

## Chapter 6. Enhanced access to publicly funded data for STI

By

Alan Paic and Carthage Smith

*Enhanced access to data can be a key enabler for science, technology and innovation (STI). It can support new scientific insights across disciplines, contribute to reproducibility of scientific results, and facilitate innovation. However, many countries have yet to develop comprehensive approaches to enhance access to data. This chapter focuses on policy concerns and potential policy action to enhance access to publicly funded research data for STI. It starts with an overview of public research data. It then outlines the specific policy dilemmas concerning enhanced data sharing. These include: (i) fostering data governance for trust and balancing the benefits and risks of data sharing; (ii) developing and implementing technical standards and practices; (iii) defining responsibility and ownership of data; (iv) changing recognition and reward systems to encourage scientists to share data; (v) implementing business models and long term funding for data provision; and (vi) developing human capital and skills to support data sharing and analysis. Finally, the chapter draws policy implications for the future by outlining two possible scenarios.*

## Introduction

Research is becoming increasingly data-intensive. “Big data” are no longer the prerogative of experimental physics and astronomy: they are spreading across all scientific domains. Access to data is a key enabler for science, technology and innovation (STI); not surprisingly, enhancing this access is a major priority for policy makers (OECD, 2006) (European Commission, 2014). As a critical element of open science, big data are expected to lead to new scientific breakthroughs, less duplication and better reproducibility of results, as well as bring about improved trust and innovation (OECD, 2015a, 2015b). The development of artificial intelligence (AI) further reinforces the importance of access to data, since AI algorithms need very large amounts of well-described data to “train”, i.e. improve their performance.

Open data can be simply defined as “data that can be accessed and reused by anyone without technical or legal restrictions” (OECD, 2015a). This does not necessarily mean data is free of cost, although in the context of open science, it is normally assumed the user bears no charges. Openness is not a binary concept: data can be made more or less open, according to the specific nature of the data and the community of stakeholders involved. “As open as possible, as closed as necessary” is gradually replacing the “open-by-default” mantra associated with the early days of the open-access movement. Opening up data can help advance the STI agenda, but this needs to be balanced against issues of costs, privacy, security and preventing malevolent uses. Enhanced access to data is a term that is used increasingly in relation to public sector data and captures some of these important caveats around openness.

Enhanced access to data can be described as encompassing any practical and lawful means through which data can be effectively accessed by, and shared with an entity (individual or organisation) other than the data holder, for the purpose of fostering data re-use by the entity or a third-party chosen by the entity, while, at the same time, taking into account the private interests of individuals and organisations concerned (e.g. their intellectual property and privacy rights) as well as national security and public interests.

This chapter focuses on enhanced access to publicly funded research data for STI. It starts with an overview of public research data. It then develops the specific policy dilemmas concerning enhanced data sharing. Finally, it draws policy implications for the future. Much has already been written on this topic, and not all the important issues can be fully addressed in one short chapter. Hence, the chapter focuses on policy concerns and potential policy action. It builds on the recent OECD data-access survey of OECD countries regarding the OECD 2006 Recommendation on Access to Research Data from Public Funding (OECD, 2017a). It also draws on responses to the 2017 European Commission-OECD STI Policy Survey (STIP Compass) and the discussions held at an OECD expert workshop on principles for enhanced access to public data held in March 2018.

### Public data for STI: An overview

Three broad categories of data are used for STI: 1) public-sector information (PSI), produced, curated and managed *by* or *for* governmental entities; 2) data resulting from publicly funded research; 3) privately owned or commercial data. This chapter covers only publicly funded data for STI, which includes both data produced by research and PSI used in research, such as meteorological or social survey data. These distinctions are somewhat artificial and partially overlapping, but they can be important in defining where policymaking responsibilities lie. For example, unlike data generated by research, access

to PSI is not principally the remit of STI policymakers, and yet research data is sometimes treated as a subset of PSI, as is the case in the latest EC Directive on the re-use of PSI (European Commission, 2018a). Ensuring that PSI-related policies and practices that affect research are consistent with policies and practices affecting other research data requires co-ordination across policy communities.

Publicly funded research data are defined in the OECD 2006 Recommendation on Access to Research Data from Public Funding (OECD, 2006) as data “that are supported by public funds for the purposes of developing publicly accessible scientific research and knowledge”. Research data can be further defined as: “factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings”. We exclude from the scope of consideration here research data gathered for the purpose of commercialisation of research outcomes, or research data that are the property of a commercial sector entity. Access to such data is subject to a range of considerations that are beyond the scope of this document.

This chapter focuses on the *data* outputs of research, rather the *publication* outputs. It makes this distinction mainly for pragmatic reasons: because some policy issues – particularly the role of commercial publishers – are distinct, issues around open-access publications are normally considered separately from those concerning research data (e.g. OECD, 2015a). Nevertheless, research data and publications are widely recognised as part of a continuum and policies need to be connected accordingly. Access to data currently lags behind access to publications: more than 92% of universities in Europe have – or plan to have – open-access policies for publications, but under 28% have established guidelines concerning open access to data. The main institutional barriers to promoting research data management and/or open access to research data are: different “scientific cultures” within the university, absence of national guidelines or policies, limited awareness of benefits, legal concerns, and technical complexity (Morais and Borrell-Damian, 2018).

### ***Rationales for sharing research data***

At least six main rationales exist in favour of enhanced access to public research data (Borgman, 2012):

1. *New scientific insights*: Providing broader access to data allows more researchers (and citizens) to analyse and link those data to other data sources, to respond to different scientific questions. For example, biodiversity data are increasingly used by the health-research community working on emerging diseases.
2. *Reproducibility of scientific results*: Sharing access to the data underpinning scientific publications allows peers to test and reproduce scientific results. In practice, data alone are often insufficient for testing reproducibility, and enhanced access to analysis software is also necessary.
3. *Public research is a public good*: Data from publicly funded research should, in principle, be available to researchers, citizens and commercial actors who wish to use and derive value from them. This is sometimes also an issue of transparency and accountability.
4. *Promote innovation*: Allowing commercial companies to access public research data enables them to use the data to accelerate innovation on products (e.g. new drugs) or new data services (e.g. in weather forecasting). Data are an essential enabler for AI and related innovations.

5. *Support meta-analyses*: Enhancing access to and sharing of data encourages meta-analysis, which combines the results of different related studies (e.g. clinical trials of a drug) to provide greater statistical power.
6. *Avoid duplication*: Sharing datasets showing positive or negative results can avoid duplication of research efforts (Rothstein et al., 2006).

When taken together, these rationales contribute to a more efficient and effective scientific enterprise. Access to data alone is insufficient to achieve all these expectations, but lack of access is a major barrier to achieving them.

There are also legitimate concerns about enhanced access to data, including privacy and intellectual property protection and national security and other public interests. These risks are discussed in the section on Future outlook. When, how and under what conditions public research data should be made accessible are important policy questions, which cut across the issues discussed in the rest of this chapter.

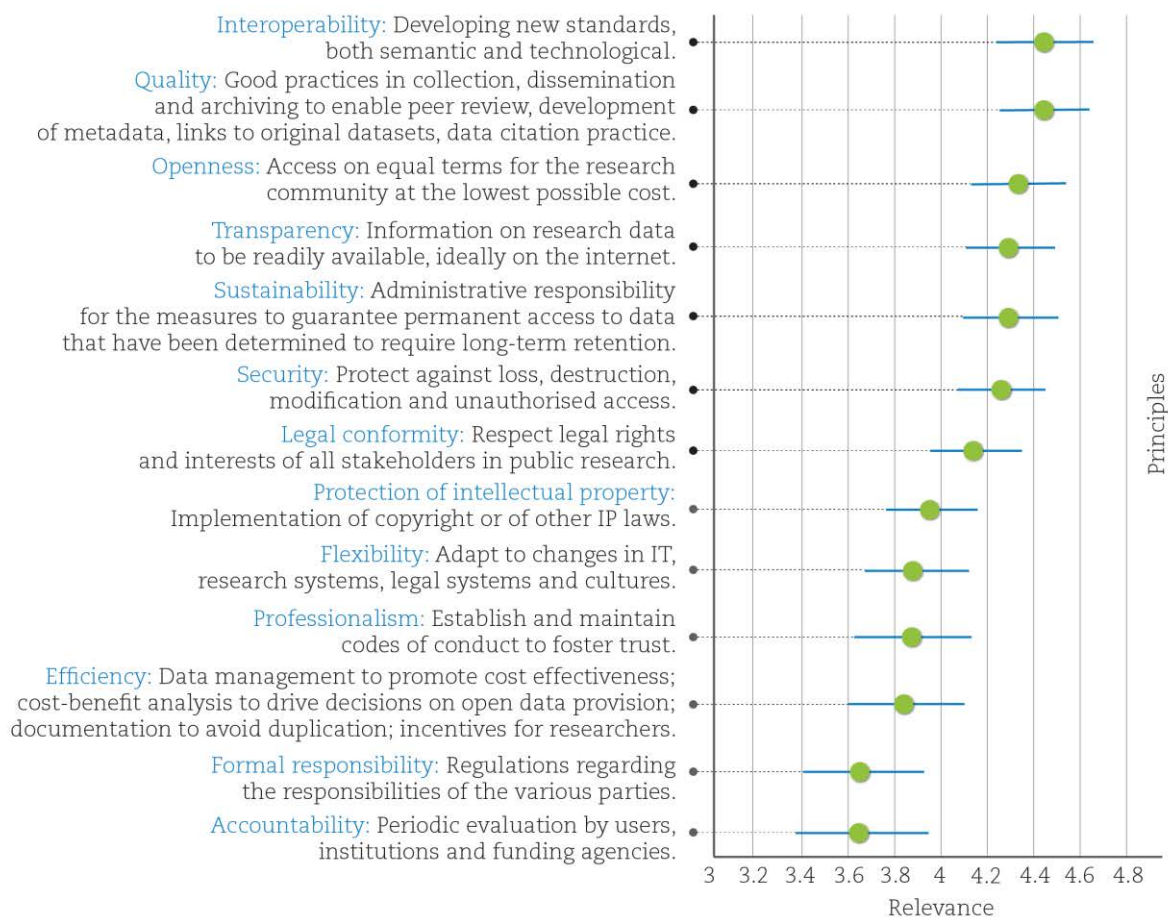
### *Policy action in favour of sharing research data*

The OECD 2006 Recommendation (OECD, 2006) and the OECD Principles and Guidelines for Access to Research Data from Public Funding (OECD, 2007) represented an important step in multilateral efforts to foster open access to data in STI. A wide range of policies were implemented fairly quickly in response to these instruments: some countries introduced laws and comprehensive policies; others issued position statements and future plans (OECD, 2009).

A 2017 OECD survey on access-to-data policies among policy makers from 27 countries identified a total of 171 policy initiatives targeting enhanced access to data. Survey respondents were also asked to assess the relevance of the 13 principles cited in the 2006 OECD Recommendation (OECD, 2006). The principles considered the most pertinent today were openness, quality, security, interoperability, transparency, sustainability and legal conformity (Figure 6.1).

Building on earlier work by OECD (OECD, 2007), the findability, accessibility, interoperability and reusability (FAIR) data principles have been developed by a diverse set of stakeholders representing academia, scholarly publishers, industry and funding agencies, and are now becoming a mainstream reference for policy makers (Wilkinson et al., 2016) (Table 6.1).

**Figure 6.1. An assessment of the relevance of the OECD principles concerning access to research data from public funding**



Additional principles quoted by the stakeholders but not included in the 2006 recommendations:

**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.

**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, foresee legal sanctions

**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 systemic network of public collaborative science and innovation (Soete, 2016).

*Note:* The 2017 survey asked respondents to assess the relevance of the 13 principles cited in the original OECD Recommendation (OECD, 2006) on a Likert scale (5 = very high relevance; 0 = no relevance). Responses were received from 55 organisations in 27 countries.

**Table 6.1. Overview of FAIR principles**

FAIR principles	Action items	Technical requirements
Findable – data should be easily found by humans and machines alike	Establish portals and open-science clouds	<ul style="list-style-type: none"> <li>• Globally unique and persistent identifiers</li> <li>• Data indexed in a searchable database</li> </ul>
Accessible – as open as possible, as closed as necessary	Use open licensing, whenever possible	<ul style="list-style-type: none"> <li>• Machine readability</li> </ul>
	Establish trusted-user access for more sensitive datasets	<ul style="list-style-type: none"> <li>• Standardised communication protocol</li> <li>• Metadata are accessible – even after the data are no longer available.</li> </ul>
Interoperable – datasets need to be combinable with other datasets	Three aspects of interoperability: semantic (taxonomy), legal (rights) and technical (machine readability)	<ul style="list-style-type: none"> <li>• Semantic interoperability – common vocabulary</li> </ul>
	Standard-setting	<ul style="list-style-type: none"> <li>• Data include relevant references to other datasets</li> </ul>
Reusable – it must be possible to re-use data in future research projects and then process these data further	Data curation	<ul style="list-style-type: none"> <li>• Metadata are exhaustive</li> </ul>
	Open Archival Information System (OAIS)-compliant repositories	<ul style="list-style-type: none"> <li>• Data describe multiple precise and appropriate properties</li> </ul>
		<ul style="list-style-type: none"> <li>• Data are released with a clear and accessible data licence</li> </ul>
		<ul style="list-style-type: none"> <li>• Data are connected to their origin</li> <li>• Data meet standards relevant to the field</li> </ul>

Source: Author's analysis, based on Oxford Research (2018) and expert opinions (OECD, 2018).

Several other multilateral initiatives have been developed to foster data sharing, particularly at the European level (Box 6.1). At the national level, the 2017 edition of the EC-OECD STI Policy survey asked OECD member and partner countries to provide information about policy initiatives supporting open science and open access. Most of the 181 policy initiatives cited concern infrastructures and strategies, and a smaller number concern governance issues (Box 6.2).

### Box 6.1. International policy initiatives to promote sharing of research data

The Transparency and Openness Promotion (TOP) guidelines were created by journals, funders and societies to align scientific ideals with practices. They include standards covering citations, data transparency, software, research materials, design and analysis, as well as preregistration of study and analysis plans, and replication. Journals select which of the eight transparency standards they wish to adopt, as well as a level of implementation for each standard (Center for Open Science, 2014).

In 2012, the European Commission issued a Recommendation on access to and preservation of scientific information, calling for co-ordinated open access to scientific publications and data, preservation and re-use of scientific information, development of e-infrastructures among EU Member States (European Commission, 2012). The Recommendation was updated in 2018 (European Commission, 2018).

In 2016, the European Commission published “Open Innovation, Open Science, Open to the World”, a vision that incorporated its ambitious plans for a European Open Science Cloud (EOSC)<sup>1</sup> (European Commission, 2016). The EOSC is conceived to provide

EU researchers an environment with free and open services for data storage, management, analysis and re-use across disciplines by connecting existing and emerging infrastructures, adding value and leveraging past infrastructure investment. The EOSC is expected to develop common specifications and tools to ensure data is FAIR and legally compliant with the European Union's General Data Protection Regulation (GDPR) and cybersecurity legislation. It also foresees mechanisms for cost recovery on cross-border access (European Commission, 2018b).

Similar “cloud” initiatives include the National Institutes of Health (NIH) Commons in the United States (NIH, 2017), the Australian Research Data Cloud (eRSA, 2014), and the African Research Cloud (ARC). All of these initiatives aim to be interconnected and interoperable.

In addition to government policy, research funders and scientific journals are increasingly demanding open-access to data. Funders require data-management plans and have specific data-release policies; some (such as the UK Economic and Social Research Council) even require researchers wishing to collect new data to demonstrate that no existing data can be used for their purpose (Economic and Social Research Council, 2015). Many scientific journals require data statements and links; some (such as *Science*) require authors to share the computer code they used to create or analyse data.

In summary, since the OECD Recommendation drew international attention to the area in 2006, several multilateral initiatives to promote access to research data have been launched. The FAIR principles have de facto become an international norm, helping to guide policy actions. The majority of OECD countries are taking actions to promote open data, sometimes in association with plans to develop science clouds linking research data with services provided to the entire research community. However, several outstanding challenges need to be overcome before open data becomes a reality (the section on Future outlook addresses these challenges).

**Box 6.2. Instruments concerning data access reported in the 2017 EC/OECD STI Policy Survey**

Of the 181 policy initiatives reported as supporting open science and open access, 74 (42%) are concerned with research infrastructures, including portals offering open access to publications; repositories and archives for scientific data; search engines; virtual networks; and clouds connecting individual physical repositories. Examples include the European Open Science Cloud, and Research Data Infrastructure for Open Science in Japan. In some cases (Australia, Estonia, Finland and France), open-data infrastructure is treated within a national strategy on research infrastructures.

55 initiatives (33%) are national strategies and policies for open access to data and publications. These include:

- dedicated strategies and policies for open access to data and publications at the policy-making level (Czech Republic, Korea, New Zealand, Norway, Slovenia and United Kingdom), as well as at the funding-agency level (Australia, Austria, Belgium-Federal, Canada, Lithuania, Nordic Council of Ministers, Netherlands, Norway, Portugal, Switzerland and United Kingdom)
- open-data access within open-science policies (e.g. Chile, Colombia, Denmark, Estonia and the Netherlands); the Open Innovation Strategy (Austria); the Innovation and Science agenda (Australia); the Law on Scientific Activity (Latvia); and a specific Law 310/2014 for Public Research which focuses on co-operation between business and academia (Greece)
- open-data access, integrated within open-government and public sector-information initiatives (Australia, Argentina, Brazil, Canada, Sweden)
- bottom-up approaches through institutions (Centre national de la recherche scientifique and Institut national de la recherche agronomique in France; University of Malta; universities in Slovenia; and Concordat on Open Research Data in the United Kingdom).

13 initiatives (7%) aim to create or reform a governance body to foster open access. These include:

- Etalab, a high-level, pan-governmental open-data platform in France co-ordinating open-data and open-government initiatives, which is chaired by the national chief data officer and reports to the French Prime Minister
- a national focal point (chief science officer Canada, national chief data officer in France, point of reference in Slovenia) for access to and preservation of scientific data
- an agency for information systems used in higher education and research (CERES – National Center for Systems and Services for Research and Studies, Norway)
- The Datacite consortium, which enables researchers to attach a digital object identifier (DOI) to research data (Estonia)
- the Data Archiving and Network Services institute, which facilitates data archiving and re-use, and provides training and consultancy (Netherlands)
- open-data institutes (Canada and the United Kingdom) supporting economic, environmental and social-value creation opportunities arising from open data



8 initiatives (4%) are concerned with networking and collaborative platforms to facilitate open access to data. These include:

- OpenAire Advance, a network of repositories with 34 National open science desks promoting open science as the default solution in Europe
- library networks (HEAL Link in Greece, HAL and Persée in France)
- think tanks sharing good practice and engaging in advocacy (EPRIST in France)
- a data-analytics initiative linking disparate government datasets (Data61 in Australia)
- cooperatives of research, educational and medical institutions (e.g. the SURF cooperative in the Netherlands), aiming to promote innovation in information technology
- a commercialisation marketplace (Open Data Exchange in Canada)

5 initiatives (3%) undertake formal consultations of stakeholder groups, including expert groups. These include:

- working groups and committees for open science and open access to scientific data (e.g. the European Commission Directorate General for Research, Technology and Innovation, and initiatives in France, Greece, Ireland, Japan, Slovenia, Turkey and the United Kingdom)
- an open-data forum advocating the development of open-data policies (United Kingdom)

Several initiatives aim to collect data about researchers, research projects and policies. For an overview of these initiatives, see Chapter 12 on digital science and innovation policy.

*Source: EC/OECD (2017)*

### Policy challenges to promoting enhanced access to data

The 2017 OECD data-access survey and a follow-up workshop in 2018 identified six key areas of policy concern with regard to enhancing access to public data for STI, as follows:

- *Data governance for trust – balancing the benefits of data sharing with the risks:* Opening up data can help advance the STI agenda, but this needs to be balanced against issues of costs, privacy, intellectual property, national security and other public interests.
- *Technical standards and practices:* Achieving FAIR goals hinges on the development and adoption of a common technical framework. The challenge is that technology development is now far outpacing standard-setting, creating regulatory gaps.
- *Defining responsibility and ownership:* Intellectual property rights and licensing arrangements associated with data need to be clearly defined and respected. IPR protection can be an important incentive for private sector investment in research and innovation. At the same time, enhanced access to data is also a driver for innovation. Public-private partnerships present a particular challenge, with the risk of “privatising” and preventing access to data derived from publicly funded research.

- *Incentives*: Recognition and rewards encourage researchers to share data. Current academic-reward systems mostly motivate researchers to publish their scientific results and do not attach enough value to the sharing of data.
- *Business models and funding*: The costs of providing open data are mostly borne by data providers, while the benefits accrue to users including those who develop “value added” data services. There are a variety of business models for providing data access and services, but these are often restrained by policy mandates and incentives.
- *Building human capital and institutional capabilities to manage, create, curate and re-use data*: A lack of skills breeds a lack of trust. It is important to ensure there are appropriate skills along the full data value chain, including data management skills of researchers, curation skills with data stewards, and data literacy among users.

The following subsections develop these six challenges.

### *Data governance for trust – balancing benefits and risks*

Balancing the potential public benefits and risks of sharing research data is a critical issue for data governance. Sound data governance is needed to ensure trust from both data providers and users, and promote a culture of sharing, with the aim of making data “as open as possible and as closed as necessary”.

Sharing data presents multiple risks related to: 1) individual privacy (e.g. in the case of clinical research data); 2) misuse (e.g. data about rare and endangered species, or rare minerals); 3) misinterpretation (particularly as concerns datasets of uncertain quality, and/or lacking the appropriate metadata); and 4) national security (e.g. data from research with potential military applications). More granular data often have higher potential research value, but the risk increases as well.

Providing access to personal data or human subject data is a particular challenge (OECD, 2013). Although anonymisation techniques can remove personally identifiable information from individual datasets, true anonymisation becomes very difficult as more and more data from different sources are integrated (President’s Advisors on Science and Technology, 2014). Moreover, the research value of personal data often stems from the ability to link it back to individual characteristics. In the United Kingdom, for example, linking information from hospitals with the cancer-data repository, and data from various screening programmes, has made it possible to recommend changes in medical protocols that are likely to improve cancer survival rates. Rules and laws can be a disincentive to breaching anonymity, but the financial incentives to do so can be high in certain industries, and legal regimes are very difficult to implement across national jurisdictions.

Alongside anonymization, informed consent is the second pillar underlying the use of personal data in research. Consent is a right recognised in many countries and enshrined in legislation, such as the recent GDPR (European Commission, 2016). However, situations exist in research where consent for using data for specific research purposes is impossible or impractical to obtain, particularly if these purposes were not envisaged when the data were originally collected. For example, when analysing new forms of data from social networks in ways the collector had not anticipated, it might be unfeasible to go back to all the individuals to ask for consent. It is notable that the GDPR<sup>2</sup> makes exceptions for the use of data in research, where consent is one consideration, but is not prescribed as the legal basis for data use. Recent OECD work on the subject stressed the need for properly

constituted independent ethics review bodies (ERBs), outlining their role in evaluating applications to access publicly funded personal data for research purposes. Well-functioning ERBs contribute to building trust (OECD, 2016). This recent work also emphasised the importance of public engagement in defining norms on the use of personal data in research. The approach adopted by the Australian Government, which aims to achieve value creation with open data while transparently managing risk, is one example (Box 6.3).

**Box 6.3. In my view – Trust is the key to unlocking data**

The Hon. Michael Keenan MP, Minister for Human Services and Digital Transformation,  
Australian Government

Data is the fuel powering our new digital economy. However, news of data breaches and misuse of personal information erodes trust and leads the public to believe that data is bad or something to be feared.

If these negative perceptions become entrenched, we risk missing out on the enormous opportunities and benefits data offers to improve people's lives, help grow the economy and become more successful as a nation.

As a Government, we have a responsibility to use data to make the best possible decisions to improve people's lives. In May 2018, the Australian Government announced reforms to simplify the way public data can be shared and used, and clarify accountabilities around the management of data. These reforms are made up of four components:

- A Consumer Data Right to give Australians greater access and control over their data, to enable them to get a better deal from their bank, energy and telecommunications companies;
- A National Data Commissioner to manage the integrity and improve how the Australian Government manages and uses data;
- A new National Data Advisory Council to provide advice on ethical data use, technical best practice, and industry and international developments; and
- Enabling legislation – the *Data Sharing and Release Act* – to improve the use and re-use of data while strengthening security and privacy protections for personal and sensitive data.

These reforms represent a tremendous opportunity to unlock national productivity. However, we will only seize this opportunity if public data is used in a safe and transparent manner and citizens trust their privacy and security is being valued and protected at all times.

To achieve that, we are working hard to secure the trust of the public at the core of our reforms.

This is the only way we can ensure the benefits of data and insights are driving effective outcomes for all people and organisations and indeed, for the entire economy and society.

Data is the fuel of growth and trust is the key that will enable us to get ahead.

If the full benefits of open data are to be realised, trust is required at multiple levels, not just as it relates to personal data. Power relations between individuals, institutions and countries are a critical component of trust, and need to be considered when developing data

access policies. The reality is that open research data can be more readily exploited by more advanced companies, institutions and countries, which master the technology and the algorithms needed to analyse extract value from the data. Less empowered stakeholders can easily be reduced to simple data providers, while the (research and monetary) value is captured elsewhere.

In order to secure public trust and accountability, the socio-economic impacts of open research data need to be monitored. Over time, such impact assessments should help society evaluate the value of open-data initiatives. The 2006 OECD Recommendation suggested considering a few core aspects for external evaluation, including overall public investments, the management performance of data collection, and the extent to which existing datasets are used and reused (OECD, 2006). This provides useful starting guidance. Nevertheless, it must be noted that such assessments are quite challenging to implement, since the methodologies are not yet well developed and standardised.

Data integration is another major opportunity. For example, New Zealand’s Integrated Data Infrastructure<sup>3</sup> allows registered researchers to access microdata about people and households, including data on education; income and work; benefits and social services; population; health; justice and housing. Such an integrated dataset enables social-science research on issues such as the life outcomes of socially disadvantaged groups, linking their educational attainment to income, health and crime outcomes.

Building on current experience and looking forward, some policy implications can be drawn for governments:

- Public data for STI should be “as open as possible, as closed as necessary”. When it comes to accessing sensitive data, governance arrangements are critical. Ethics review boards that include data experts can play an important role in this respect.
- Governments should strive to enhance trust among different stakeholders, and create consensus around data sharing and re-use. Risks of privacy breaches cannot be completely avoided, but should be managed, and the procedures to this end should be clear and transparent.
- Specific initiatives can be launched to support data integration, exploring ways in which data from different sources can be combined transparently across different institutions. These initiatives should explore important issues relating to sensitive data, such as anonymization and informed consent.
- Socio-economic assessments should be undertaken to monitor the impact of open research data, with specific attention to where – and to whom – benefits accrue.

### *Technical standards and practices*

As the volume and variety of research data increases, the resources required by data providers to make their data available, and the time invested by users to discover available data, also increase proportionally (OECD, 2015a). Insufficient information exists on what data are available, both for and from research. When data can be found, they are not always useable, because they do not conform to standards, lack metadata or are not machine-readable.

At the national scale, a large variety of institutional and domain-specific data catalogues, search engines and repositories are being established to enhance the findability of data (Box 6.1 and Box 6.2). At the international scale, increased efforts to co-ordinate and

support global data networks are necessary (OECD, 2017c), to provide the foundation for developing open-science cloud initiatives that will facilitate data usage (Box 6.1).

Scientific publications are another major channel of discoverability. Many researchers first read about potentially interesting data in a journal article; the question then is how to obtain access to that data. Persistent links should appear in published articles, which should also include a permanent identifier for the data, code and digital artefacts underpinning the published results. Data citation should be standard practice. Broken links or inadequate metadata are common challenges, especially as journals tend to be lenient on data requirements for fear of losing good papers to competing journals. Several publishers have recently developed data journals, which can play an important role in promoting the use of published datasets.

Formal standard-setting through bodies, such as the International Standards Organisation, is a slow iterative process of negotiation that can take several years. As a result, pro-active commercial or public players in a position of power can set de facto standards. One example is Google's General Transit Feed Specification, a common format for public transportation schedules and associated geographic information (OECD, 2018).

The research community can turn this into an advantage if it takes the lead in developing appropriate standards and in so doing, consults fully with all concerned stakeholders. This is the approach taken by organisations that are helping to build the social and technical infrastructure to enable open sharing of data across national and disciplinary borders. For example, the Research Data Alliance produces recommendations – which can be adopted as standards – on a broad range of issues related to interoperability, data citation, data catalogues or workflows for publishing research data (Research Data Alliance, 2017).

Good metadata are essential for data interoperability and re-use (Table 6.1). Provenance information tracks the history of a dataset and is an essential part of metadata, necessary to understand both the source of the information and the history of the dataset (it is also important for incentivising data access, as discussed in the section 'A recognition-and-reward system for data producers'). In this regard, the Open Archival Information System (OAIS) reference model is of particular interest. OAIS was initially developed in the context of archival of data from space missions. It is designed to preserve information over the long term and disseminate it to a designated community that should be able to understand the data independently in the form in which it is preserved. OAIS covers the steps of ingesting, preserving and disseminating the data. It is universally accepted as the common language of digital preservation (Lavoie, 2014). An increasing number of repositories strive to be OAIS-compliant, since this ensures the possibility of re-using data in the long term.

Going forward, some policy implications can be drawn for governments:

- The development and adoption of community agreed standards is critical for FAIR data. Individuals and bodies (such as the Research Data Alliance) that work in this area should be supported accordingly.
- Good metadata are critical for data interoperability and re-use. The compliance of data controllers with standardised reference models (such as OAIS) should be encouraged.

### *Definition of responsibility and ownership*

Issues of ownership and responsibility, including copyright and intellectual property need to be considered when enhancing access to public research data, as they can have important implications for how – and by whom – data can be used. Data creators may not necessarily hold the intellectual property rights (IPR) to the data they collect: in the case of human-subject data, for example, the participants themselves may hold those rights.

Most saliently, any IPR associated with research data, and the licensing arrangements for the use of that data, must be clearly specified. In the absence of such specification, data acquire the statutory IPR of the jurisdiction in which they are used. This may include copyright and *sui generis* database rights (e.g. as in Europe), which can greatly inhibit the further use of data. Such protections arise automatically unless expressly excluded, waived or modified (Doldirina et al., 2018).

Legislation and other rules for managing research data are not harmonised across organisations and countries. Data custodians often operate under various legal frameworks governing the collection and use of research data (e.g. Box 6.4 on South Africa). In the United States, for example, different research-funding agencies have different IPR policies (EARTO, 2016). In the European Union, copyright can be claimed on data that may not be copyrightable in other jurisdictions (such as the United States), with implications for the use of text and data mining in research. According to Hargreaves (2011), “Copyright, once the exclusive concern of authors and their publishers, is today preventing medical researchers studying data and text in pursuit of new treatments.”

Tensions between public- and private-sector actors over access to research data are a concern, bearing in mind that one of the main drivers for open data is to improve knowledge transfer and innovation. Enormous potential exists for combining public research data with private-sector data (including social-media data); However, IPR and/or licensing arrangements ensuring both adequate protection of legitimate commercial interests, and the openness and transparency necessary to promote reproducibility and public confidence, are required (OECD, 2016). In this regard, the OECD Principles and Guidelines for Access to Research Data from Public Funding state that: “Consideration should be given to measures that promote non-commercial access and use while protecting commercial interests, such as delayed or partial release of such data” (OECD, 2007).

Going forward, there are a number of policy implications:

- Information about ownership and licensing should be contained within the metadata and specified for all prospective data products in research data management plans. Open-use licences, such as those developed by Creative Commons, should be used, wherever appropriate (OECD, 2015c).
- The implications of any amendments to copyright legislation and IPR regimes, as they relate to access to publically funded data for research, should be carefully considered. They should not inhibit research and innovation in new areas, such as text and data mining, and deep learning.

**Box 6.4. In my view: Greater clarity in intellectual property (IP) and data-management policies can contribute to promoting open-data practice**

Michelle Willmers, Curation and Dissemination Manager of the Global South Research on Open Educational Resources for Development (ROER4D) project, University of Cape Town, South Africa

The ability of researchers to legally share outputs arising from their work is dictated by institutional IP policies, which are in turn largely influenced by national copyright acts. In the African context, many universities have nascent policy environments, meaning that they may not have an IP policy, or it is out of date and inadequate to cover the intricacies of online content sharing – particularly as relates to open data transfer and publication. There are also instances in which policy environments provide conflicting or contradictory stipulations. This situation makes for confusion on the part of academics in terms of what their actual rights are in the context of data sharing ... or, in some cases, may lead to flagrant disregard for policies and mandates.

Both the IP Policy and the Research Data Management Policy of the University of Cape Town (UCT) state that research data are owned by UCT, unless otherwise agreed in research contracts. This may lead many academics to assume they do not have the legal rights to share their data, which is not the case. UCT promotes the use of Creative Commons licensing in its IP Policy, and has a concerted campaign underway to promote responsible data sharing at all levels of the academic enterprise.

Possible confusion in this regard is compounded by the fact that the institutional terms of deposit for sharing data in repositories state that: “UCT grants the Principal Investigator (PI) of a research project the right to upload UCT research data supporting a publication required by a journal publisher or a funder and all UCT project data where this is a specific funder requirement, as long as the data complies with any ethics requirements (e.g. patient confidentiality, consent, etc.)”

This caveat raises questions around the rights of academics who are not operating in research contexts led by PIs, or are functioning in a context where there is no publisher or funder requirement in this regard. The fact that the caveat only exists on a website designed to promote data sharing and is not captured in any of the formal institutional policies regulating data sharing makes the institutional open data policy landscape confusing for academics to navigate, and may serve to build reluctance and confusion, rather than promote a culture of sharing where academics are certain of their legal rights.

Grant agreements and repository deposit terms do increasingly provide exceptions and caveats to restrictive or confusing IP policies, but these agreements are often not adequately scrutinised by academics, and the lack of cohesion between institutional policies, the dictates of funding entities and the intricacies of repository terms and conditions can ultimately amplify the distrust of – and therefore the reluctance to engage with – open-data practice.

National and regional initiatives to assess and revise institutional IP policies so that they are conducive to open data sharing and form part of a set of clear, cohesive institutional stipulations would be extremely valuable in terms of promoting open data practice, and ensuring a functional understanding of the legal and ethical aspects of the process – the uncertainty of which often inhibits academics’ practice in this regard.

### *A recognition-and-reward system for data producers*

Data sharing entails a cultural change among researchers in many fields of science. Appropriate acknowledgement and reward systems need to counterbalance perceived barriers and risks of providing open access to data. The emphasis on competition in research, including the way in which it is evaluated and funded, can be a strong disincentive to openness and sharing.

Researchers have incentives to publish (preferably positive) scientific results. Incentives to publish data are less developed, and usually seen as a constraint imposed by funding agencies and/or publishers. Data citation has not been widely implemented; although the prerequisites for achieving this (e.g. standard formats and citation metrics) already exist, they are not being broadly adopted. Data activities (including those relating to negative results) need to be embedded in evaluation systems, to ensure that researchers who provide high-quality research data are rewarded.

Despite the progress achieved, sharing of research data remains suboptimal. In a 2016 OECD Survey of scientific authors,<sup>4</sup> only 20-25% of corresponding authors had been asked to share data after publication. If asked, a significant share (30-50%) said they would grant access to the data, or at least undertake steps to grant them; about 30% of authors said they would seek to clarify the request. Depending on the discipline, 10-20% of authors would refuse to share data on legal grounds (Boselli and Galindo-Rueda, 2016). Authors of scientific papers are more reluctant to share their data openly than to access data from other research groups (Elsevier and Centre for Science and Technology Studies, 2017).

The TOP Guidelines (Box 6.1) recognise data citation as one of the levers to incentivise data sharing. They propose making data citation mandatory, and citing and referencing all datasets and the codes used in a publication with a DOI (Center for Open Science, 2014). The adoption of unique digital identifiers for researchers, such as the Open Researcher and Contributor ID (see Chapter 12), is also important in this context, as it would greatly simplify provenance mapping and related citation.

Adopting data citation as standard practice so that it can be used to incentivise and reward data sharing also requires developing appropriate data-citation metrics. These could then be used alongside other assessment measures, such as bibliometrics, in recruitment and evaluation processes (OECD, 2018). The approach adopted by the National Science Foundation (NSF) in the United States is an interesting example in this regard. The NSF has implemented an incremental strategy for accessing research data over the past decade. Since 2013, datasets and publications are treated equally as products in the context of an individual researcher's "biosketch". In 2016, the NSF added to the proposal section a requirement to discuss evidence of research products and their availability, including data, in prior NSF-funded research. In France, the newly published national Open Science Plan (Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, 2018) adopts similar principles, pleading for a more qualitative rather than purely quantitative, approach to evaluating researchers. The Open Science Plan is based on the San Francisco Declaration on Research Assessment, which calls for a more holistic evaluation of scientists considering all their research outputs, including data and software (DORA, 2012).

Although recognising data citation and data products in academic evaluation processes may incentivise researchers, it will not necessarily value the critical contribution of data stewards. These are the people who curate and manage data, and ensure their long-term availability and usability (see the section on Future outlook). Career paths for this cohort of data professionals are unclear. Mechanisms to assess their performance should be



distinct from the evaluation mechanisms applied to researchers, but should be linked to the data that they manage. New measures, incentives and reward systems will be required for data stewards.

Going forward, possible policy measures to incentivise and promote data sharing by researchers include:

- developing new indicators/measures for data sharing, and incorporating these into institutional assessment and individual researcher-evaluation processes
- promoting the use of unique digital identifiers for individual researchers and datasets, to enable citation and accreditation
- developing attractive career paths for data professionals, who are necessary to the long-term stewardship of research data and the provision of services.

### *Business models and funding for open data provision*

“Open access” does not necessarily imply “free of charge”. However, many experts agree that public research data should ideally be free at the point of usage (OECD, 2018), implying that the costs of the stewardship and provision will be assimilated by the data provider or repository. These costs can be substantial and require long-term financial commitment, often over several decades. Ultimately, most of the funding for open research data is likely to come from the public purse, although alternative revenue streams exist for some types of data (OECD, 2017b). A key question from the science-policy or funder perspective is how best to allocate this funding. The answer depends on a full understanding of the business models and value propositions of specific data repositories and of the networks in which they are integrated (Figure 6.2).

This must consider multiple factors, including the role of the repository, national and domain contexts; the stage of the repository's development or lifecycle phase; the characteristics of the user community; and the data product required by this community (influencing the level of investment necessary to curate and enhance the data). Business models are constrained by – and need to be aligned with – policy regulation (mandates) and incentives (including funding) (OECD, 2017b).

Many different kinds of data repositories provide a large variety of services, ranging from raw data to complex online analyses. Institutional repositories, national repositories, domain-specific repositories and international repositories are all parts of a complex landscape. This landscape is constantly changing as valuable new data resources arise from projects and transition into longer-term sustainable infrastructures, with longer-term funding requirements. At the level of the individual research system, potential economies of scale can be obtained by centralising or federating the management of data resources; this is common practice in some fields. However, not all data can be transferred across institutional or national boundaries for legal, proprietary or ethical reasons; a certain amount of redundancy in the system can also present some advantages, by making it more resilient. Federated networks can provide some of the benefits of scale, while respecting diversity (OECD, 2017c).

Even when business models are well-developed, and long-term funding is identified, there are limits on how data repositories can operate to provide FAIR access to increasing volumes of data. Priorities need to be established and choices made, e.g. between providing immediate online access or putting data into deep storage. With very big data from experimental facilities (such as the Square Kilometre Array telescope), it is impossible to

provide open online access to all users; thus, tiered access systems have been developed. Prioritisation and data selection will be an increasingly significant challenge in the future. Addressing this challenge will require dialogue with data provider and users, as well as more systematic cost-benefit analyses (bearing in mind that data that may be of little value today can be very valuable tomorrow, and today's users may be different tomorrow).

**Figure 6.2. Creating a value proposition for data repositories**



Source: OECD (2017b).

Research-data repositories and services can also be developed as public-private partnerships. Some private companies are opening their data for non-monetary gain (e.g. for recruiting, improving their image or exchanging data). For instance, medical researchers may want to combine data about people's medical history, genomics, food intake and mobility. Here, medical and genomic data may come from the public sector, while mobility and food data could depend on access to private-sector data. Provided that IPR and ethical issues can be agreed on, public-private partnerships built around such themes should be

encouraged, as they can support the development of data infrastructure and the creation of value-added services. The governance arrangements of such public-private partnerships need to be carefully designed to promote trust among all stakeholders, and ensure transparency and accountability (OECD, 2016).

Going forward, some policy implications can be drawn for governments:

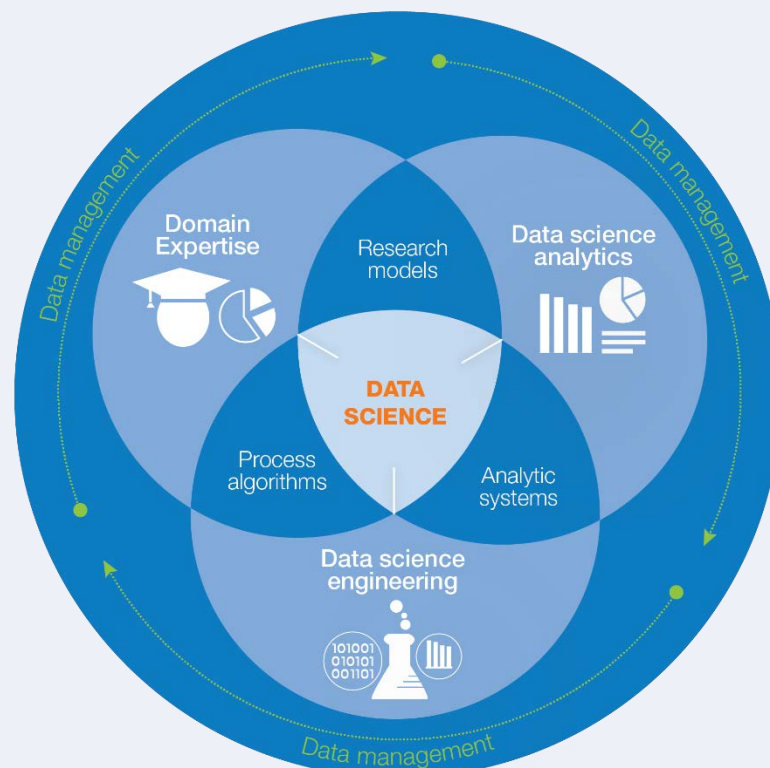
- Develop strategies and roadmaps, including long-term funding plans and business models, to build sustainable research-data infrastructure (i.e. data repositories and services).
- Explore how public investment in research data and infrastructure can be used to leverage private investment (as well as skills and data resources), while ensuring openness and accountability.

### *Building human capital*

Depending on the scientific domain, researchers normally have some training in data analysis, but often lack data-management skills. Users (who may be from different academic sectors or from the private sector) do not always have the appropriate skills to interpret and analyse the data correctly. The effective operation of data repositories requires specialised skills in data curation and stewardship. Various other skills – related to ethical, legal and security issues, as well as risk management, communication and design – should be included in any well-functioning open-data ecosystem. A lack of these skills breeds lack of trust.

“Data science” and “data scientists” are overarching terms encompassing a wide range of skill needs. The National Institute of Standards and Technology Big Data Interoperability Framework (Volume 1)<sup>5</sup> defines a data scientist as “a practitioner who has sufficient knowledge in the overlapping regimes of business needs, domain knowledge, analytical skills and software and systems engineering to manage the end-to-end processes in the data life cycle.” In reality, very few individuals exist in most scientific fields who fit this definition and are leaders in each of these skill areas. Research increasingly depends on collaboration and co-operation between individuals with different data skillsets. Defining the needs and gaps for these skillsets in different scientific fields is a challenge.

Several detailed analyses exist of the data-skill requirements for science, e.g. the Data Science Framework developed by the EDISON project funded by the European Commission<sup>6</sup> (Box 6.5).

**Box 6.5. Data skills**

This diagram illustrates the main competence groups within data science, as defined in the EDISON project: data-science analytics, data-science engineering, and domain knowledge and expertise. Data management, including curation and long-term stewardship, is sometimes classified as part of data science or as a separate competence group. These various competences need to be integrated into the different aspects of the research process, from design to experimentation, analysis and reporting.

Different scientific domains are equipped to varying degrees when it comes to data skills. Traditionally data-intensive fields, such as experimental physics or astronomy, are generally well-positioned (although competition for data scientists from commercial actors is affecting recruitment and retention in academia). Other areas, such as medical research, have significant skill gaps. Moreover, the additional burden of curating and stewarding data to make it available for secondary use creates a human-resource challenge that cuts across all areas of science.

Identifying skill needs and gaps across different research domains is a necessary first step. Meeting these needs is an even greater challenge, which requires a combination of retraining existing personnel (e.g. retraining librarians and archivists to perform data-stewardship functions), and providing new education and training opportunities for researchers and professional research-data support roles. Many initiatives are already taking place in this regard, presenting considerable opportunities for mutual learning across countries and different scientific domains.

Data scientists are in high demand in industry, and academic research competes for the best talent. An urgent need exists to develop recognition-and-reward structures and attractive career paths for all the specialists needed to realise the value of public research data. As in other research areas, workforce diversity will be an important determinant of success (Chapter 7) that should be considered at the outset when developing human-resource strategies for the digital research age.

Going forward, some policy implications can be drawn for governments:

- Develop a national data-skill strategy for STI, identifying specific skill gaps, and the education and training requirements needed to fill them.
- Facilitate co-operation across different education and research actors, to ensure coherence and complementarity in data-skill capacity-building activities.

### Future outlook

The significance of data for STI will undoubtedly continue to increase over the next decade. The volume of data produced globally amounted to 16 zettabytes (ZB)<sup>7</sup> in 2016 and is expected to grow to 163 ZB by 2025 (Reinsel, Gantz and Rydning, 2017). The importance of artificial intelligence in assisting scientific discovery is also expected to grow significantly; access to well-managed data is a key enabler of this development (Kitano, 2016).

Enhanced access to research data holds considerable promise for increasing research productivity and innovation, and developing solutions to complex societal challenges. However, realising this potential, and minimising the potential risks, will require strategic planning and policy interventions. The OECD Recommendation (OECD, 2006) and the more recent FAIR principles for data access provide a broad guiding framework for policy development and co-operation across communities. Many countries have already taken up the challenge and have adopted open-science policies and/or open access to research-data strategies; at the European level, the European Commission has taken the lead in ensuring policy coherence across countries.

Box 6.6 references two possible scenarios. Successful implementation of open-data policies and strategies crucially requires establishing governance systems and processes that ensure transparency and foster trust across the research community and society at large. Mandates and incentives will need to be used judiciously to support and facilitate changes in research behaviour, without stifling creativity and innovation. Long-term investment in technical infrastructure and human capital will be required. Technical standards need to be developed, and legal and ethical concerns addressed.

A lot needs to be done, and a lot is already being done. Understandably, policy intervention focuses on realising the exciting opportunities presented by enhanced access to research data. Open data can help address issues related to the reproducibility and accountability of scientific research; it can help provide solutions to pressing socio-economic challenges; and it can unite the global scientific community around these issues. Looking to the future, however, it is also important to properly consider and mitigate some of the potential risks. The advent of data-driven science coincides with a crisis of confidence in science and the advent of the “post-truth” era.

Opening up public research data means that new actors will be able to analyse and interpret the data from their own perspectives, and not necessarily with the critical objectivity

expected from scientists. The old adage that “if you have enough data, you can prove anything” is not unfounded.

In the new world of open science, the scientific community will need to work rigorously, communicate clearly the scientific method and limitations of its analyses, and engage in honest discourse and dialogue with public and policy makers. In a hyper-competitive research enterprise characterised by enormous pressure to succeed and growing hype around scientific breakthroughs, there is a need to ensure that open science and data can be trusted. Technological developments can assist in this regard. Ultimately, however, trust is a social construct that needs to be carefully nurtured over time.

#### **Box 6.6. Possible future scenarios for access to data for STI**

In a desirable future scenario, trust would be earned across society through strong governance initiatives. These would ensure robust risk management and mitigation, elaborated in transparent consultation with stakeholders. Ethics review boards would credibly represent individual interests and arbitrage consent issues. On the technical side, strong global standards would emerge, akin to Transmission Control Protocol/Internet Protocol for Internet communication, complemented by more specialised standards for specific applications. IPR and licensing provisions would promote responsible data access and re-use and be a standard part of machine-readable metadata. Data citation would become ubiquitous and an integral part of researcher evaluation. Financing of repositories would be based on long-term infrastructure strategies and sustainable models. Finally, digital skills would be addressed through a strategic approach encompassing initial education and lifelong learning for data producers, stewards and users.

A “worst-case” scenario is also possible, in which repeated security and privacy breaches would be inadequately managed, fostering a general level of mistrust. Standards would continuously lag behind technology development while IPR would be insufficiently defined to support widespread data re-use. Incentives for researchers to publish their data would remain weak, and initiatives to support data skills development would be poorly designed.

## Notes

<sup>1</sup> “EOSC Declaration: European Open Science Cloud – New Research & Innovation Opportunities”: [https://ec.europa.eu/research/openscience/pdf/eosc\\_declaration.pdf#view=fit&pagemode=none](https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf#view=fit&pagemode=none).

<sup>2</sup> Regulation 2016/679 defines “consent” of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her.

<sup>3</sup> [http://archive.stats.govt.nz/browse\\_for\\_stats/snapshots-of-nz/integrated-data-infrastructure.aspx](http://archive.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure.aspx).

<sup>4</sup> <http://oe.cd/issa>.

<sup>5</sup> [https://bigdatawg.nist.gov/V1\\_output\\_docs.php](https://bigdatawg.nist.gov/V1_output_docs.php).

<sup>6</sup> <http://edison-project.eu/>.

<sup>7</sup> 1 ZB = 1 trillion gigabytes, or  $10^{21}$  bytes.

## 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.
- Boselli, B. and F. Galindo-Rueda (2016), “Drivers and Implications of Scientific Open Access Publishing: Findings from a Pilot OECD International Survey of Scientific Authors”, *OECD Science, Technology and Industry Policy Papers*, No. 33, OECD Publishing, Paris, <http://dx.doi.org/10.1787/5jlr2z70k0bx-en>.
- Center for Open Science (2014), “Transparency and Openness Promotion (TOP) Guidelines”, Center for open Science, Charlottesville, VA, <http://cos.io/top>.
- Doldirina, C. et al. (2018), “Legal Approaches for Open Access to Research Data”, *LawArXiv Papers*, Cornell University, Ithaca, NY, <http://dx.doi.org/10.31228/OSF.IO/N7GFA>.
- DORA (2012), *San Francisco Declaration on Research Assessment (DORA)*, <https://sfedora.org/> (accessed on 07 July 2018).
- EARTO (2016), “EARTO Background Note: Overview of US Federal Agencies Data Sharing Policies”, Defense Technical Information Center, Fort Belvoir, VA, [http://www.dtic.mil/dtic/pdf/dod\\_public\\_access\\_plan\\_feb2015.pdf](http://www.dtic.mil/dtic/pdf/dod_public_access_plan_feb2015.pdf) (accessed on 19 July 2018).
- Elsevier and Centre for Science and Technology Studies (2017), *Open data: The researcher perspective*, Elsevier and Centre for Science and Technology Studies, Amsterdam and Leuven, [https://www.elsevier.com/\\_data/assets/pdf\\_file/0004/281920/Open-data-report.pdf](https://www.elsevier.com/_data/assets/pdf_file/0004/281920/Open-data-report.pdf) (accessed on 26 July 2018).
- EC/OECD (2017), *STIP Compass: International Science, Technology and Innovation Policy (STIP)* (database), edition 2017, <https://stip.oecd.org>.
- Economic and Social Research Council (2015), “Research Data Policy”, webpage, <https://esrc.ukri.org/funding/guidance-for-grant-holders/research-data-policy/> (accessed on 19 June 2018).
- eRSA (2014), “The Australian Research Cloud: Your most powerful eResearch tool yet – eRSA”, webpage, <https://www.ersa.edu.au/the-australian-research-cloud-your-most-powerful-eresearch-tool-yet/> (accessed on 8 November 2018)
- European Commission (2018a), “Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the re-use of public sector information (recast)”, <http://ec.europa.eu/transparency/regdoc/rep/1/2018/EN/COM-2018-234-F1-EN-MAIN-PART-1.PDF>
- European Commission (2018b), “COMMISSION RECOMMENDATION (EU) 2018/790 of 25 April 2018 on access to and preservation of scientific information”, Official Journal of the European Union, Vol. L/134, pp. 12-18, <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2018:134:FULL&from=EN>.
- European Commission (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)”, Official Journal of the European Union, <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN>.
- European Commission (2014), “Validation of the results of the public consultation on Science 2.0: Science in Transition”, European Commission, Brussels, [http://ec.europa.eu/research/consultations/science-2.0/science\\_2\\_0\\_final\\_report.pdf](http://ec.europa.eu/research/consultations/science-2.0/science_2_0_final_report.pdf).
- European Commission (2012), “COMMISSION RECOMMENDATION of 17.7.2012 on access to and preservation of scientific information”, European Commission, Brussels, [https://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/recommendation-access-and-preservation-scientific-information\\_en.pdf](https://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf).

- Hargreaves, I. (2011), “Digital Opportunity: A Review of Intellectual Property and Growth An Independent Report”, ORCA Online Research @Cardiff, Cardiff University, [https://orca.cf.ac.uk/30988/1/1\\_Hargreaves\\_Digital%20Opportunity.pdf](https://orca.cf.ac.uk/30988/1/1_Hargreaves_Digital%20Opportunity.pdf).
- Kitano, H. (2016), “Artificial Intelligence to Win the Nobel Prize and Beyond: Creating the Engine for Scientific Discovery”, *AI Magazine*, Vol. 37/1, p. 39, Association for the Advancement of Artificial Intelligence, Palo Alto, CA, <http://dx.doi.org/10.1609/aimag.v37i1.2642>.
- Lavoie, B. (2014). The Open Archival Information System (OAIS) Reference Model: Introductory Guide (2nd Edition) Principal Investigator for the Series Neil Beagrie. Retrieved 07 03, 2018, from <https://www.dpconline.org/docs/technology-watch-reports/1359-dpctw14-02/file>
- Ministère de l’Enseignement supérieur, de la Recherche et de l’Innovation (2018), “Plan national pour la Science ouverte, mercredi 14 juillet 2018”, Ministère de l’Enseignement supérieur, de la Recherche et de l’Innovation, Paris, [http://cache.media.enseignementsup-recherche.gouv.fr/file/Actus/67/2/PLAN\\_NATIONAL\\_SCIENCE\\_OUVERTE\\_978672.pdf](http://cache.media.enseignementsup-recherche.gouv.fr/file/Actus/67/2/PLAN_NATIONAL_SCIENCE_OUVERTE_978672.pdf) (accessed on 17 September 2018).
- Morais, R. and L. Borrell-Damian (2018), “2016-2017 EUA SURVEY RESULTS”, European University Association, Brussels, <http://www.eua.be/Libraries/publications-homepage-list/open-access-2016-2017-eua-survey-results> (accessed on 19 June 2018).
- NIH (2017). ERA Commons - NIH. Retrieved 08 November 2018, from <https://public.era.nih.gov/commons/public/login.do>.
- OECD (2018), “Towards New Principles For Enhanced Access To Public Data For Science, Technology And Innovation”, Joint CSTEP-GSF Workshop, March 13, 2018, Paris.
- OECD (2017a), “Open access to data in science, technology and innovation – initial survey findings”, OECD, Paris.
- OECD (2017b), “Business models for sustainable research data repositories”, *OECD Science, Technology and Industry Policy Papers*, No. 47, OECD Publishing, Paris, <https://doi.org/10.1787/302b12bb-en>.
- OECD (2017c), “Co-ordination and Support of International Research Data Networks”, webpage, <http://www.oecd.org/going-digital> (accessed on 16 July 2018).
- OECD (2016), “Research Ethics and New Forms of Data for Social and Economic Research”, *OECD Science, Technology and Industry Policy Papers*, No. 34, OECD Publishing, Paris, <http://dx.doi.org/10.1787/5jln7vnpxs32-en>.
- OECD (2015a), “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>.
- OECD (2015b), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264229358-en>.
- OECD (2015c), “Assessing government initiatives on public sector information: A review of the OECD Council Recommendation”, *OECD Digital Economy Papers*, No. 248, OECD Publishing, Paris, <http://dx.doi.org/10.1787/5js04dr9147j-en>.
- 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, Paris, <http://www.oecd.org/sti/sci-tech/new-data-for-understanding-the-human-condition.pdf>.
- OECD (2009), “Access to Research Data: Progress on Implementation of the Council Recommendation”, *Science, Technology and Innovation for the 21st Century. Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level, 29-30 January 2004 - Final Communiqué*, OECD, Paris, [http://www.oecd.org/document/15/0,3343,en\\_2649\\_34269\\_25998799\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/15/0,3343,en_2649_34269_25998799_1_1_1_1,00.html).
- OECD (2007), *OECD Principles and Guidelines for Access to Research Data from Public Funding*, OECD Publishing, Paris, <https://doi.org/10.1787/9789264034020-en-fr>.
- OECD (2006), *Recommendation of the Council concerning Access to Research Data from Public Funding*, OECD, Paris, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347>.



- President’s Advisors on Science and Technology (2014), “Report to the President: Big data and privacy: a technological perspective”, The White House, Washington DC, [https://bigdatawg.nist.gov/pdf/pcast\\_big\\_data\\_and\\_privacy\\_-\\_may\\_2014.pdf](https://bigdatawg.nist.gov/pdf/pcast_big_data_and_privacy_-_may_2014.pdf).
- Research Data Alliance (2017), “All Recommendations & Outputs – RDA Endorsed Recommendations”, webpage, <https://www.rd-alliance.org/recommendations-and-outputs/all-recommendations-and-outputs> (accessed on 25 September 2017).
- Wilkinson, M. et al. (2016), “The FAIR Guiding Principles for scientific data management and stewardship”, *Scientific Data* 3, Article No. 160018 (2016), *Nature Communications*, <http://dx.doi.org/10.1038/sdata.2016.18>.





**From:**  
**OECD Science, Technology and Innovation  
Outlook 2018**  
Adapting to Technological and Societal Disruption

**Access the complete publication at:**  
[https://doi.org/10.1787/sti\\_in\\_outlook-2018-en](https://doi.org/10.1787/sti_in_outlook-2018-en)

**Please cite this chapter as:**

OECD (2018), "Enhanced access to publicly funded data for STI", in *OECD Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption*, OECD Publishing, Paris.

DOI: [https://doi.org/10.1787/sti\\_in\\_outlook-2018-11-en](https://doi.org/10.1787/sti_in_outlook-2018-11-en)

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to [rights@oecd.org](mailto:rights@oecd.org). Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at [info@copyright.com](mailto:info@copyright.com) or the Centre français d'exploitation du droit de copie (CFC) at [contact@cfcopies.com](mailto:contact@cfcopies.com).