicable, please specify)					Comments on 7:			

Data and indicator availability

8.1 Data, indicators, measurements	Document contains data	**X**	If yes, please specify (including page numbers in document)	The document extensively describes data repositories (pp.39 onwards)
Comment on 8.1	The document does not contain data by itself, but it includes substantial links to other data repositories and databases that can have practical use in the MoRRI project.			
8.2 Reference made to data, indicators measurements in other sources	Document refers to relevant sources	**X**	If yes, please list source(s): (URLs, data banks, reports, statistics, etc.)	Appendix 2 in the report (pp. 39 onwards) lists and systematically evaluates data repositories. The report includes abundant references and links to relevant sources, including initiatives (e.g. Research Data Alliance, International Council of Science, STAR METRICS), repositories (e.g. DataCite, Dryad, CrossRef, etc.) and relevant other literature and documents.
Comment on 8.2:				

Guiding questions for review

- please add page numbers where appropriate -

9. How is RRI characterized?

(For literature dealing explicitly with responsible (research) and innovation. If the publication deals with one of the 5 key dimensions, please proceed to 11.)

9.1 Which definition of RRI is being used? (author's definition or reference to other source)	
9.2 Which aspects of RRI receive special emphasis? (e.g., certain normative goals, procedural approaches, reference to	

one or more of the 5 key dimensions, …)	
9.2 Which arguments are presented in support or rejection/criticism of RRI?	

9.3 To which concepts, theories, approaches, schools of thought, communities (scientific or practice) in the area of research and innovation does the literature relate or make reference to? (e.g., STS, constructive TA, anticipatory governance, foresight, deliberative democracy, …)	
Comments on 9.	
10. Policy context of RRI (For literature dealing explicitly with responsible (research) and innovation. If the publication deals with one of the 5 key dimensions, please proceed to 11.)	
10.1 Which RRI-related developments (international, EU, national, sub-national) are mentioned, how are they characterized and what are they aiming at (strategies, funding initiatives, regulation etc.)?	
10.2 Which approaches, instruments are discussed to facilitate the uptake of RRI?	
10.3 Which problems, barriers, potential drawbacks for RRI are brining discussed, how could they be addressed?	
Comments on 10.	
11. Claims regarding the effects of RRI and / or the key dimension (benefits, costs, disadvantages, trade-offs)	
11.1 What claims are being made?	The main objective of the report is to present an overview on the state of the arte in the area of data metrics, as well as existing solutions, providing also a critical assessment of possibilities for their use and suggestions for further actions. The report points to the main challenges related with data metrics. A kind of 'vicious circle' is described: data sharing is not a common activity among scholars, this implies that the development of data metrics is limited by the low incidence of data sharing activities, thus it is difficult to establish metrics to measure it and reward it and keeps low the commitment of scholars to share and cite data. The report also describes main challenges for the development of data metrics: - Data sharing is not very common among scholars. - Low commitment of scholars to cite data. - Barriers perceived by scholars are lack of time and lack of funding. - Data publication and data citation is not an element considered for

	promotion and research assessment. - 'Data sharing vicious circle' (circular paradox): scholars do not share their data because they feel that they are not rewarded by this, this implies that the volume of data publications and data citations is poor, thus limiting the development of a reward system that would incentivize data sharing. - Scientists fear the ' loss of control' over their data. - Potential decrease in the quality of science, with more researchers simply reusing datasets of other instead of collecting new ones. - Embargoes needed. - Disciplinary differences in needs for data reuse. Some solutions are also discussed: - Need for a reward system of scientists that considers data metrics. - Development of standards for data citation. - Institutional commitment. Recommendations proposed: - General adoption of data sharing and data publication among scholars. - Development of reward systems that include data metrics. - Reduce costs and make the process of data publication more efficient. - Reduce the negative perceptions of researchers regarding data publication, via policies that inform and create awareness among scholars on the importance of data sharing. - Solution of the most important technical problems and lack of standards for preservation. - Solution of organizational problems. - Reduction of data repositories and coordination of initiatives. - Develop standards and interoperability protocols across the different actors. Recommendations for the different stakeholders are also proposed (table 6):

<table>
<tr><td></td><td>
<table>
<tr><th>Stakeholders</th><th>Recommendations</th></tr>
<tr><td>Funders</td><td>* Demand and reward data sharing activities
* Consider data metrics in assessments
* Inform policy about the importance and benefits of data sharing
* Promote open access of data</td></tr>
<tr><td>Research infrastructures</td><td>* Promote policies of data sharing
* Promote arguments and incentives in favour of data sharing
* Provide options and alternatives to the different types of data sharing activities
* Professionalize staff and standardize data sharing activities (collection, curation, dissemination)</td></tr>
<tr><td>Scientists</td><td>* Include data sharing as good scientific and scholarly practice
* Promote data citation as the formal way of acknowledging data sharing
* Perform more research on benefits and possibilities of data sharing
* Define codes of conducts for disciplines considering appropriate regulations, i.e. embargo periods, anonymisation etc.</td></tr>
<tr><td>Data centres</td><td>* Inform the scientific community about data activities and services
* Contribute to reduce the dispersion of data repositories
* Develop robust solutions for the preservation and standardisation of the data storage and citations
* Develop tools for tracking the users of the repositories</td></tr>
<tr><td>Publishers</td><td>* Promote data sharing in their publications and journals
* Inform authors about other data sharing stakeholders (e.g. repositories, data centres)
* Support open access to data</td></tr>
<tr><td>Libraries</td><td>* Promote data publications and data citations
* Coach scholars and research managers in their data publication and citation activities
* Inform authors about other data sharing stakeholders (e.g. funders, repositories, data centres)
* Develop tools to find data repositories
* Develop and test appropriate metrics</td></tr>
<tr><td>Publication databases</td><td>* Collect and measure data publications and data citations
* Facilitate the analysis and metrics of data publications and data citations</td></tr>
</table>
sdf
</td></tr>
<tr><td>11.2 Which arguments are used to support the claim(s)?</td><td></td></tr>
<tr><td>11.3 What evidence is presented to support the claims?
(e.g., data, indicators, research results, case studies, anecdotal evidence)</td><td>Desk research, exploration and analysis of online tools, literature review and interviews with main stakeholders in data sharing and open data (e.g. repositories, funders, etc.).</td></tr>
<tr><td>11.4 According to the author(s), which type of evidence/data is missing to better support the claim? (e.g. data gaps, limitations with regard to analytical levels, lack of indicator specifications etc.)</td><td></td></tr>
<tr><td>Comments on 11.</td><td></td></tr>
<tr><td colspan="2">12. Key dimensions of RRI
(For literature dealing with one or more of the 5 key dimensions.)</td></tr>
<tr><td>12.1 How is the key dimension defined?
(terminology applied, central features/characteristics)</td><td></td></tr>
<tr><td>12.2 Does the document reach beyond one single dimension / are more than one of the key dimensions discussed? If yes, what is the proposed relationship between</td><td></td></tr>
</table>

different dimensions (complementary, contradictory…)?	
12.3 To which concepts, theories, approaches, schools of thought, communities (scientific or practice) in the area of research and innovation does the literature relate or make reference to? (e.g., STS, constructive TA, anticipatory governance, foresight, deliberative democracy, …)	Citation and metrics theories.
Comments on 12.	
13. Are other important "dimensions" / aspects of RRI discussed, presented which are so far not covered by MoRRI?	
14. Anything else deemed relevant?	
15. General comments and remarks	
16. Relevant sources cited (Please list references to other sources cited in the literature which seem to be highly relevant for MoRRI and/or represent important contributions in the field)	Most of the other references discussed in this analytical report are also discussed in this report regarding Data Sharing, thus making the report a central piece in the discussion and understanding of the challenges and opportunities regarding the development of indicators for data sharing.