DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INNOVATION
COMMITTEE FOR SCIENTIFIC AND TECHNOLOGICAL POLICY

OPEN ACCESS TO DATA IN SCIENCE, TECHNOLOGY AND INNOVATION – INITIAL SURVEY FINDINGS

Current use of the OECD Recommendation concerning Access to Research Data from Public Funding and policy challenges related to open access to data

24-25 October 2017, OECD Conference Centre, Paris

This document provides an initial overview of findings related to the topic of open access to data in science, technology and innovation, based on the analysis of responses to a questionnaire regarding current usage and future needs.

CSTP delegates are invited to note and discuss those findings, in view of a potential future policy report, and the development of a possible overarching Recommendation on data access.

The document will be discussed under agenda item 15 of the 111th session of the CSTP.

Contact: Alan Paic, alan.paic@oecd.org

JT03419788

Table of contents

Figures

Boxes

1. Introduction

1.   At its 109th session in October 2016, the CSTP discussed and approved a proposed joint development of a possible new overarching recommendation on enhanced access to data, together with the Committee for Digital Economy Policy (CDEP) and the Public Governance Committee (PGC) [COM/DSTI/CDEP/STP/GOV/PGC(2016)1].

2.   As pointed out in the Summary Record of CSTP's 109th Session [DSTI/STP/M(2016)2]:

> "In view of the importance taken by data access in all social and economic areas and of changes in the situation since 2007 (new technology, new uses), one of the proposed key deliverables of the CSTP project on digitalisation will be for the CSTP, CDEP and Public Governance Committee (PGC) to develop jointly a new OECD recommendation on access to data covering all types of public data.
>
> […]
>
> The Committee agreed on the proposed key outputs of the project, including the three-committee recommendation on access to data."

3.   A Joint Steering Group has been established to support this process. The CSTP members for this Steering Group include delegates of Belgium, Brazil, Latvia, Spain and the United States.

4.   As a first step, the Joint Steering Group and the Secretariat proposed to assess current usage of CSTP's central instrument in this domain, the Recommendation of the Council concerning Access to Research Data from Public Funding [C(2006)184]. The objectives of such an assessment are threefold:

- Provide feedback on usage of the Recommendation, including its strengths and weaknesses which need to be addressed;
- Identify potential gaps arising from the evolution of data usage in the past decade; and
- Highlight areas to be covered in a potential future recommendation covering all types of public sector data, beyond public research.

5.   Please note that this assessment does not represent a formal Review of the Recommendation. The instrument itself does not specify a review date, and instructs the CSTP to review the instrument as necessary. An initial formal Review was already conducted in 2009 [DSTI/STP(2009)3] and may be included in the 2019-2020 PWB, if deemed relevant by Delegates, and would lead to a potential update of the Recommendation itself.

6.   On June 13, 2017, a questionnaire was circulated to the CSTP with a request to provide input. The Secretariat received input from 27 Delegations, which is analysed in this paper.

2. Policy context of open access to data

2.1. Overall context of open public sector information

7.   The exponential growth of data production is well documented, amounting to 8 zettabytes (ZB)[1] in 2015, and is expected to grow to 40 ZB by 2020. Data is seen as a key resource of the knowledge economy, as close to real-time analysis accelerates knowledge and value creation across society (OECD, 2015b).

8.   In 2013, McKinsey Global Institute estimated the potential value creation from open access to data in seven sectors[2] to USD 3-5 trillion, or 4-7% of global GDP (McKinsey Global Institute, 2013). An estimate of the impact of public sector information, commissioned by the European Commission, has estimated the aggregate economic impact of public sector information at about EUR 140 bn in 2008, or about 1.1% of EU27 GDP (Vickery, 2011).

9.   Enhanced data access and use offer opportunities for individuals, businesses and governments alike. Individuals are able to access valuable services at negligible cost – for example using a satellite navigation system included in any smartphone at virtually no marginal cost provides a service which is superior to previous costly option of buying a paper map and painstakingly finding one's way. Such services are based on geospatial data which is collected by the public sector.

Box 2.1. Open access to data definition

Open data are data that can be used by anyone without technical or legal restrictions. The use encompasses both access and reuse.

(OECD, 2015a)

---

10.   Businesses use data to learn about consumer preferences, to create new products and services, to streamline their business processes, and increase overall productivity. Moreover, data itself is the commodity being sold. According to McKinsey Global Institute, data flows account for a larger impact on GDP than global trade in goods, i.e. more than 5 percent of global GDP (Manyika, et al., 2016).

11.   Governments are both producers and users of data. Some examples include data on geographical maps, weather, traffic, statistics about businesses, consumption, demography, food, safety, education, government budgets, etc. Such data can be used by governments themselves, by businesses and citizens. Governments can also use big data techniques to understand and predict the behaviour of citizens. Data are used in policy implementation, and could be used to model effects of policies on human behaviour, and

---

[1] 1 ZB = 1 trillion gigabytes

[2] Education, transportation, consumer products, electricity, oil and gas, health care and consumer finance

thus optimise policy making to achieve predictable and favourable outcomes (Pentland, 2014).

12. Potential benefits of open access to data include increased transparency, democratic accountability, empowerment of citizens, increased trust in government, better delivery of public services, facilitation of innovation, use of the wisdom of crowds, avoiding duplication of data collection efforts, optimization of administrative processes, access to external problem-solving capacity and many more (Janssen, Charalabidis, & Zuiderwijk, 2012). Open government data can help combat corruption, both by reducing private returns to corruption, and making it easier to detect (OECD, 2017).

13. However, strong barriers exist to open access to data from the public sector. These include unclear trade-offs between transparency and privacy rules, risk-averse culture of civil servants, the absence of uniform policies on open access to data, lack of resources for data curation and storage, lack of a process to deal with user inputs which could potentially add value to the dataset, lack of user ability to discover appropriate data, lack of metadata, and many more.

14. In addition, publicizing data does not automatically yield benefits, since re-use needs to be stimulated by awareness raising and training – users do not necessarily have the right skills for correct interpretation and analysis. In addition, not all data can be publicised, and reasons to refrain from publishing include privacy considerations, situations where the cost seems excessive compared to the benefit, the quality is too low or the complexity too high, or when such publication may hurt the economic interest of organisations who base their financial existence on the sale of data. Data curation and inclusion of metadata are an essential (and costly) ingredient of open access to data provision, which needs to be factored into any open access to data initiative. (Janssen, Charalabidis, & Zuiderwijk, 2012).

15. The OECD Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information [C(2008)36] provides policy guidelines to improve access and increase use of public sector information through greater transparency, enhanced competition and more competitive pricing.

16. In June 2013, the G8 leaders signed the Open Data Charter, which sets out five principles: (i) Open Data by default; (ii) Quality and quantity; (iii) Usable by all; (iv) Releasing data for improved governance; (v) Releasing data for innovation. (G8, 2013).

17. The European Union followed up with an implementation plan, with specific action points, such as the commitment to publish key datasets under open licence, including the budget of the EU, EU Parliament election results, data about EU staff, finance and contracts, which are available on the European Open Data Portal, which by now contains 11,600 datasets. The EU also committed to promote the application of the G8 Open Data Charter to all Member States (EU implementation of the G8 Open Data Charter).

18. The International Open data Charter, initiated at the Open Government Partnership summit in 2015, also promotes open access to data. It is a collaboration between governments and experts, around six principles for how governments should publish information: (i) key datasets[3] open by default; (ii) timely and comprehensive; (iii)

---

[3] Such as: budget, spending, contracting, land ownership, company registries, legislation, election results

accessible and usable (meaning online availability in machine-readable format, available in bulk for easy downloading, free of charge and open-licensed); (iv) comparable and interoperable; (v) for improved governance and citizen engagement; (vi) for inclusive development and innovation. The Charter has been joined by 17 national[4] and 30 local and subnational governments. (International Open data charter, 2017).

19.   However, progress towards the achievement of those principles has been uneven at best, and the latest edition of the Open access to data Barometer shows that only 7% of the data is fully open, with many fragmented, incomplete and outdated datasets, which also lack the necessary metadata. Open access to data catalogues or portals are informally maintained, and the most complete datasets are often found on sources different from the official open data portal. Some specific categories of datasets are particularly important for innovation, such as map data, public transport timetables, international trade data, and crime data which can be used by entrepreneurs to provide specific services to end users. In these categories as well, the degree of openness is low (representing 8-11% of all datasets), and is reported to be declining, rather than growing (The World Wide Web Foundation, 2017).

20.   Data integration is another major challenge for governments. This implies linking different data sets to exploit the significance of data to its fullest extent. Linking information from hospitals with the cancer data repository and data from various screening programmes in the UK made it possible to recommend changes in medical protocols which should improve cancer survival rates. Data integration needs to be done in a way which preserves personal privacy of citizens. The Australian government launched an AUD 130 million data integration partnership initiative to improve linkages between government data sets. (Australian Government Productivity Commission, 2017).

2.2. Open access to data in science, technology and innovation

21.   In the domain of science and technology, open access to data is commonly linked to open science and open innovation (OECD, 2015a). An extensive European stakeholder consultation conducted in 2014 revealed the priority given to open access to data to publications and open access to data within open science issues, as the top issues in need for policy intervention, ahead of research infrastructure, assessment of quality of research and other issues (EC, 2014).

22.   The development of data acquisition, storage and analysis capabilities provide for scientific disciplines to be ever more data-driven. Scientific data is very diverse: it includes *observational data* which record natural phenomena (such as astronomy, geoscience, demography and others); *experimental data* which records the outcomes of man-made experiments, such as laboratory experiments in physics, chemistry, biology, or clinical trial data; *computational data* which is generated through large-scale simulations; and *reference data*, highly curated datasets such as the human genome for example. Simulation is used to generate data based on theoretical predictions, and the results are compared to actual experiments, in order to verify the validity of theoretical concepts, and adjust them accordingly. Increasingly, the development of artificial intelligence will make it possible for algorithms to detect patterns by themselves (OECD, 2015b).

---

[4] Australia, Argentina, Chile, Colombia, Costa Rica, France, Guatemala, Italy, Mexico, Panama, Paraguay, Philippines, Sierra Leone, Korea, Ukraine, United Kingdom and Uruguay - only seven of these are OECD member countries.

23.   A well-known issue in science is the *publication bias*, whereby negative results, or results not deemed of sufficient significance, are often not published, due to the fact that such publication is not deemed worth the researcher's time and effort, since she will receive very little or no recognition for such a "non-result". However, the failure to publish such data causes additional time and effort elsewhere, where such experiments may be duplicated since it is not known that such an avenue of research does not lead to positive discovery (Rothstein, Sutton, & Borenstein, 2006). Adoption of open access to scientific data would help resolve this bias.

24.   Data driven innovation is transforming society through far reaching effects on resource efficiency, productivity and competitiveness. It also helps address global challenges, such as climate and demographic changes, scarce resources and many others. In this context, the issue of text and data mining is hot topic, since this technique allows extraction of valuable knowledge and information from large digital data sets, but is not allowed in many countries due to concerns about personal privacy (OECD, 2015b).

25.   The OECD Principles and Guidelines for Access to Research Data from Public Funding (OECD, 2007) and the OECD Recommendation on Access to Research Data from Public Funding [C(2006)184] resulted from work of the CSTP, and represent an important step in multilateral efforts to create conditions for the development of open access to data in the domain of science, technology and innovation. The CSTP subsequently reviewed progress by countries in the implementation of the principles in 2009 DSTI/STP(2009)3, and concluded that a wide range of policies were implemented based on those instruments. Some countries had introduced laws and comprehensive policies, while others had position statements and planned to implement further policies in the future.

26.   In the following years, additional multilateral efforts materialised. In 2011, UNESCO adopted a Revised Draft Strategy on UNESCO's Contribution to the Promotion of Open Data to Scientific Information and Research. The document calls for UNESCO to provide policy advice on the development of comprehensive national open access to data policies, strengthen the capacities to adopt open access to data, serve as a clearing house and inform the global open access to data debate. (UNESCO, 2011).

27.   Another multilateral initiative involved European Commission's 2012 Recommendation on access to and preservation of scientific information references the OECD Recommendation on Access to Research Data from Public Funding [C(2006)184], and calls for open access to scientific publications, open access to data, preservation and reuse of scientific information, development of e-infrastructures, in a coordinated way among EU Member States (EC, 2012).

28.   The *Research Data Alliance* was initiated in 2013 by the EC, The US National Science Foundation and National Institute of Standards and the Australian Government's Department of Innovation with the goal of building the social and technical infrastructure to enable open sharing of data. As of September 2017 it has 43 organisational members and 6000 individual members from 130 countries. RDA operates working groups which work on Recommendations, which can be adopted as standards, as well as a number of supporting outputs. (Research Data Alliance Council, 2014). RDA recommendations address a broad range of issues related to interoperability, data citation, data catalogues, workflows for research data publishing, (Research Data Alliance, 2017).

29.   A recent initiative which is quickly gaining momentum is the set of *FAIR principles*, where FAIR stands for Findability, Accessibility, Interoperability, and

Reusability (Wilkinson, Dumontier, Aalversberg, Appleton, & Axton, 2016). These principles are championed by FORCE11 (Future Of Research Communications and E-scholarship), a non-government, multidisciplinary community, established after the 2011 Beyond the PDF meeting in San Diego, dedicated to the transformation of scholarly communications through technology. FAIR principles are being quickly adopted as a new reference by academics and policy makers in many countries.

30.   An important recent development is the Amsterdam call for Open Science in 2016, which formulates twelve action points to ensure open access to all scientific publications by and a new approach for optimal reuse of research data until 2020. The actions are grouped under five pillars: (i) Removing barriers to open science; (ii) developing research infrastructures; (iii) fostering and creating incentives for open science; (iv) mainstreaming and further promoting open science; (v) stimulating and embedding open science in science and society (Netherlands' EU Presidency, 2016).

31.   The years 2015-16 saw the emergence of supranational Science Cloud concepts as a practical solution to creating open access to data for researchers in the form of a single window, and practically implement the FAIR principles.

32.   In June 2016, the European Commission published a report "Open Innovation, Open Science, Open to the World", which outlines a vision for the EU research and innovation policy. The vision is to materialise through the development of the European Open Science Cloud, as well as dedicated Horizon 2020 projects to support open access to scientific data (European Commission, 2016).

33.   The European Open Science Cloud (EOSC) thus aims to offer seamless access to a commons based on scientific data. It is conceived as a European infrastructure, with global accessibility and interoperability. It comprises human expertise, resources, standards, best practices as well as the underpinning technical infrastructures. (EC, 2016). The GO-FAIR initiative is a proposal for the practical implementation of the EOSC. It has been championed by Germany and the Netherlands, as "early movers" in the implementation of the European Open Science Cloud recommendation (ZBW-Mediatalk, 2017).

34.   Other similar initiatives involve the NIH Commons in the United States (NIH, 2017), the Australian Research Data Cloud (Australian Government, 2016), and the African Research Cloud (ARC), which all ultimately aim to be interconnected and interoperable.

35.   The G7 Science and Technology Ministers' Meeting in Tsukuba, Ibaraki in 2016 decided to establish a working group on open science. The working group identified three action points essential to the transition to Open Science: (ia common vision; (ii) adaptation of the incentive and reward systems; and (iii) the development of a federated research data infrastructure. (ZBW – Leibniz-Informationszentrum Wirtschaft, 2016).

36.   In parallel, the OECD Global Science Forum is implementing two projects to inform policies to promote open data for science. The first of these addresses business models for sustainable data repositories, identifying financing needs over the repositories' lifetime and revenue sources from a combination of structural funding, institutional funding and user fees, offering a distinct value proposition to each type of stakeholder [DSTI/STP/GSF(2017)1/REV1]. The second project focuses on internationally coordinated data networks. The overall aim of this project was to identify principles and policy actions that can enable the establishment and maintenance of effective international data networks that are necessary to support a global open science enterprise.

addressing issues of data infrastructure, including sustainable business models for data repositories [DSTI/STP/GSF(2017)5]. These two projects are a follow-up to earlier GSF work on the use of data, including personal data, for social science research (OECD, 2013) and (OECD, 2016). In addition some initial work has been carried out on incentives for sharing data in the specific area of research on dementia [DSTI/STP/MS(2015)16].

3. Survey findings: Open access to data from public research

3.1. CSTP survey on open access to data in science, technology and innovation

37.   As mentioned in the introduction, the Joint Steering Group and the Secretariat proposed to assess current usage of CSTP's central instrument in this domain, the Recommendation of the Council concerning Access to Research Data from Public Funding [C(2006)184]. The objectives of such an assessment are threefold:

- Provide feedback on usage of the Recommendation, including its strengths and weaknesses which need to be addressed;
- Identify potential gaps arising from the evolution of data usage in the past decade; and
- Highlight areas to be covered in a potential future recommendation covering all types of public sector data, beyond public research.
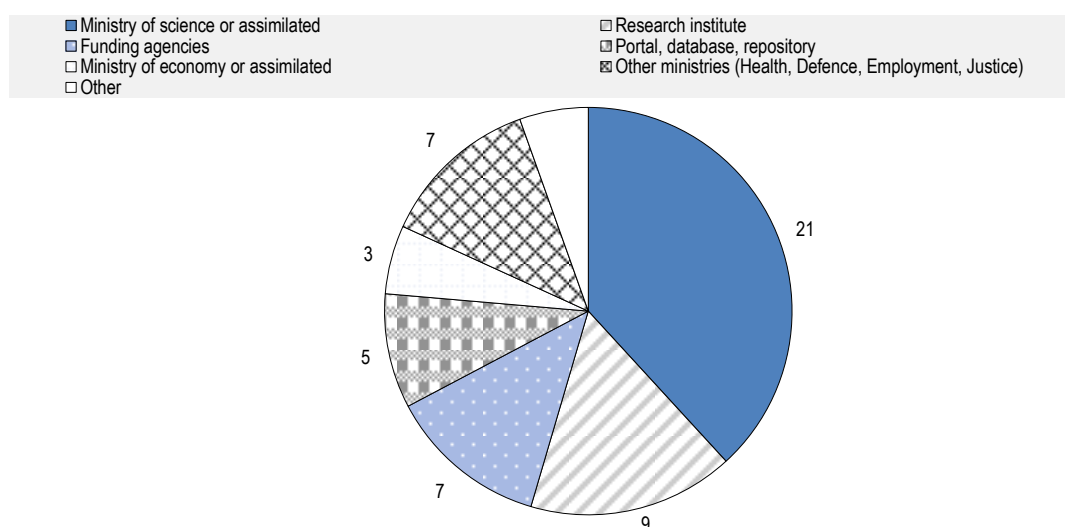
38.   Please note that this assessment does not represent a formal Review of the Recommendation. The instrument itself does not specify a review date, and instructs the CSTP to review the instrument as necessary. An initial formal Review was already conducted in 2009 [DSTI/STP(2009)3] and may be included in the 2019-2020 PWB, if deemed relevant by Delegates, and would lead to a potential update of the Recommendation itself.

39.   On June 13, 2017, a questionnaire was circulated to the CSTP with a request to provide input. The Secretariat received input from 27 Delegations, which is analysed in this paper. The questionnaire is available in Annex 1.

40.   The questionnaire asked respondents to list policy initiatives (reforms or changes to existing legislation) the country has implemented in the domain of access to research data from public funding and public sector information as relevant to science, technology and innovation.

41.   This is an initial survey on this topic, which has no ambition of statistical significance. The objective was to collect official positions from Delegations, who involved a limited number of institutions (an average of 2 per country, although some have mobilised up to 11 institutions).

42.   A total of 55 institutions from 27 countries have responded to the questionnaire. For most countries the response came from Ministries of Science or assimilated (Technology, Innovation, Education and Science,…). However, representatives of research institutes, funding agencies and repositories were also broadly represented (Figure 3.1).

Figure 3.1. Respondent institutions

Legend:
- Ministry of science or assimilated
- Funding agencies
- Ministry of economy or assimilated
- Other
- Research institute
- Portal, database, repository
- Other ministries (Health, Defence, Employment, Justice)
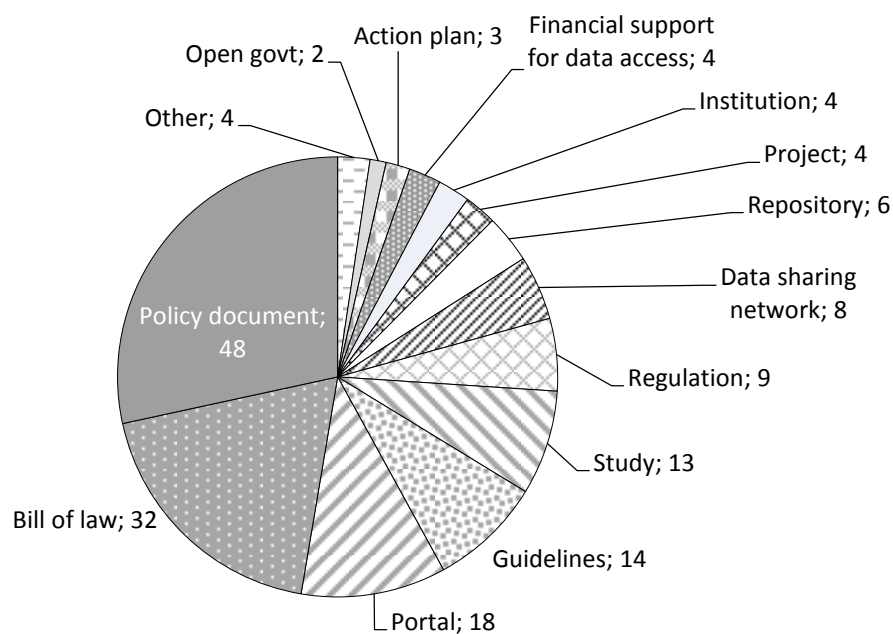
Source: Answers from OECD and Partner delegations

3.2. Policy initiatives in favour of Open access to data relative to STI

43.   A total of 171 policy initiatives were listed from 27 countries. Roughly one half of the initiatives concerned public research in general, one in six initiatives was sector-specific, and one third were broader initiative concerning public sector information, also covering public research data (Figure 3.2).

44.   In terms of initiative types, the sample is quite diversified (Figure 3.3). Roughly half of the initiatives concern policy documents (including strategies) and legal initiatives.

Figure 3.2. Scope of policy initiatives reported in the survey
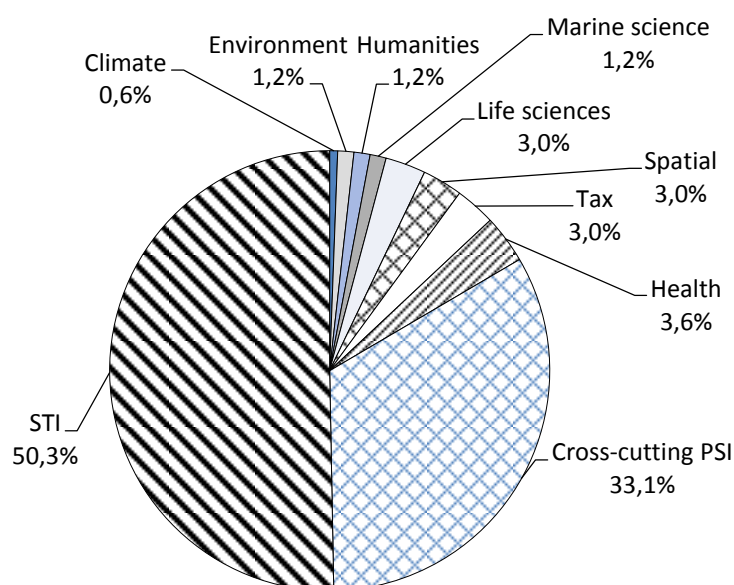
Source: Answers from OECD and Partner delegations.

Figure 3.3. Initiatives by type

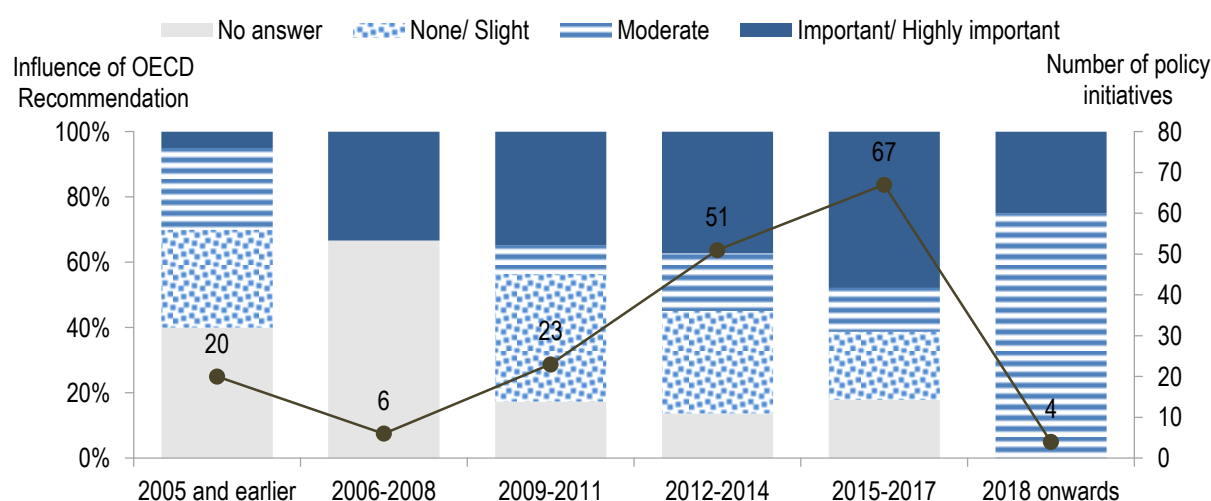*Source*: Answers from OECD and Partner delegations.

3.3. Influence of the OECD Recommendation [C(2006)184]

45.   Respondents to our survey were also asked to assess the influence the Recommendation of the Council concerning Access to Research Data from Public Funding [C(2006)184] for each of those initiatives. Results are shown in Figure 3.4.

46.   Overall, we observe an increasing number of initiatives over time, showing the rising importance of the issue of data access. Even though there might be some selection bias, whereby respondents are more likely to include more recent initiatives[5], the trend seems to be quite strong, with only six initiatives in the triennium 2006-2008 against 67 initiatives in 2015-2017.

47.   Concerning the degree of influence of the Recommendation, we observe a relatively stable proportion of answers saying the influence was 'important' or 'highly important', at around 40% of all initiatives.

Figure 3.4. Influence of OECD Recommendation [C(2006)184].

Source: Answers from OECD and Partner delegations

48.   Examining the qualitative responses, we observe that the Recommendation is seen to have a strong influence in the following types of initiatives:

- Strategies, policies and laws focused on open access to data in research
- Policies and strategies addressing knowledge based economy and society
- Studies and guidelines on research data management

49.   The Recommendation is likely to have less impact in the following situations:

- Broader open government and public sector information initiatives, which do concern data from public research, but have little or no specifics on this sector

---

[5] It should be noted that the questionnaire concerns a decade of policy implementation, and that in some cases the respondents may not be familiar with the full history of implementation

- Initiatives which deal with data collected by private sector entities
- In some countries, focus was given on other international guidance, such as the European Commission policies (which have been influenced by the OECD Recommendation C(2006)184, for example (EC, 2012)), or Open Government Partnership.
- Initiatives in partner economies, who have not yet adhered to the Recommendation (but are starting to take it into consideration as a policy reference)
- In a few cases, lack of awareness of the Recommendation (even among Member States)

50.	Respondents also mentioned that although the Recommendation remains an important reference, there are also other international guidelines which are being taken into account, such as the European Commission 2012 Recommendation on access to and preservation of scientific information (EC, 2012), the European Open Science Cloud (ZBW, 2016) and the Amsterdam call for open science (Amsterdam Call for Action on Open Science, 2016).

Figure 3.5. Influence of the Recommendation per country
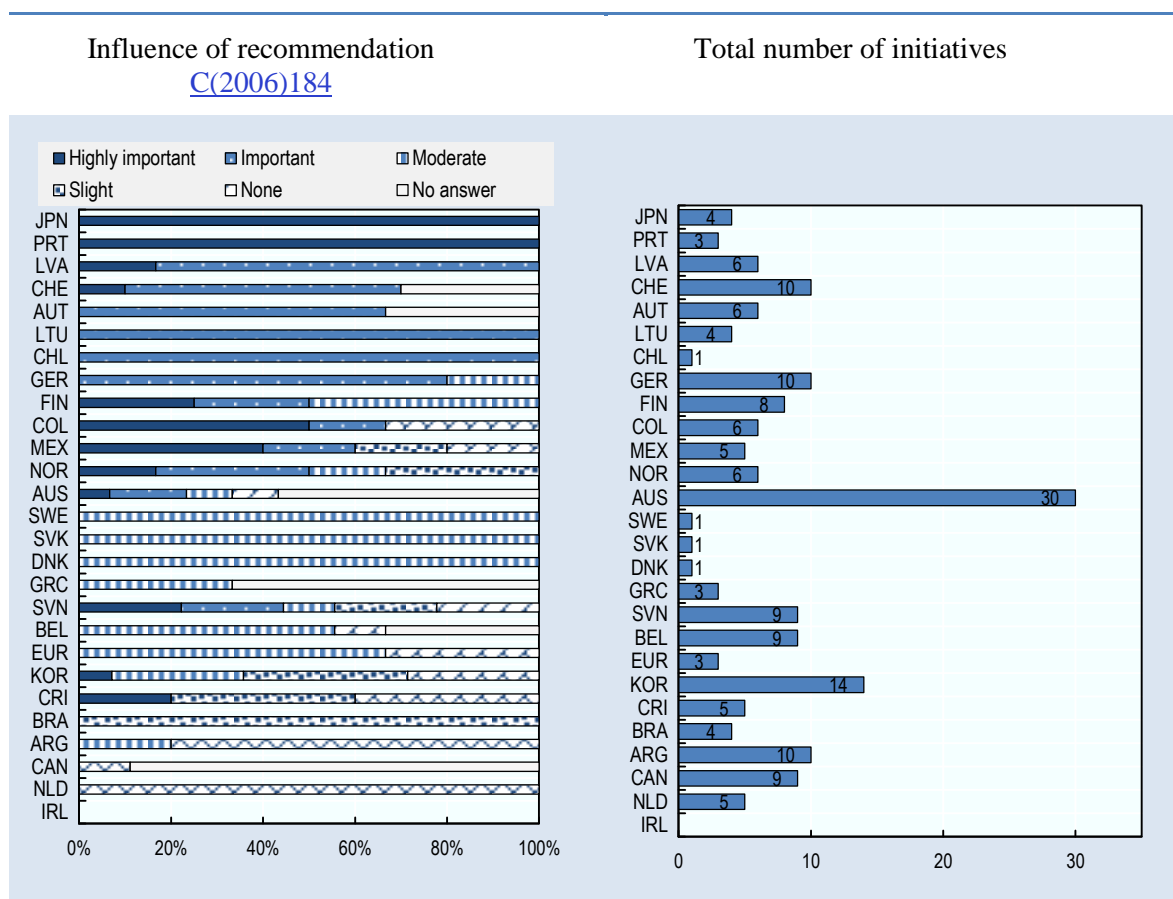
Note: The graph on the left represents the proportion of initiatives with the respective level of influence of the Recommendation. The graph on the right represents the total number of initiatives per country. Countries are ordered according to a weighted average score computed from the responses on a Likert scale: (1='none'; 2='slight'; 3='moderate'; 4='high'; 5='very high').

Source: Answers from OECD and Partner delegations Assessment of the objectives of the OECD Recommendation [C(2006)184]

51.   We also see a strong dispersion of the average influence of the Recommendation per country (Figure 3.5.).

52.   The stated aims and objectives of the Recommendation for the advancement of the data access agenda include:

   i.   promoting a culture of openness and sharing of research data among the public research communities within Member countries and beyond;
  ii.   stimulating the exchange of good practices in data access and sharing;
 iii.   raise awareness about the potential costs and benefits of restrictions and limitations on access to and the sharing of research data from public funding;
  iv.   highlighting the need to consider data access and sharing regulations and practices in the formation of Member countries' science policies and programmes;
   v.   providing a commonly agreed upon framework of operational principles for the establishment of research data access arrangements in Member countries;
  vi.   offering recommendations to Member countries on how to improve the international research data sharing and distribution environment.

53.   Respondents were asked to assess the relevance of each of these objectives, as well as the current policy effort related to those objectives. The results are shown in Figure 3.6.

54.   Overall, the relevance of the objectives is judged between 3.9 and 4.5 on the Likert scale, meaning that the relevance is high, tending towards 'very high' in some cases. Concerning policy effort, the average is between 3 and 4, ranging from moderate to high.

55.   The highest relevance is assigned to objective 2 concerning good practice sharing. However it is interesting to note that the policy effort dedicated to this objective is by far not the highest, with an average closer to 'moderate' than 'high'. Respondents report that good practice sharing occurs mostly in seminars and conferences, and some countries report the establishment of communities of practice as well as membership in international fora. However, many respondents report insufficient structured institutional or policy effort in this respect.

56.   The objective which is seen as commanding the highest policy effort is objective 1: promoting a culture of openness and data sharing. This is the core objective of most open access to data access initiatives. Some respondents (Japan, Korea) point to the differences across disciplines in the spirit of openness, with physics and biomedical sciences leading the way.

Figure 3.6. Relevance and current policy effort dedicated to the achievement of the objectives
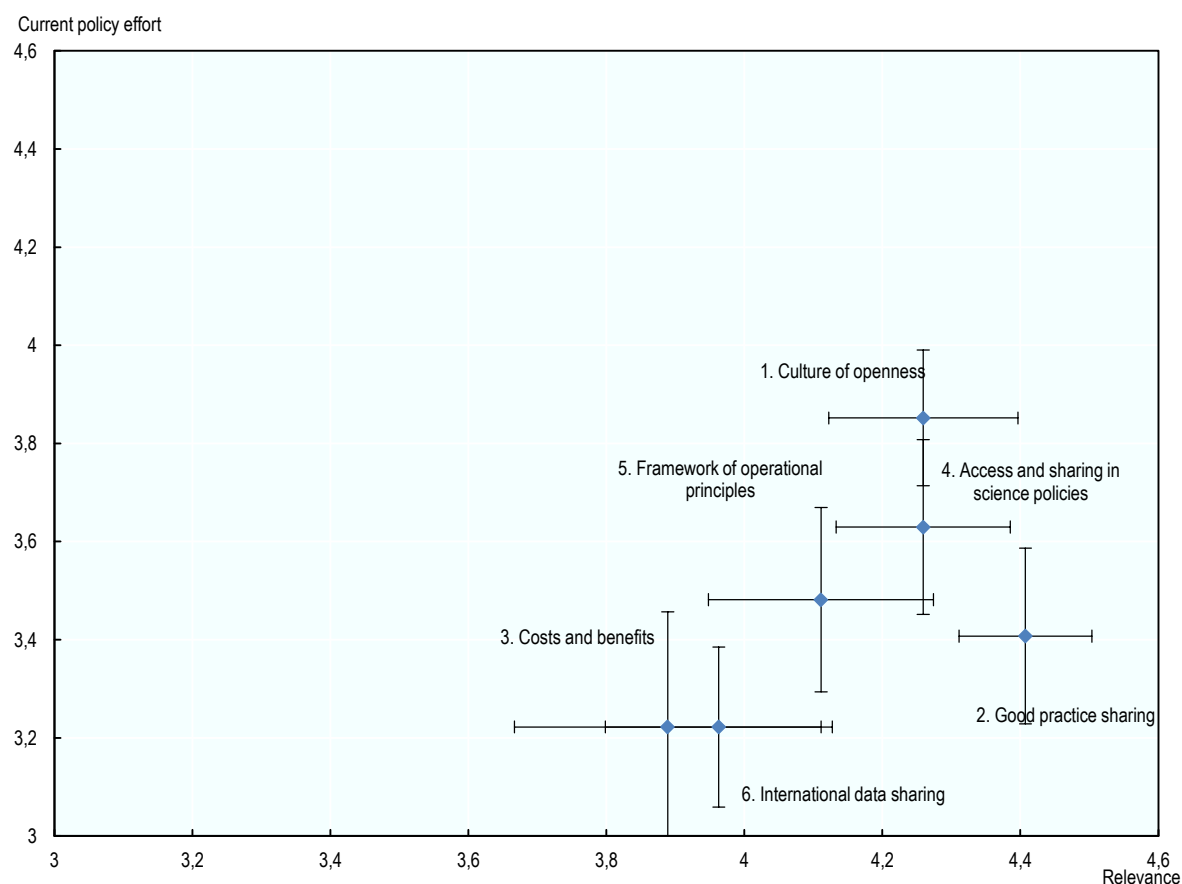
Note: An average score was computed from the responses on a Likert scale: (1='none'; 2='slight'; 3='moderate'; 4='high'; 5='very high'). The average gives equal weight to countries, i.e. if several assessments are submitted by country, those were averaged out first, and the average score of the country was used for the overall average. The error bars show the statistical error on the mean score, at 68% confidence level.

Source: Answers from OECD and Partner delegations

57.   Objective 4 concerning the inclusion of access to data and sharing in science policies is also ranked highly, both in relevance and policy effort. Some countries such as Finland and the Netherlands have implemented comprehensive open science policies, while others include data access aspects in individual policies, including funding regulations (such as the Research Council of Norway). It must be noted however that open access to data is not always specifically addressed in *science* policies, as reported by Denmark, Korea and Canada (in Canada those policies are dealt through the Open Government Partnership, a Tri-Agency Open Access Policy on Publications and Statement of Principles on Data Management).

58.   The framework of operational principles (objective 5) comes next in the order of priority. In this respect, respondents quote the formulation of the FAIR[6] principles (Wilkinson, Dumontier, Aalversberg, Appleton, & Axton, 2016) as a good basis, but with further efforts needed to operationalise those principles. Canada has issued a specific guidance document for releasing scientific data, and is addressing the issue of licencing under an Open Government licence. The European Commission has published the Guidelines to the Rules on Open access to Scientific Publications and Open access to Research Data in H2020. (EC, 2016), and requires a Data Management Plan as a key element for good data management in research projects.

59.   Awareness about potential costs and benefits of limitations on access to data sharing (objective 3) ranks relatively low, compared to other objectives, especially considering the policy effort which is seen as moderate. Respondents agree that the objective is important, but policies in place are relatively limited. The EC is ready to run a study on the costs of non-FAIR data to the EU Economy. The study will also estimate costs and benefits for implementing the FAIR principles. Moreover, a number of pertinent research projects are being funded via H2020.

60.   Improvement in the international data sharing and distribution environment also ranks relatively low. Countries report participation in many international fora, such as the Research Data Alliance, the European Science Cloud and the GO FAIR initiative, the G7 Working Group on Open Science, the Group of Senior Officials on Global Research Infrastructure, the Global Science Forum, and others. However, data access is primarily addressed at national level.

3.4. Assessment of the OECD Principles for Access to Research Data

61.   The OECD Recommendation C(2006)184 is complemented by the OECD Principles and Guidelines for Access to Research Data from Public Funding (OECD, 2007). Those principles are: A. Openness; B. Flexibility; C. Transparency; D. Legal conformity; E. Protection of intellectual property; F. Formal responsibility; G. Professionalism; H. Interoperability; I. Quality; J. Security; K. Efficiency; L. Accountability; M. Sustainability (Box 3.1).

62.   Respondents gave their view on the following:

- The relevance of the Principle
- The need for the principle to be included in the potential new overarching recommendation, as opposed to just keeping it within the current instrument
- Comment on issues related to implementation of the principle, and potential new focus dues to the evolution of the needs of stakeholders and policy makers.

---

[6] FAIR stands for Findability, Accessibility, Interoperability, and Reusability

Box 3.1. OECD Principles and Guidelines for Access to Research Data from Public Funding

Ministers of science and technology of OECD countries met in Paris in 2004 and adopted a Declaration on Access to Research Data from Public Funding. Subsequently, the CSTP proceeded to develop and approve set of principles and guidelines in October 2006, and then endorsed by the OECD Council. The CSTP subsequently reviewed progress by countries in the implementation of the principles in 2009 [DSTI/STP(2009)3].

The Principles can be summarised as follows:

• Openness – Open access to research data from public funding should be easy, timely, user-friendly and preferably Internet-based.

• Flexibility – Flexibility requires taking into account the rapid and often unpredictable changes in ICTs, the characteristics of different research fields, and the diversity of research systems, legal frameworks and cultures of each member country.

• Transparency – Information on research data and data-producing organisations, documentation on the data and conditions attached to the use of data should be internationally available in a transparent way, ideally through the Internet.

• Legal conformity – Data access arrangements should respect the legal rights and legitimate interests of all stakeholders in the public enterprise. Restrictions on access may be for reason of: national security, privacy and confidentiality, trade secrets and intellectual property rights, protection of rare, threatened or endangered species, legal processes.

• Protection of intellectual property – Data access arrangements should consider the applicability of copyright and other intellectual property laws that may be relevant to publicly funded research databases (as in the case of public-private partnerships).

• Formal responsibility – Access arrangements should promote rules and regulations regarding the responsibilities of the parties involved; should be developed in consultation with stakeholders; should be responsive to factors such as the characteristics of the data, e.g. their potential value for research purpose. Data management plans and long-term sustainability should also be considered.

• Professionalism – Institutional arrangements for the management of research data should be based on the relevant professional standards and values embodied in the codes of conduct of the scientific communities involved.

• Interoperability – Access arrangements should consider the relevant international data standards.

• Quality – The value and utility of data depend on the quality of the data themselves. Particular attention should be paid to ensuring compliance with explicit quality standards.

• Security – Attention should be devoted to supporting the use of techniques and instruments to guarantee the integrity and security of research data.

• Efficiency – One of the central goals of promoting data access and sharing is to improve efficiency of publicly funded scientific research so as to avoid expensive and unnecessary duplication of effort. This also involves cost and benefit analysis to define data retention protocols; the engagement of data management specialist organisations; and the development of new reward structures for researchers and database producers.

• Accountability – The performance of data access arrangements should be subjected to periodic evaluation by user groups, responsible institutions and research funding agencies.

• Sustainability – Due consideration should be given to the sustainability of access to publicly funded research data as a key element of the research infrastructure.

Source: (OECD, 2007)

63.   The results are summarised in Figures 3.7 and 3.8. Overall, we see a clustering of the Principles into 2 groups:

- A group with high to very high relevance, and high to very high need to be included in the possible future overarching recommendation, including openness, quality, security, interoperability, transparency, sustainability and legal conformity.
- Another group with moderate to high relevance, and also moderate to high need to be included in the overarching recommendation, including IP protection, flexibility, professionalism, efficiency, formal responsibility and accountability.

64.   Some respondents distinguish between core principles which enable the behaviour which is sought, and should thus be included in the overarching recommendation, and enablers of such behaviour, which could be kept in the current recommendation, or even addressed within guidelines.

Figure 3.7. Assessment of the principles concerning access to research data from public funding
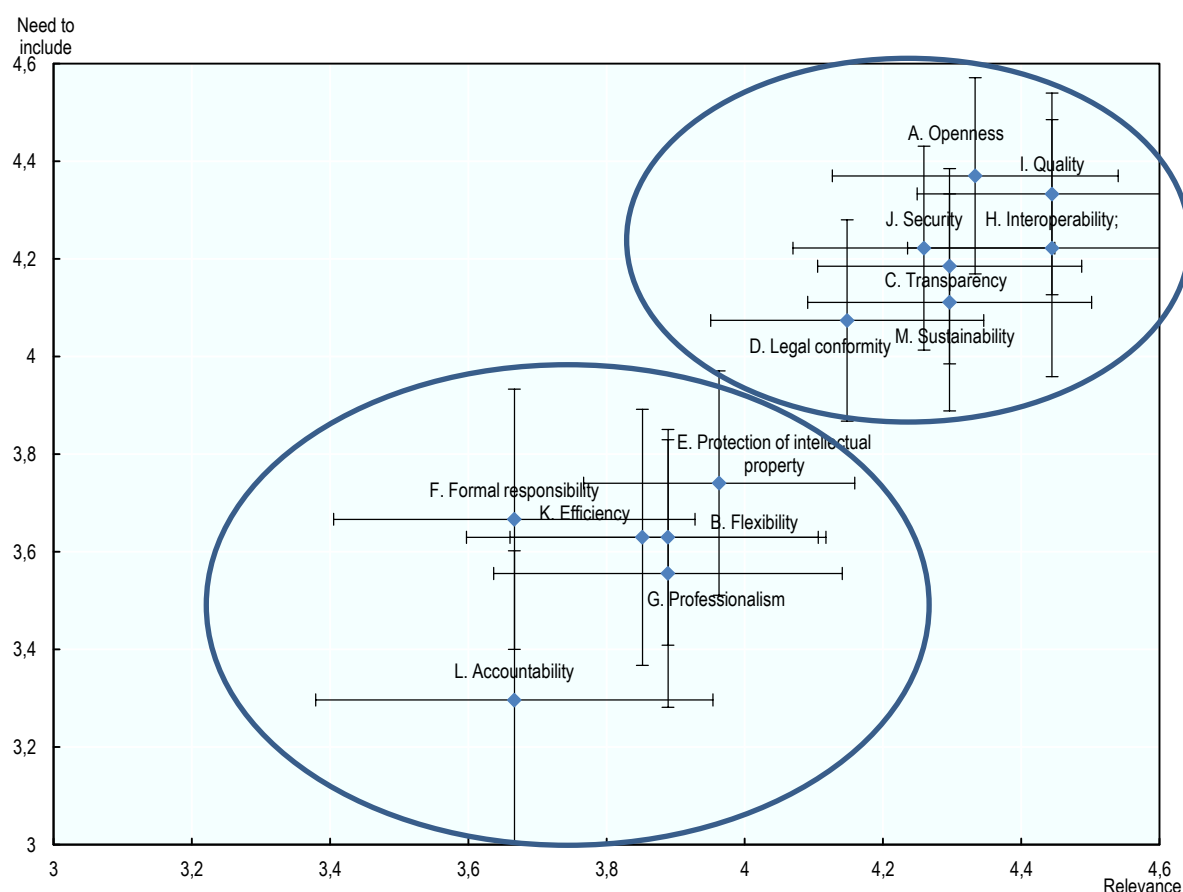
Note: An average score was computed from the responses on a Likert scale: (1='none'; 2='slight'; 3='moderate'; 4='high'; 5='very high'). The error bars show the statistical error on the mean score.
Source: country responses to the questionnaire
Source: Answers from OECD and Partner delegations.

Figure 3.8. Assessment of the principles concerning access to research data from public funding

Relevance of principle

Need to include in overarching recommendation
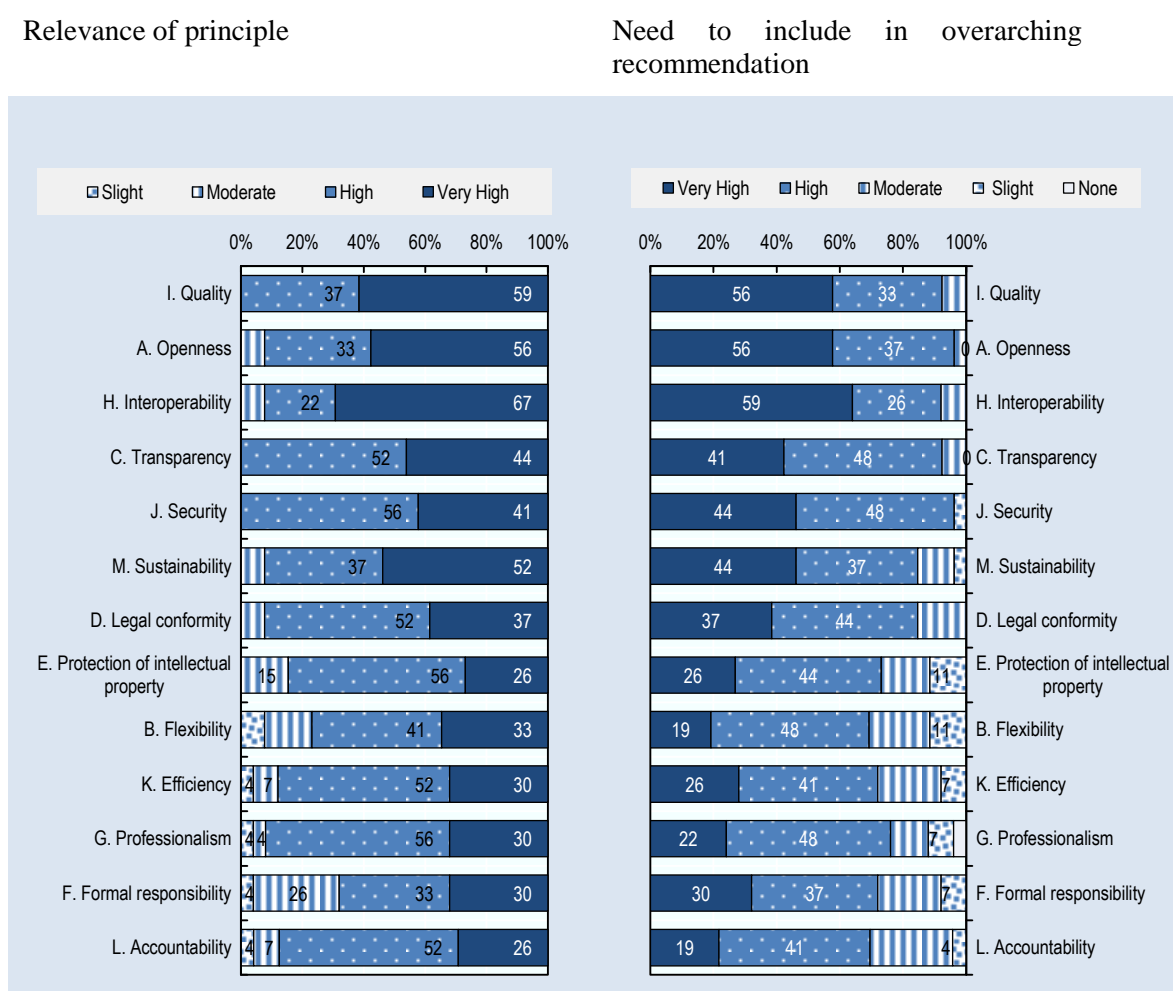
Notes : Principles are ordered according to a weighted average score computed from the responses on a Likert scale: (1='none'; 2='slight'; 3='moderate'; 4='high'; 5='very high').
Source: Answers from OECD and Partner delegations.

65.   ***Quality*** ranked among the highest both in relevance and with respect to the need to include in any future recommendation. This principle comprises aspects of quality control through peer review, documentation of the origin of sources, links to original research materials and datasets, and data citation practices. There is a sentiment that more needs to be done on quality assurance in general, with clear quality standards which could be explicit and verifiable, and captured quantitatively where possible. The new Recommendation could provide guidelines for determining data quality and a standard for labelling datasets with a confidence value.

66.   ***Openness*** was defined in the 2007 Principles as "access on equal terms for the international research community at the lowest possible cost, preferably at no more than the marginal cost of dissemination. Open access to data to research data from public funding should be easy, timely, user-friendly and preferably Internet-based."

67.   It was quoted as one of the most relevant principles by respondents, and clearly one that should be built into the overarching recommendation. EU promotes 'open by default', efficient and cross-disciplinary research data environment, allowing for proportionate limitations only in duly justified cases of personal data protection, confidentiality, IPR concerns, national security or similar (e.g. 'as open as possible and as closed as necessary'). Australia includes open access to data within its Open Government National Action Plan. However, respondents emphasise the need to maintain limits of openness in case of legitimate reasons to keep data closed, and warn against potential disincentive to data acquisition under an 'open by default' policy. Further, the need for a 'cultural change' among researchers is emphasised, in order to achieve more openness.

68.   ***Interoperability*** was cited among the most relevant principles. In the 2007 Principles it comprised the explicit mention of standards used, adoption of best practices among professional organisations dealing with data collection and preservation, and taking into account the development of more general ICT standards.

69.   While there is progress in interoperability within disciplines, cross-disciplinary interoperability remains to be developed. Interoperability is also the 'I' of the FAIR principles put forward by the EU. It is proposed that the recommendation provides guidance on ontologies and translation. The establishment of supranational open science clouds such as the European, Australian, African and NIH Commons in the U.S. will provide a leap in findability for scientists in those regions. Going a step further, the interoperability between those clouds needs to be established to develop global access.

70.   ***Transparency*** ranks very high in terms of relevance and inclusion in the overarching recommendation. In the 2007 Principles it was defined through the following aspects: (i) information on data-producing organisations and their holdings, documentation on available data sets; (ii) dissemination of information on research data policies to stakeholders; (iii) agreements on standards for cataloguing data, and application thereof; (iv) information on data management and access conditions to be shared among data archives and data producing institutions.

71.   ***Security*** is another highly ranked principle. In the 2007 Principles, security was meant to encompass both integrity (completeness and absence of errors) and security (protection against loss, destruction, modification and unauthorised access).

72.   Respondents see security as essential to foster trust. It is proposed that the new recommendation include guidelines for repositories on data provenance and versioning to address data integrity. Benefits, profits, advantages as well as possible disadvantages and risks of open access to data should be addressed in order to look for the ways how to avoid them.

73.   ***Sustainability*** is seen as highly relevant and highly needed in the overarching recommendation. Sustainability has to be ensured throughout evolutions of technology and standards. Therefore datasets need to be preserved across technology changes. Expectations of users need to be set however, to ensure that they understand the scope and re-use potential in particular of older data which do not conform to the latest data standards.

74.   In this respect, regular evaluations of electronic infrastructures and services are needed. The data lifetime and deletion policy should be specified.

75.   *Legal conformity* under the 2007 principles includes aspects of national security, privacy and confidentiality, intellectual property rights, protection of biodiversity and legal process. Respondents agree that this is one of the high priority principles to be included in the overarching recommendation. Privacy is currently mentioned under this principle, but may need an enhanced focus going forward.

76.   *Intellectual property (IP) protection* has a high relevance, but the need to include it in the overarching is moderate to high. The 2007 principles recommend considering the applicability of copyright or other IP laws, including the cases of data which are the result of public-private partnerships, and should nevertheless facilitate open access to data if possible, for public research, or other public-interest purposes.

77.   Respondents agree that when partnering with private parties, the public good nature of publicly-funded research data should not be compromised - meaning that it must be freely available for the use of all - and private parties cooperating with the public sector must acknowledge this special status and abide by these principles. They also see IP protection of public research as a complex issue, and suggest that data from public research be protected under Creative Commons, allowing data to be open, but restricting any derivative or commercial use.

78.   Other principles are seen as having moderate to high relevance, and the need for including in the overarching recommendation is equally moderate to high:

- *Flexibility* is considered with respect to technology evolution, evolving needs of scientific disciplines, legal systems and culture of each country.
- *Efficiency*, covering cost effectiveness, cost-benefit analysis, engagement of data management specialists as appropriate, as well as reward structures for researchers.
- *Professionalism* is defined in the 2007 Principles as the use of codes of conduct in order to simplify the regulatory burden on access to data, inducing mutual trust between researchers, institutions and other stakeholders involved, as well as clear rules for temporary exclusive use of data.
- *Formal responsibility* including explicit rules and regulations regarding the responsibilities of parties involved in data related activities, pertaining to authorship, producer credits, ownership, dissemination, usage restrictions, financial arrangements, ethical rules, licensing terms, liability, and sustainable archiving. If used in the overarching recommendation, should provide guidelines on the handover of data curation responsibility from a national laboratory to the institution of the principal investigator. It is also noted that in some countries formal agreements on terms of access and use of data are absent, reducing the motivation for researchers, and increasing personal burden for them.
- *Accountability*, meaning evaluation according to criteria of overall public investments, management performance of data collection and archival agencies, extent of re-use of data sets, knowledge generated from re-use of existing data, and targeted foresight for future needs related to data preservation and re-use.

79.   Respondents were also asked to provide additional principles not covered in (OECD, 2007). Some of the principles quoted include:

- *Discoverability* is crucial if data is to be re-used. Therefore there is a need to establish ontologies and appropriate semantics to enable scientist from different scientific domains to find relevant data sets.
- *Machine-readability* should be the norm, to facilitate treatment of large quantities of data
- *FAIR data principles*. Findable, accessible, interoperable, re-usable. A strong link needs to be established between the new OECD recommendation, and those principles which are spreading to a large number of countries.
- *Strict rules for the financial support* of open access to data: (i) responsibility for financing at international, national, regional and institutional level; (ii) definition of roles between financiers, operators and users; (iii) supervision mechanism for compliance with open access to data rules, including legal sanctions for non-compliance.
- *Definition of responsibility and ownership*, including legal and ethical issues
- *Explicit recognition and reward system* for data authorship
- *Publicly funded research data treated as Commons*. Licencing under Creative Commons could be used to provide a framework which ensures openness while restricting re-use as needed
- *Setting an embargo period* for the exclusive use of data. The embargo could vary according to the nature of the output, and provide reassurance to authors.
- *Implications of blockchain technologies* on open access to data should be investigated, since opinions of relevance are very different in the open science field. Blockchain technology is mentioned as a potential tool which could help re-allocate private monopoly rents from innovation back into the network of public collaborative science and innovation (Soete, 2016).

80.   Some respondents suggested re-organising principles into two groups: one for core principle, and another one for enablers. In addition, it was pointed out that some principles presented a degree of overlap (e.g. between interoperability and quality), to be avoided in the future recommendation. In general, the feeling is that there may be too many principles which could be streamlined for the future.

3.5. Addressing challenges in access to research data

81.   Progress has been achieved in the decade since the Recommendation C(2006)184 was issued. Data associated with publications is now expected to be made available, and the 'reproducibility' agenda is getting more attention, e.g. in clinical trials. A number of research funders are now mandating open access to data.

82.   Nevertheless, data sharing still seems to be less widespread than would be expected, and is still limited to a small number of research fields. Reasons for this seem to be disincentives to sharing research data, such as lack of reward or credit for sharing data, the substantial effort needed to upload and maintain data in a form which is usable by others, risks of misinterpretation or misuse, control of intellectual property and personal data protection issues, i.e. the need of anonymising data samples. In addition, demand for shared data seems limited to a few scientific disciplines (OECD, 2015)

83.   The OECD Global Science Forum has identified nine challenges related to data-driven and evidence-based research in social and economic sciences (OECD, 2013). A synthesis of those findings can be found in Box 3.2.

Box 3.2. Challenges to data-driven and evidence-based research identified in social sciences

A. Infrastructure and skills (or lack thereof) at country level:

A.1 Lack of data management planning to make datasets available for reuse

Researchers have a responsibility to ensure that they have the skills and the resources necessary to ensure that the data they use for research are available for re-use and that plans are made to this effect before they engage in research.

[The recommendation in (OECD, 2013) was for national funding agencies to ensure that new research awards have accompanying data management plans and assign resources for this purpose. Agencies should co-operate at both national and international levels to share this information, publishing details about data and metadata to be created and plans for their preservation in formats that make this information readily accessible to the research community.

The international community should agree on a standard semantic dictionary describing common elements of a data management plan to facilitate the discovery and use of the information contained in such plans.]

A.2 Investments in personnel and infrastructure needed for data creation and curation

Not all countries have invested in the skills, resources or infrastructure required to curate important datasets. This places such data at risk of loss, minimises the potential for their re-use and precludes their inclusion in an evolving global data ecosystem.

[The recommendation in (OECD, 2013) was for research communities without institutional support for data curation or supporting infrastructure to conduct an assessment of their national needs and assets in this area that will contribute to national plans of action. Working with researchers in such countries, established science data archives should assist them by developing an assessment instrument and providing expert advice in preparing plans.]

B. Legal and regulatory barriers/challenges at country level:

B.1 Lack of information on what data exists and lack of adoption of international standards for data documentation

While many countries have vast amounts of more traditional forms of administrative, survey, and census data collected by and held by national statistical agencies and government departments, knowledge about the existence of such data as micro-data records is a precondition for the efficient and effective planning of international research.

[The recommendation in (OECD, 2013) was for national statistical organisations and international organisations to ensure that all data they collect and process are documented to agreed and common standards. Such documentation should be easily discoverable on their websites.]

B.2 Individual privacy issues, absence of a recognised framework governing the use of personal data

Personal data, such as medical records or social networking data, can provide insights into the human condition. However, the use of those data as research resources may pose risks to individuals' privacy, particularly in case of inadvertent disclosure of the identities of the individuals concerned. There is a need for greater transparency in the research use of new forms of data, maximising the gains in knowledge derived from such data while minimising the risks to individuals' privacy, seeking to retain public confidence in scientific research which makes use of new forms of data.

[The recommendation in (OECD, 2013) was for research funding agencies and data protection authorities to collaborate and develop an internationally recognised framework code of conduct covering the use for research of new forms of personal data, particularly those generated via

network communication. This framework, built on best practice procedures for consent from data subjects, data sharing and re-use, anonymisation methods, etc., could be adapted as necessary for specific national circumstances.]

B.3 Legal, cultural, language and proprietary rights barriers

Barriers to access to data hinder national and cross-national collaboration which could exploit their research value. These barriers relate to a variety of obstacles (legal, cultural, language, proprietary rights of access) all of which have to be identified and removed if public-private and cross-national research is to be promoted.

[The recommendation in (OECD, 2013) was for national statistical agencies to establish mechanisms to improve access for comparative research by the research community to social science microdata in their possession. These same agencies should collaborate internationally to further their efforts in this area, particularly with respect to the use of novel methods to facilitate secure access to such data where there is a risk of disclosure of identities. Leading international organisations were also invited collaborate in the formulation of a strategic approach towards the removal of obstacles to improved access to and sharing of micro-data in their possession and should provide a coordinated plan for the creation of data discovery and data management tools on their websites.]

C. Researchers' incentives and careers at country level:

C.1 Incentives for researchers to ensure effective data sharing

Data sharing, including the creation of appropriate metadata to international standards is fundamental to the process of scientific enquiry. Researchers need incentives to ensure effective data sharing.

[The recommendation in (OECD, 2013) was for research funding agencies to collaborate at the international level to ensure that a common system is adopted for referencing datasets in research publications. They should also ensure that the intellectual effort required for the creation and sharing of data is recognised in their evaluation of research activities.

Publishers of research should be encouraged to adopt guidelines for publications, stipulating that a common and internationally agreed referencing system for datasets is used within all scientific publications that have made use of data.

The employers of researchers should recognise the intellectual efforts that have been made by researchers who generate significant data resources. This could be reflected via merit awards, promotions and other ways which acknowledge the professional contributions that have been made.]

D. Data quality and characteristics:

D.1 Reliability, statistical validity and generalisability of different data sources

Massive amounts of digital data are being generated at unprecedented scales and velocity, much of it from new sources such as the internet. The reliability, statistical validity and generalisability of new forms of data are not well understood. This means that the validity of research based on such data may be open to question.

[The recommendation in (OECD, 2013) was for national research funding agencies to collaborate internationally to provide resources for researchers to assess the research potential and develop new methods to understand the opportunities and limitations offered by new forms of data to address important research areas.]

D.2 Need for greater harmonisation of social sciences data sources across countries

Comparative research in the social sciences, based upon data pertaining to different countries and regions, is an essential part of the research process and is set to become increasingly important.

Without collaboration and sharing of experience between countries in the development of comparable data resources, the full benefits from comparative research will not be achieved.

[The recommendation in (OECD, 2013) was for national and international statistical agencies to strengthen the efforts they are making to harmonise social and economic data at the international level, seeking to prioritise these activities in specific areas.

Researchers involved in the creation and maintenance of data resources designed specifically for comparative research and those funding such research should collaborate to provide mechanisms to foster an integrated approach to data design and harmonisation, access and sharing as stated in the OECD (2007) Council Recommendation concerning Access to Research Data from Public Funding.]

D.3 Increasing need for international coordination

The drive to address what is increasingly an international research agenda requires the use of existing capacity to its full potential.

[The recommendation in (OECD, 2013) was for mechanisms to be established which build upon and enhance further the efforts being made by producers of data (e.g. official data producing agencies, businesses, researchers) and the users of data (e.g. researchers, policy-makers) to share expertise, knowledge and resources, particularly in the areas of data access, linkage and integration and analysis.]

D.4 Increasing need for interdisciplinary coordination for global challenges

The drive to address what is increasingly an interdisciplinary research agenda, in particular when addressing grand challenges such as those represented by the 2030 Sustainable Development Agenda.

Source: adapted from (OECD, 2013)

84.   Respondents were asked to assess the relevance of each of these objectives, as well as the current policy effort related to those objectives. The results are shown in Figure 3.9.

Figure 3.9. Assessment of challenges related to data-driven and evidence-based research

Assessment of challenges originally identified by the Global Science Forum for the domain of social sciences
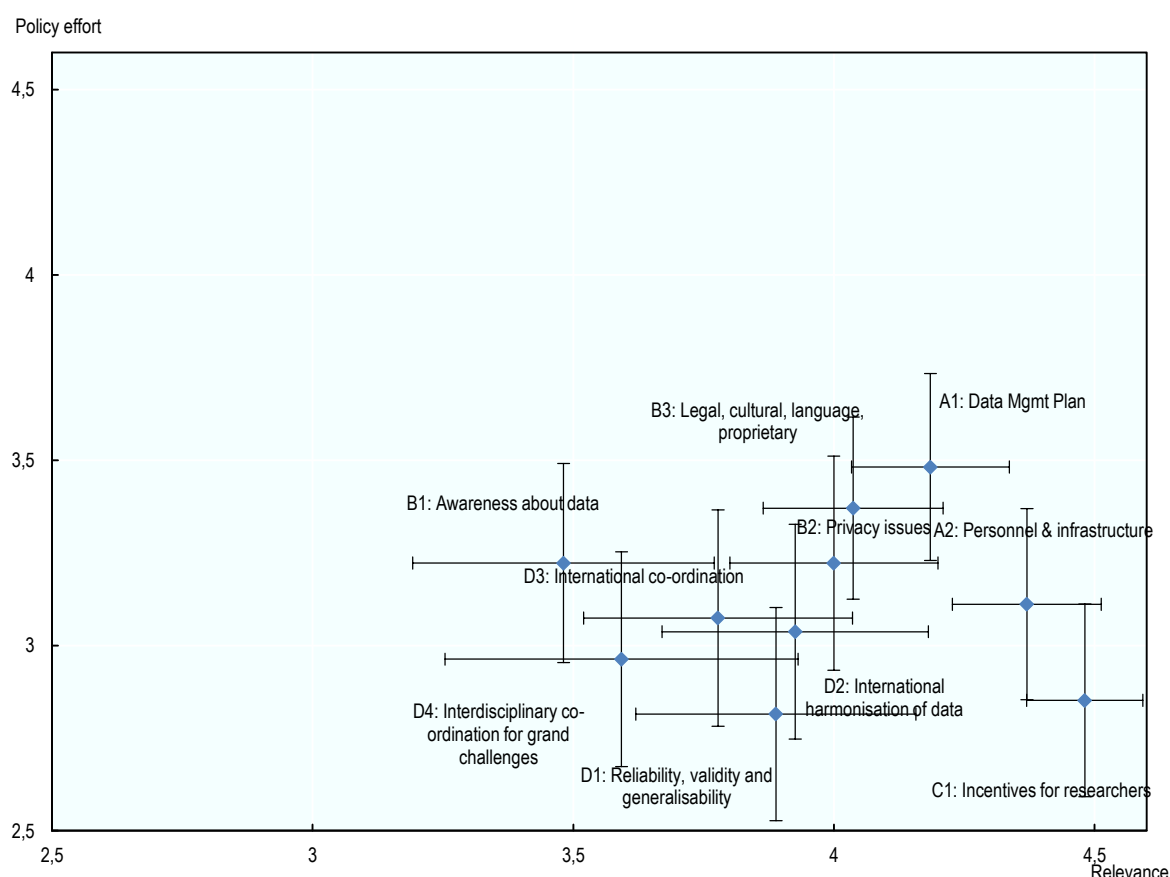
Note: An average score was computed from the responses on a Likert scale: (1='none'; 2='slight'; 3='moderate'; 4='high'; 5='very high'). The errors bars show the statistical error on the mean score.
Source: Answers from OECD and Partner delegations.

The most relevant of all challenges is C1 which concerns ***researcher's incentives*** to ensure effective data sharing, but there is consensus that policy efforts to overcome that challenge are still weak. Cultural change is a long process. Perceived barriers and risks of open access to data need to be counterbalanced by appropriate acknowledgement and reward systems. Data citation does not seem to have been widely implemented, and the prerequisites for it are questioned by some respondents. Some countries also shared the view that Open Science should be embedded in evaluation systems to ensure that researchers who provide high quality research data should be rewarded (e.g. Brazil, Canada, EC, Japan, Netherlands). Training in data literacy and data management is also an important aspect. An interesting initiative was mentioned by Australia whose

Department of Employment organises a yearly GovHack competition to mashup, reuse, and remix government data. Raising awareness and communicating about a 'cultural shift' towards data stewardship and sharing throughout research data lifecycle.

85. The strongest policy effort goes into challenge A1 concerning the lack of ***data management planning*** to make datasets available for reuse. Most governments report activity on addressing the issue by making recurrent research funding contingent on data sharing and data management plans (e.g. in Australia, Canada, Netherlands, Sweden). Adequate training in data management planning is also quoted as an important aspect.

86. Another challenge of very high relevance is A2, concerning ***investments in personnel and infrastructure*** needed for data creation and curation. An important initiative in this respect is the GO-FAIR initiative, led by Germany and the Netherlands, which is a proposed approach towards the establishment of the European Open Science Cloud. The initiative rests on three pillars: (i) Go-Change to foster culture change, open science promotion and reward systems; (ii) Go-Train for education and training; and (iii) Go-Build for technical infrastructure (ZBW, 2016).

87. ***Individual privacy issues*** (Challenge B2) also command high policy effort, and are seen as highly relevant. Governments strive to ensure a balance between maximising data sharing whilst ensuring the privacy and security of information, in particular the use of "anonymised", "non-sensitive" data. This issue is taken up in many projects and policies, at the level of government as well as funding agencies, data centres. In an effort to harmonise data protection across Europe, the EC has adopted a General Data Protection Regulation (GDPR) (EC, 2016) which will enter in force in spring of 2018, and there are concerns that those stricter rules might negatively affect an ever-more collaborative and data-intensive scientific research sector.

88. Equally relevant, but with slightly less policy effort dedicated to it is challenge B.3 concerning ***legal, cultural, language and proprietary rights barriers***. In this respect, some countries have modified copyright law accordingly. Respondents also point to the necessity of clarifying and addressing the legal uncertainty of open access to research data, as well as the correct legal implementation of the FAIR principles. Issues of ownership must also be addressed, particularly where institutions have created services and resources. Australia also points to the necessity of harmonising legislation across data custodians, who often operate under varying legal frameworks that govern the collection and use of sensitive data.

89. Still highly relevant, but with only moderate policy effort is the issue (D2) of greater ***international harmonisation of data sources across countries***. Respondents reported addressing the issue within the Research Data Alliance and also through application of FAIR principles which should be the future reference. Respondents agree that more needs to be done in this respect.

90. Even less policy effort seems to be dedicated to the very important challenge of ***reliability, statistical validity and generalisability of different data sources*** (D1). The respondent from the EC proposes to put in place an accreditation or certification mechanism based on agreed processes to ensure FAIR compliance, and an accreditation or certification body to maintain an up-to-date and accessible catalogue of certified repositories.

91. Finally, in the group of less relevant challenges we find:

- D3: *International co-ordination*. Countries report participation in the Research Data Alliance, CODATA, DDI Alliance as well as in networks of repositories to work on this issue (e.g. the Confederation of Open Access Repositories, LA Referencia in Latin America). Such for a should serve to define global standards for the implementation of FAIR principles

- D4: Increasing need for *interdisciplinary coordination for global challenges*. Respondents propose to establish cross-disciplinary agreements and protocols which will lead to specific standards, inspired directly by relevant domain-specific needs. Variations across scientific disciplines and their specific efforts of making research data open and FAIR should be respected. Development of best practice projects of interdisciplinary co-ordination could demonstrate benefits in addressing global challenges.

- B1: *Lack of information on what data exists* and *lack of adoption of international standards for data documentation*. Respondents recommend raising awareness about standards published by the Research Data Alliance, better implementation of data management plans and a landscape analysis of data repositories. The respondent from the EC proposes the creation of catalogues for datasets, services and standards, based on machine readable metadata and identifiable by means of a common and persistent identification mechanism that will make research data findable. Ongoing work of the Global Science Forum on Open and Inclusive Science addresses the issue of data catalogues [DSTI/STP/GSF(2017)4].

92.   Respondents were also asked to provide additional challenges not covered in (OECD, 2013). Some of the challenges quoted include:

- *Measurement of the status quo of data access*. General and specific indicators need to be established to measure sharing and re-use of open access to data. Such measurement would: (i) enable demonstration of value added of open access to data; (ii) provide a basis for acknowledgement and reward to the researchers and institutions involved; (iii) help monitor quality and sustainability of the data sets

- *Large infrastructure solutions* for big data to be addressed nationally and internationally with adequate governance arrangements. Existing physical infrastructures need to be strengthened and new ones created to accommodate rapidly growing needs of big data. *Repositories* with tools for publishing datasets are preferable to read-only portals.

- *Data re-use, data portability and interoperability.* Physical infrastructures need to be complemented by internationally accepted and agreed-upon standards, which need to be widely disseminated. An overarching recommendation on enhanced access to data could ideally pave the way towards a more uniform political vision of these issues, in order to trigger the needed action at the national level.

- *Funding models*. The responsibility for data curation is implicitly transferred to the home institution of the researcher, which may not have appropriate repositories for that data type. This calls for the establishment of mechanisms for cross-institutional and cross-border use and the involved compensation of costs. Provision of open access to data by publishers at a cost is seen as problematic.

- *Cost-benefit analysis and priority setting of open access to data*. Scarcity of resources implies that not all data can be made openly accessible in the short term. Efforts should thus focus on open access to data to the data which are likely to provide the most impact, to the extent that such impact can effectively be

predicted (this is not always the case). The justification of investments in infrastructure needs to be made on the basis of the value expected from open access to data, which links to the issue of measurement and funding models. Linked to this is the issue of selection of data for long-term preservation which involves complex decision-making about the choice of priority data, and the timespan of the data preservation (some research communities make use of data which is even centuries old);

- *Operationalise FAIR principles* in a pragmatic and technology-neutral way, equally encompassing all four dimensions: findability, accessibility, interoperability and reusability. Apply FAIR principles to all digital research objects, incl. data-related algorithms, tools, workflows, protocols & services. A single window for access to all data would be desirable. Promotion of FAIR principles and associated FAIR services. These services also need to be sustainably maintained.

- *Statistical and methodological training* for using and interpretation of data, data management, as well as training for data standards. Increased data access and enhanced impacts from data will require new skills, and this fact should be given attention. Delivering the new skills will require actions from policy makers, data producers and users, and higher education institutions, in the form of co-operation and partnerships, training, new education programmes and curricula, etc. (and probably great opportunities for digital learning and MOOCs). The required skills are not only technical, there is probably a need for a wide range of skills from the domains of statistics, computer science, information science (data librarians etc.) as well as law and other social sciences. Many countries report limitation of current curricula in addressing those skills needs.

- *Build trust* between all stakeholders, e.g. scientific communities, e-infrastructures, research infrastructures, funders - "look outside the organisational boxes and work together". Open and FAIR access to research data to be integrated in the wider context of Open Science (as in e.g. the Dutch National Plan for Open Science, interconnecting Open access, Open data and reward systems).

- *Intellectual property of data.* Some publishers require data supporting the publication to be handed over to them. Others offer platforms for facilitating the research process, where everything, from annotations, methods, data and publications, can be performed or disseminated through them. In the short term, this is positive and enhances data accessibility, but there is no guarantee whatsoever that it will always be freely available.

- *Accessibility to content mining.* Due to the availability of research data through privately held platforms, automated content mining performed in these platforms is frequently hindered by their proprietors, with the excuse of technical limitations of the platforms.

- *Integrating open access to data within a broader Open Science approach,* including citizen science with citizens as both providers and users of data. Transpose primary research data and scientific information to make it intelligible to the general public as a debt to taxpayers and citizens

4. Public sector information in science, technology and innovation

93.   Beyond data from public research, respondents were asked about the relevance of broader sources of public sector information to scientists, and also the perceived barriers to accessing this information? The results are summarised in Figures 4.1 and 4.2.

94.   Overall, the relevance of public sector information is considered mostly as high to very high, while the barriers to accessing the information are seen as moderate to high.

Figure 4.1. Relevance of public sector information to public research
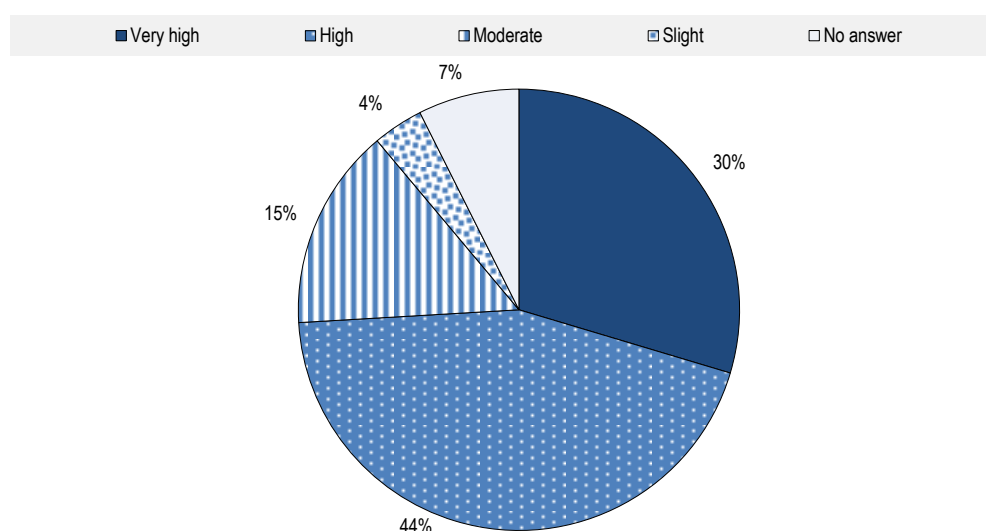
Source: Answers from OECD and Partner delegations.

Figure 4.2. Barriers accessing public sector information vs. relevance

Source: Answers from OECD and Partner delegations.
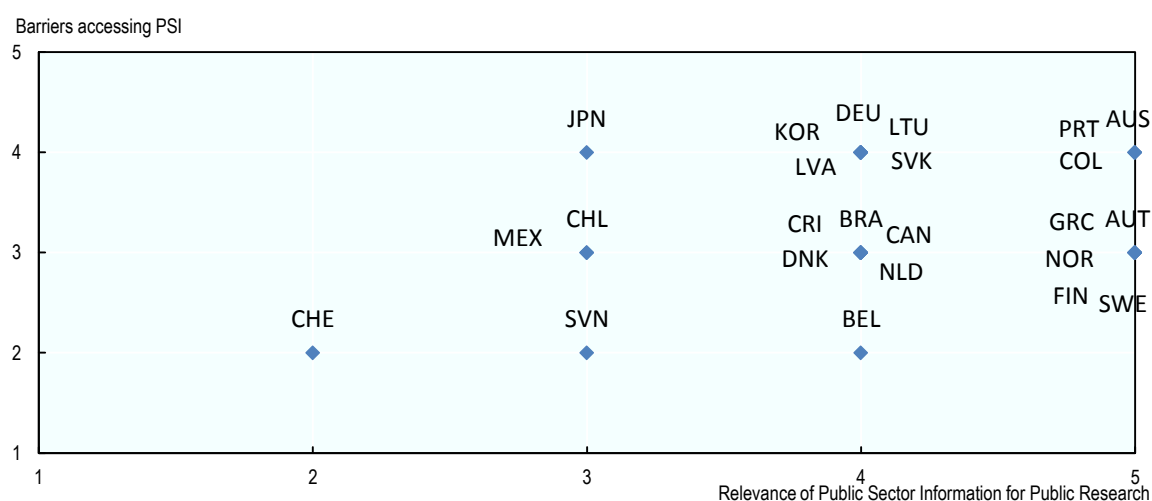
OPEN ACCESS TO DATA IN SCIENCE, TECHNOLOGY AND INNOVATION – INITIAL SURVEY FINDINGS

.

95.   Concerning overall issues to be addressed concerning access to public sector information for scientists, respondents have quoted the following issues:

- *Commitment of governments to release non-sensitive data by default.* Public institutions should adopt the "as open as possible, as closed as necessary" principles in regards to data; they should also seek to implement and operationalise the FAIR principles; and lastly, should incorporate into their mission the duty to provide and make freely available the data they generate or manage thus incorporating funding and operational provisions in their work plans for that effect.
- *Striking a balance between the benefits of sharing and linking of personal data against privacy and confidentiality considerations.* This balance is crucial in maintaining trust and thereby maximising the value of data.
- *Clear roles and administrative responsibilities* and arrangements at a whole of government level and agency level.
- *Collaboration across government, as well as with the research sector and private sector* in order to exchange good practices, harmonise data standards, explore synergies and enlarge the scope of open access to data and thus create more value for society as a whole.
- *Building institutional capabilities* at public agencies to manage, create, curate and reuse data.
- *Awareness raising and training for re-use of open access to data* among researchers, to enhance effective re-use of available datasets.
- *Clearly identified sources of funding* (e.g. for European Open Science Cloud initiatives)
- *Middleware* needed to connect between dispersed databases.
- *Emphasis on quality, rather than quantity of open access to data*

5. Conclusion

96.   In this paper we exploited the responses to the questionnaire on the usage of the Recommendation of the Council concerning Access to Research Data from Public Funding [C(2006)184], in view of further approaches to enhanced access to data.

97.   The Recommendation of the Council concerning Access to Research Data from Public Funding [C(2006)184], and the OECD Principles and Guidelines for Access to Research Data from Public Funding remain relevant and important references with respect to access to research data from public funding.

98.   Nevertheless, numerous other initiatives have appeared in the past decade. A prominent one is the FAIR principles initiative, supported by many policy makers around

the globe. Open science clouds are being established in Europe, Australia, U.S. (NIH Commons), and Africa, and since November 2016 a G7 working group on Open Science has been established to establish a common vision, create incentive and rewards systems and develop a federated international research data infrastructure. The Research Data Alliance is an important stakeholder with global reach, and significant achievements in the domain of interoperability and standards.

99.   While some of the objectives of Recommendation C(2006)184 are well addressed by policy makers, such as fostering the culture of openness and taking into account access and sharing within science policies, others have received less policy attention, such as sharing of good practices which is performed mostly in seminars and conferences, rather than through structured policy initiatives.

100.   With regards to principles, there is a group of high priority principles from the 2007 document which are still seen as highly important for inclusion in the overarching recommendation. These include:

 i.   Openness,
 ii.   Quality
 iii. Interoperability
 iv.   Security
 v.   Transparency
 vi.   Sustainability; and
 vii. Legal conformity.

101.   A number of additional principles were suggested by respondents, and could be included, for example:

 viii. Discoverability/findability
 ix.   Machine-readability
 x.   Regulation of financial support of open access to data
 xi.   Definition of responsibility and ownership
 xii. Recognition and reward system for data authors
 xiii. Possibility of licencing
 xiv. Definition of an embargo period

102.   Some of these principles may be merged in order to avoid having a too large number of principles. In addition, a grouping of Principles could be envisaged, with a group of core principles, and another one addressing enablers.

103.   Considering challenges for open access to data, some of the most prominent challenges are:

 i.   Creating incentives for researchers to prioritise provision of data as an integral part of their research. Currently this is addressed more through a 'stick', with funding agencies imposing data publication as an obligation. In the future, a system of recognition (through data citation) and rewards (inclusion of data publication in evaluation and advancement) may foster greater enthusiasm for open data among researchers.

 ii.   An overall investment plan. Big data requires big investments in physical and human capital. Not all data can be published immediately, and thus an investment plan based on prioritisation and estimates of costs and benefits should be established. This also includes a clear 'business plan' for

repositories, work already initiated by the Global Science Forum [DSTI/STP/GSF(2017)1/REV1].

iii.   Creating internationally co-ordinated data networks with user-friendly single-window entry points to enhance findability of data. Such efforts are already ongoing with the establishment of Open Science clouds in Europe, Australia, the U.S. and elsewhere as discussed in Section 2.2. above. The Global Science Forum is also investigating the issue [DSTI/STP/GSF(2017)5].

iv.   Individual privacy issues, as well as business interests, national security considerations and other barriers to data publication are very high on the agenda. Ensuring an adequate balance between those considerations which may restrict publication of datasets and the potential benefits which can be derived from such publication is one of the main political challenges faced by governments.

v.   Measurement of open data is necessary both for impact assessment, quality assurance, and as an enabler for the incentive system. Therefore clear indicators will need to be established and data collection organised regularly. Blockchain technology is mentioned as a potential tool which could help re-allocate private monopoly rents from innovation back into the systemic network of public collaborative science and innovation (Soete, 2016).

vi.   Interoperability is still a very strong challenge. Even though work is done on establishing data standards, these are not always global, and those that are (e.g. the ones developed by the Research Data Alliance) are not widely implemented yet.

vii.   Training in data literacy and data standards is essential, for researchers as well as technical staff operating the repositories. Specific curricula including statistical skills, computer science and information science are needed.

viii.   Co-operation with the private sector is an additional challenge to be addressed, with sensitive issues of data ownership for data created through public private partnerships.

104.   Access to public sector information is very relevant for public research purposes, even though the awareness among scientists may be insufficient in this respect, and many barriers remain to open access to PSI.

105.   The main issues to be addressed for PSI as relates to its reuse in science, technology and innovation are:

- Commitment of governments to release non-sensitive data by default
- For sensitive data, strike a balance between the benefits of sharing against privacy and confidentiality considerations, applying adequate levels of anonymization and aggregation as appropriate.
- Organise collaboration across governmental agencies, but also with the private sector and academia, in order to exchange good practices and identify opportunities for synergies and data sharing.
- Build institutional capabilities at public agencies, to manage, create, curate and reuse data.

References

Ahmad, N. a. (2016). *Measuring GDP in a Digitalised Economy.* OECD Publishing.

Amsterdam Call for Action on Open Science. (2016). Retrieved 09 12, 2017, from file:///C:/Users/paic_a/Downloads/amsterdam-call-for-action-on-open-science.pdf

ARC. (n.d.). *African Research Cloud – African Research Cloud.* Retrieved 09 20, 2017, from https://www.arc.ac.za/

Australian Government. (2016). *2016 NATIONAL RESEARCH INFRASTRUCTURE ROADMAP.* Retrieved 09 20, 2017, from https://docs.education.gov.au/system/files/doc/other/ed16-0269_national_research_infrastructure_roadmap_report_internals_acc.pdf

Australian Government Productivity Commission. (2017). Data Availability and Use - Productivity Commission Inquiry Report. Retrieved 09 07, 2017, from https://www.pc.gov.au/inquiries/completed/data-access/report

EC. (2012). *COMMISSION RECOMMENDATION of 17.7.2012 on access to and preservation of scientific information.* Retrieved 09 12, 2017, from https://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf

EC. (2014). Validation of the results of the public consultation on Science 2.0: Science in Transition. Retrieved 09 20, 2017, from http://ec.europa.eu/research/consultations/science-2.0/science_2_0_final_report.pdf

EC. (2016). H2020 Programme Guidelines on FAIR Data Management in Horizon 2020 History of changes. Retrieved 09 11, 2017, from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

EC. (2016). *Realising the European Open Science Cloud.*

EC. (2016). *REGULATION (EU) 2016/ 679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL - of 27 April 2016 - on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/ 46/ EC (General Data Protection Regulation).* Retrieved 09 13, 2017, from http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

EU implementation of the G8 Open Data Charter. (n.d.). Retrieved 09 29, 2017, from file:///C:/Users/paic_a/Downloads/EUimplementationofG8OpenDataCharter.pdf

European Commission. (2016). *Open innovation, open science, open to the world - a vision for Europe | Digital Single Market.* Retrieved 09 29, 2017, from https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe

G8. (2013). *G8 Open Data Charter and Technical Annex - GOV.UK.* Retrieved 09 29, 2017, from https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex

International Open data charter. (2017). *Open Up Guide: Using Open Data to Combat Corruption.* Retrieved 09 05, 2017, from https://opendatacharter.net/wp-content/themes/proparty/assets/anticorruption_guide.pdf

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management, 29*(4), 258-268. doi:10.1080/10580530.2012.716740

Manyika, J., Lund, S., Bughin, J., Woetzel, J., Stamenov, K., & Dhringra, D. (2016). Digital Globalization : The New Era of Global Flows. *McKinsey Global Institute*.

McKinsey Global Institute. (2013). *Open data: Unlocking innovation and performance with liquid information | McKinsey & Company.* Retrieved 09 07, 2017, from http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information

Netherlands' EU Presidency. (2016). *Amsterdam Call for Action on Open Science.*

NIH. (2017). *Big Data to Knowledge | NIH Common Fund.* Retrieved 09 20, 2017, from https://commonfund.nih.gov/bd2k

NIH. (n.d.). *ERA Commons - NIH.* Retrieved 09 20, 2017, from https://public.era.nih.gov/commons/public/login.do?TYPE=33554433&REALMOID=06-1edb031f-46c7-44b3-b803-60b537de74d2&GUID=&SMAUTHREASON=0&METHOD=GET&SMAGENTNAME=-SM-938PYmoLVb4VrDeXo04LZUDVDvc%2b3899ByInEAjuSUvWNIGfB2zRpWiCivYGCogG&TARGET=-SM-http%3a%2f%2fpu

OECD. (2007). OECD Principles and Guidelines for Access to Research Data from Public Funding. Retrieved 09 18, 2017, from http://www.oecd-ilibrary.org/docserver/download/9207043e.pdf?expires=1505727205&id=id&accname=guest&checksum=98FA3DE023D88AB61F2281628EAC7DC2

OECD. (2013). New Data for Understanding the Human Condition: International Perspectives OECD Global Science Forum Report on Data and Research Infrastructure for the Social Sciences. Retrieved 09 12, 2017, from http://www.oecd.org/sti/sci-tech/new-data-for-understanding-the-human-condition.pdf

OECD. (2015). Making Open Science a Reality. In *OECD Science, Technology and Industry Policy Papers* (Vol. 2015). OECD Publishing, Paris. doi:http://dx.doi.org/10.1787/5jrs2f963zs1-en

OECD. (2015a). Making Open Science a Reality. In *OECD Science, Technology and Industry Policy Papers* (Vol. 2015). OECD Publishing, Paris. doi:http://dx.doi.org/10.1787/5jrs2f963zs1-en

OECD. (2015b). *Data-Driven Innovation: Big Data for Growth and Well-Being.* OECD Publishing, Paris. doi:http://dx.doi.org/10.1787/9789264229358-en

OECD. (2016). Research Ethics and New Forms of Data for Social and Economic Research. In *OECD Science, Technology and Industry Policy Papers* (Vol. 2016). OECD Publishing, Paris. doi:http://dx.doi.org/10.1787/5jln7vnpxs32-en

OECD. (2017). *Compendium of good practices on the use of open data for Anti-corruption: Towards data-driven public sector integrity and civic auditing.* Retrieved 09 07, 2017, from http://www.bmjv.de/SharedDocs/Downloads/EN/G20/Compendium_Open_Data.pdf?__blob=publicationFile&v=2

Open Knowledge International. (n.d.). *The Open Definition - Open Definition - Defining Open in Open Data, Open Content and Open Knowledge.* Retrieved 09 07, 2017, from http://opendefinition.org/

Pentland, A. (2014). Social physics: How good ideas spread-the lessons from a new science. Retrieved 09 04, 2017

Research Data Alliance. (2017). *All Recommendations & Outputs | RDA*, Mendeley. Retrieved 09 25, 2017, from https://www.rd-alliance.org/recommendations-and-outputs/all-recommendations-and-outputs

Research Data Alliance Council. (2014). RDA Governance Document. doi:http://dx.doi.org/10.15497/RDA00001

Rothstein, H., Sutton, A., & Borenstein, M. (2006). Publication Bias in Meta-Analysis. In *Publication Bias in Meta-Analysis* (pp. 1-7). John Wiley & Sons, Ltd. doi:10.1002/0470870168.ch1

Schaffers, H., Komninos, N., Pallot, M., Trousse, B., Nilsson, M., & Oliveira, A. (2011). LNCS 6656 - Smart Cities and the Future Internet: Towards Cooperation Frameworks for Open Innovation. *Future Internet Assembly LNCS, 6656*, 431-446. Retrieved 09 20, 2017, from https://s3.amazonaws.com/academia.edu.documents/43890480/Smart_cities_and_the_future_internet.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1505922071&Signature=39FiGFNohjmoxPgcaJzZoRKV7Xw%3D&response-content-disposition=inline%3B%20filename%3DSmart_Cities_and_the_Future_Internet_Tow.pdf

Soete, L. (2016). *A sky without horizons - presentation at Blue Sky Conference 2016*. Retrieved from LinkedIn Slideshare: https://www.slideshare.net/innovationoecd/soete-a-sky-without-horizons

The World Wide Web Foundation. (2017). *Open Data Barometer 4th Edition - Global Report.* Retrieved 09 06, 2017, from http://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-GlobalReport.pdf

UNESCO. (2011). Revised Draft Strategy on UNESCO's Contribution to the Promotion of Open Access to Scientific Information and Research. Retrieved 09 20, 2017, from http://www.unesco.org/fileadmin/MULTIMEDIA/HQ/CI/CI/images/GOAP/OAF2011/213342e.pdf

Vickery, G. (2011). *Review of Recent Studies on Psi Re-Use and Related Market Developments.*

Wilkinson, M., Dumontier, M., Aalversberg, I., Appleton, G., & Axton, M. (2016). Comment : The FAIR Guiding Principles for scienti fi c data management and stewardship. *Nature Communications*. doi:10.1038/sdata.2016.18

ZBW – Leibniz-Informationszentrum Wirtschaft. (2016). *The G7 Open Science Working Group Action Points – Speeding up Open Science? | ZBW MediaTalk.* Retrieved 09 20, 2017, from https://www.zbw-mediatalk.eu/en/2016/11/the-g7-open-science-working-group-action-points-speeding-up-open-science/

ZBW. (2016). *GO-FAIR: A Member States-Up strategy for the EOSC implementation | ZBW MediaTalk*, Mendeley. Retrieved 09 12, 2017, from https://www.zbw-mediatalk.eu/2017/01/go-fair-a-member-states-up-strategy-for-the-eosc-implementation/

ZBW-Mediatalk. (2017). *GO-FAIR: A Member States-Up strategy for the EOSC implementation | ZBW MediaTalk.* Retrieved 09 20, 2017, from https://www.zbw-mediatalk.eu/2017/01/go-fair-a-member-states-up-strategy-for-the-eosc-implementation/

Annex 1: Questionnaire

Current use of the OECD Recommendation concerning Access to Research Data from Public Funding and perspective for a future overarching Recommendation on Data Access

Question 1: Please briefly describe your experience in relation to the issue of data access, as regulator, supplier of data or user of data.

Response box (feel free to expand box as needed)

Question 2: Please describe the main recent ***policy initiatives (reforms or changes to existing legislation)*** your country has implemented in the domain of ***access to research data from public funding*** and ***public sector information as relevant to science, technology and innovation***.

Question 2 response table

Policy initiative number	Year	Title	Brief description	Source (URL if available)
1	2005 and earlier			
2	2005 and earlier			
3	2005 and earlier			
4	2005 and earlier			
5	2005 and earlier			
6	2005 and earlier			
7	2005 and earlier			
8	2005 and earlier			
9	2005 and earlier			
10	2005 and earlier			

Question 3: For the policy initiatives above, please assess the influence of OECD Recommendation C(2006)184 by using a Likert scale 1= None; 2= Slight; 3= Moderate, 4=Important, and 5= Very important

Question 3 response table (please tick the appropriate box to inform degree of influence and relevance, then add free text to last column.)

Policy Initiative number	Degree of influence					Comment on influence of Recommendation C(2006)184 (or lack thereof)
	None	Slight	Mode-rate	Impor-tant	Highly impor-tant	
1	☐	☐	☐	☐	☐	
2	☐	☐	☐	☐	☐	
3	☐	☐	☐	☐	☐	
4	☐	☐	☐	☐	☐	
5	☐	☐	☐	☐	☐	
6	☐	☐	☐	☐	☐	
7	☐	☐	☐	☐	☐	
8	☐	☐	☐	☐	☐	
9	☐	☐	☐	☐	☐	
10	☐	☐	☐	☐	☐	

Question 4: The stated aims and objectives for the advancement of the data access agenda include: (i) promoting a culture of openness and sharing of research data among the public research communities within Member countries and beyond; (ii) stimulating the exchange of good practices in data access and sharing; (iii) raise awareness about the potential costs and benefits of restrictions and limitations on access to and the sharing of research data from public funding; (iv) highlighting the need to consider data access and sharing regulations and practices in the formation of Member countries' science policies and programmes; (v) providing a commonly agreed upon framework of operational principles for the establishment of research data access arrangements in Member countries; (vi) offering recommendations to Member countries on how to improve the international research data sharing and distribution environment.

Please give us your views as to the relevance of each of these objectives, as well as the current policy effort related to those objectives.

Question 4 response table (please tick the appropriate box to inform relevance and current policy effort of objectives, then add a comment to the last column.)

Objective	Relevance					Current policy effort					Comment
	None	Slight	Mode-rate	High	Very High	None	Slight	Mode rate	High	Very High	
Culture of openness and sharing of research data	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
Exchange of good practices in data access and sharing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
Awareness about the potential costs and benefits of limitations on access to data and sharing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
Consider data access and sharing regulations and practices in formation of science policies and	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
Framework of operational principles for the establishment of research data access arrangements	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
Improving the international research data sharing and distribution environment	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

Question 5: Addressing challenges in access to research data

Progress has been achieved in the decade since the Recommendation C(2006)184 was issued. Data associated with publications is now expected to be made available, and the 'reproducibility' agenda is getting more attention, e.g. in clinical trials. A number of research funders are now mandating open data.

Nevertheless, data sharing still seems to be less widespread than would be expected, and is still limited to a small number of research fields. Reasons for this seem to be disincentives to sharing research data, such as lack of reward or credit for sharing data, the substantial effort needed to upload and maintain data in a form which is usable by others, risks of misinterpretation or misuse, control of intellectual property and personal data protection issues, i.e. the need of anonymising data samples. In addition, demand for shared data seems limited to a few scientific disciplines (Borgman, 2012).

The OECD Global Science Forum has identified**Invalid source specified.** nine challenges related to data-driven and evidence-based research in social and economic sciences. We are synthesising[7] those findings below, and ask you to react regarding the relevance and policy effort invested in overcoming those challenges:

- In the table below, please inform the relevance of the challenge to your country, as well as the perceived effort of policy makers to address the challenge.

- Please also add any suggestions you might have on how to address the identified challenges going forward.

---

[7] For the synthesis, please refer to Box 3.2 in the main document

Question 5 response table (please tick the appropriate box to inform relevance and policy effort, then add free text to last column.)

Category	Challenge	Relevance of challenge to your country					Policy effort in addressing the challenge					How to address the challenge in a potential overarching recommendation on enhanced data
		None	Low	Med	High	Very high	None	Low	Med	High	Very high	
A. Infrastructure and skills	A.1 Lack of data management planning to make datasets available for reuse	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
	A.2 Investments in personnel and infrastructure needed for data creation and curation	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
B. Legal and regulatory barriers	B.1 Lack of information on what data exists and lack of adoption of international standards for data documentation	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
	B.2 Individual privacy issues	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
	B.3 Legal, cultural, language and proprietary rights barriers	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
C. Researchers' incentives	C.1 Incentives for researchers to ensure effective data sharing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
D. Data quality and characteristics	D.1 Reliability, statistical validity and generalisability of different data sources	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
	D.2 Need for greater harmonisation of data sources across countries	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
	D.3 Increasing need for international coordination	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
	D.4 Increasing need for interdisciplinary coordination for global challenges	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

Question 6: Are there any other _challenges_ or _opportunities_ you would like to bring up? Some examples may include:

- Support to use of open data through: (i) Raising awareness about the existence of datasets; (ii) Standardization of meta-data to facilitate search and interpretation; (iii) educating users to provide statistical skills needed to use the data.

- Cost-benefit analysis. For extensive datasets costs of data curation, publishing and curation may exceed the benefits and not represent good use of taxpayers' funds.

- Public-private partnerships in data curation and provision.

Question 6 response box (feel free to expand box as needed)

Question 7: The OECD Recommendation C(2006)184 is articulated around specific Principles. Please give us your view on the following:

- The relevance of the Principle to your country

- The need for the principle to be included in the potential new overarching recommendation, as opposed to just keeping it within the current instrument

- Comment on issues related to implementation of the principle, and potential new focus dues to the evolution of the needs of stakeholders and policy makers.

Question 7 response table (please tick the appropriate box to inform degree of relevance of Principle, the need to include it in the overarching Recommendation and add free text to last column.)

Principle	Relevance of Principle to your country					Need to include Principle in overarching Recommendation					Comment on implementation of Principle and potential new focus *(feel free to expand box as needed)*
	None	Low	Med	High	Very high	None	Low	Med	High	Very high	
A. Openness	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
B. Flexibility	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
C. Transparency	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
D. Legal conformity	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
E. Protection of intellectual property	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
F. Formal responsibility	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
G. Professionalism	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
H. Interoperability;	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
I. Quality	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
J. Security	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
K. Efficiency	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
L. Accountability	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
M. Sustainability	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

Question 8: Are there any other Principles you would suggest for an overarching Recommendation, and why? Please be specific.

Question 8 response box (feel free to expand box as needed)

Question 9: Taking into account the evolution of usage of *research data from public funding* over the past decade, which are some additional issues which would need to be addressed in the potential new overarching recommendation on enhanced access to data?

Question 9 response box (feel free to expand box as needed)

Question 10: Beyond data from public research, how relevant is access to ***public sector information*** to scientists in your country? What are the barriers to accessing this information?

Question 10 response table (please tick the appropriate box to inform degree of relevance of Principle, the need to include it in the overarching Recommendation and add free text to last column.)

Relevance of Public Sector Information to science					Percieved barriers to accessing Public Sector Information					Please list the main difficulties encountered in accessing Public Sector Information *(feel free to expand box as needed)*
None	Low	Med	High	Very high	None	Low	Med	High	Very high	
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

Question 11: Which are the issues which would need to be addressed for ***public sector information*** in the potential new overarching recommendation on enhanced access to data?

Question 11 response box (feel free to expand box as needed)

References

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, *63*(6), 1059-1078. http://onlinelibrary.wiley.com/doi/10.1002/asi.22634/epdf

OECD (2013), "New data for understanding the human condition: International Perspectives", OECD Global Science Forum Report on Data and Research Infrastructure for the Social Sciences, OECD Publishing, Paris, www.oecd.org/sti/sci-tech/new-data-for-understanding-the-human-condition.pdf.

OECD (2015), "Making Open Science a Reality", OECD Science, Technology and Industry Policy Papers, No. 25, OECD Publishing, Paris. http://dx.doi.org/10.1787/5jrs2f963zs1-en