

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



The Current Policies of Research Funders and Publishers

Vera J. Lipton

This chapter provides a review of the principles underpinning open scientific data and the policies mandating open access to scientific data. It has a specific focus on the policies of research funders and journal publishers.

The chapter consists of five parts, as follows:

1. Main international developments
2. Key policies of research funders
3. Selected policies of publishers
4. Issues covered in the open data policies
5. Open scientific data in emerging and developing countries

Introduction

Increased data sharing among scientists and with non-scientists can generate vast benefits to society and to the economy. Yet creating conditions conducive to data sharing remains a challenge. Inspired by the positive experience with open publications, similar policies have been introduced in recent years with a view to facilitating greater sharing of research data.

This chapter surveys open data policies, paying particular attention to the scope of the open data mandates. It starts with an overview of major international developments and declarations that have inspired governments and research funders to introduce open data policies. This is followed by an analysis of the policies of research funders and publishers in several jurisdictions. Next is identification of the components of ideal data sharing policies. The final section surveys the open data landscape in emerging and developing countries.

3.1 Main international developments

3.1.1 Early policies in the United States

Some of the world's leading research organisations are based in the United States. Many of them were also among the first in the world to recognise the potential of open science. The first policy statement for open access to research data consists of the Bromley Principles issued by the United States Global Change Research Program in 1991 [169]. Five years later the Bermuda Principles—developed

as part of the Human Genome Project—established an international practice in the sharing of genomic data prior to publication of research findings in scientific journals [170]. These principles of free release and data sharing have been one of the major outputs of the Human Genome Project and have established the practice of genomic data sharing globally.

The *Access to Databases Principles* first published by the International Council for Science/Committee on Data for Science and Technology (ICSU/CODATA) in 2002 provided a further impetus for promoting open access to scientific data among policymakers [171]. The principles were developed to facilitate the evaluation of legislative proposals that may affect the use of scientific databases.

3.1.2 The Berlin Declaration

The Human Genome Project was declared complete in 2003. In the same year, open access to scientific data was first codified internationally, in the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. The declaration emerged from a conference hosted by the Max Planck Institute in Munich and represents a landmark statement on open access to scientific contributions¹ including:

... original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material [63].

Such scientific contributions need to satisfy two conditions to qualify as ‘open’:

First, the author(s) and right holder(s) of such contributions grant(s) to all users a free, irrevocable, worldwide, right of access to, and a licence to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship ... as well as the right to make small numbers of printed copies for their personal use.

Second, a complete version of the work and all supplemental materials, including a copy of the permission as stated above, in an appropriate standard electronic format is deposited (and thus published) in at least one online repository using suitable technical standards (such as the Open Archive definitions) that is supported and maintained by an academic institution, scholarly society, government agency, or other well-established organisation that seeks to enable open access, unrestricted distribution, interoperability and long-term archiving.²

Organisations committed to implementing these objectives and the two key principles can sign the declaration. As of October 2007, there were over 240 signatories, mostly research organisations. As of early June 2018, the number of signatories had reached 620 [172].

3.1.3 UNESCO and open science

The United Nations Educational, Scientific and Cultural Organization (UNESCO) is the only UN agency with a specific mandate for science. One of its main functions, articulated in the UNESCO constitution, is to:

¹ The Berlin Declaration does not use the term ‘open research data’ but rather refers to ‘open knowledge contributions’ which represent a broad definition of open research data. See also discussion concerning the definition of research data in the next chapter.

² *Ibid.* This definition of open data and open access is further discussed in Chapters 4 (Section 4.5) and Chapter 7 (Section 7.3) of this book.

*... maintain, increase and diffuse knowledge: by assuring the conservation and protection of the world's inheritance of books, works of art and monuments of history and science, and recommending to the nations concerned the necessary international conventions.*³

At the same time, facilitating the sharing of scientific outcomes is only one of the many responsibilities assigned to UNESCO. Perhaps for this reason the organisation has not played a pivotal role in recommending any international conventions for open science in recent years. Many provisions of the UNESCO Declaration on Science and the Use of Scientific Knowledge—adopted in 1999—are now outdated due to rapid technological developments and changing methods of science production and dissemination.⁴

Having said that, one of the key objectives articulated in the Strategy on UNESCO Contribution to the Promotion of Open Access to Scientific Information and Research is to convene an international congress on scholarly communication to examine the feasibility of developing a UNESCO convention on open access for scientific information and research ([29], p. 13).

More recently, UNESCO endorsed several open science initiatives, including the Open Science for the 21st Century Declaration by All European Academies,⁵ which encourage scientists and their organisations, particularly publicly funded organisations, to apply open-sharing principles to the data underpinning research publications, including negative results. The Declaration also calls for measures to ensure data quality and preservation to enable future reuse.⁶

In addition, UNESCO supports several public education projects aimed at raising awareness of open access, including in developing countries. In 2012, UNESCO issued Policy Guidelines for the Development and Promotion of Open Access written by Swan [27]. The report notes that:

Research data are increasingly covered by policies and often these policies are being implemented by smaller, niche players as well as large research funders. These policies are not usually, however, the same (Open Access) policies that cover the text-based literature. Data are exceptional because policies must take into account issues of privacy and special cases where data cannot be released for other reasons. Developing and wording Open Data policies is therefore a specialised issue that is not as straightforward as developing policies for Open Access to the literature. Where

³ Article 1, Clause 2 of the UNESCO Constitution.

⁴ Article 38 of the Declaration states: 'Intellectual property rights need to be appropriately protected on a global basis, and access to data and information is essential for undertaking scientific work and for translating the results of scientific research into tangible benefits for society. Measures should be taken to enhance those relationships between the protection of intellectual property rights and the dissemination of scientific knowledge that are mutually supportive. There is a need to consider the scope, extent and application of intellectual property rights in relation to the equitable production, distribution and use of knowledge. There is also a need to further develop appropriate national legal frameworks to accommodate the specific requirements of developing countries and traditional knowledge and its sources and products, to ensure their recognition and adequate protection on the basis of the informed consent of the customary or traditional owners of this knowledge'.

⁵ A declaration of ALL European Academies (ALLEA) presented at a special session with Mme Neelie Kroes, Vice-President of the European Commission, and Commissioner in charge of the Digital Agenda on occasion of the ALLEA General Assembly held at Accademia Nazionale dei Lincei, Rome, on 11–12 April 2012.

⁶ *Ibid*, 5.

*there is Open Access policy development now, Open Data policy development will follow.*⁷

In recent years, UNESCO has taken a more active role in developing open scientific repositories. One recent example is the World Library of Science [173], an online repository of short e-books and articles, developed in partnership with the publishers Nature Education and the pharmaceutical company Roche. This currently contains resources in the field of genetics intended for university undergraduate faculties and students. The platform enables science teachers and students from all parts of the world to exchange views, information and knowledge.

3.1.4 The OECD principles for access to research data from public funding

In January 2004, the ministers of science and technology from OECD countries and from China, Israel, Russia and South Africa adopted a Declaration on Access to Research Data from Public Funding. They also called on the OECD to develop a set of guidelines based on commonly agreed principles to facilitate optimal cost-effective access to digital research data [174]. The OECD responded with such a set of principles, published in late 2006, which highlighted the importance of open access to publicly funded research data.⁸ The principles held that open access has a vast potential to improve the scientific and social return on public investment [175]. The OECD noted, however, that the level of public research funding varies significantly across countries, as do data access policies and practices at the national, disciplinary and institutional levels. The OECD Principles, summarised below, were developed with a view to providing broad policy recommendations to governments, research organisations and funding bodies:

Principle A. Openness—access on equal terms and at the lowest possible cost. Open access to research data should be easy, timely, user-friendly and preferably Internet-based.

Principle B. Flexibility—recognising the rapid and often unpredictable changes in information technologies, the characteristics of each research field and the diversity of research systems, legal systems and cultures of each member country.

Principle C. Transparency—information on research data and data-producing organisations and the conditions attached to the use of the data should be available in a transparent way, ideally through the Internet.

Principle D. Legal conformity—data access arrangements should respect the legal rights and legitimate interests of all stakeholders in a public research enterprise. Subscribing to professional codes of conduct may facilitate meeting legal requirements.

Principle E. Protection of intellectual property—data access arrangements should consider the applicability of copyright and other intellectual property laws that may be relevant to research databases. At the same time, the fact that there is private sector involvement in the data collection or that the data may be protected by intellectual property laws should not be used as a reason to restrict access to the data.

Principle F. Formal responsibility—formal institutional practices should be promoted. These include rules and regulations regarding the responsibilities of the

⁷ *Ibid*, 47.

⁸ The Principles define research data from public funding as the research data obtained from research conducted by government agencies or departments or conducted using public funds provided by any level of government.

various parties involved in data-related activities. The issues to be covered include authorship, producer credits, ownership, dissemination, usage restrictions, financial arrangements, ethical rules, licencing terms, liability and sustainable archiving.

Principle G. Professionalism—institutional arrangements for the management of research data should be based on relevant professional standards and values embodied in the codes of conduct of the scientific communities involved.

Principle H. Interoperability—technological and semantic interoperability is the key consideration in enabling and promoting international and interdisciplinary access to, and use of, research data. Member countries and research institutions should cooperate with international organisations in developing data documentation standards.

Principle I. Quality—data managers and data collection organisations should pay particular attention to ensuring compliance with explicit data quality standards.

Principle J. Security—supporting the use of techniques and instruments to guarantee the integrity and security of research data. Data integrity means completeness of the data and absence of errors. Security means that the data, along with relevant metadata and descriptions, should be protected against intentional or unintentional loss, destruction, modification and unauthorised access.

Principle K. Efficiency—improve the overall efficiency of publicly funded scientific research by avoiding unnecessary duplication of data collection efforts.

Principle L. Accountability—data access arrangements should be subject to periodic evaluation by user groups, responsible institutions and research funding agencies.

Principle M. Sustainability—research funders and research institutions should consider long-term preservation of data at the outset of each new project and determine appropriate archiving mechanisms for the data.

These core OECD Principles were the early guidelines for policymakers to promote open data, including open research data. These principles have been widely adopted. However, the definition of research data in this source is very narrow, referring to research data as:

...factual records used in primary sources ... that are commonly accepted in the scientific community as necessary to validate research findings.

Later documents have adopted a far broader approach to research data. These more recent policies are discussed in the following sections.

3.1.5 The Denton Declaration (2012)

In May 2012 [176], at the University of North Texas, a group of technologists and librarians, scholars, researchers and university administrators gathered to discuss best practices and emerging trends in research data management. Resulting from this discussion was a vision for openness in research data titled ‘The Denton Declaration: An Open Data Manifesto’. The declaration includes 6 declarations, 13 principles and 7 intentions.

The principles set out general guidelines for open data in science:

1. Open access to research data benefits society and facilitates decision-making for public policy.
2. Publicly available research data helps promote a more cost-effective and efficient research environment by reducing redundancy of efforts.

3. Access to research data ensures transparency in the deployment of public funds for research and helps safeguard public goodwill towards research.
4. Open access to research data facilitates validation of research results, allows data to be improved by identifying errors and enables the reuse and analysis of legacy data using new techniques developed through advances and changing perceptions.
5. Funding entities should support reliable long-term access to research data as a component of research grants due to the benefits that accrue from the availability of research data.
6. Data preservation should involve sufficient identifying characteristics and descriptive information so that others besides the data producer can use and analyse the data.
7. Data should be made available in a timely manner: neither too soon to ensure that researchers benefit from their labour nor too late to allow for verification of the results.
8. A reasonable plan for the disposition of research data should be established as part of data management planning, rather than arbitrarily claiming the need for preservation in perpetuity.
9. Open access to research data should be a central goal of the lifecycle approach to data management, with consideration given at each stage of the data lifecycle to what metadata, data architecture, and infrastructure will be necessary to support data discoverability, accessibility and long-term stewardship.
10. The costs of cyberinfrastructure should be distributed among the stakeholders—including researchers, agencies and institutions—in a way that supports a long-term strategy for research data acquisition, collection, preservation and access.
11. The academy should adapt existing frameworks for tenure and promotion and merit-based incentives to account for alternative forms of publication and research output including data papers, public datasets and digital products. Value inheres in data as a stand-alone research output.
12. The principles of open access should not be in conflict with the intellectual property rights of researchers, and a culture of citation and acknowledgement should be cultivated rigorously and conscientiously among all practitioners.
13. Open access should not compromise the confidentiality of research subjects and will comply with principles of data security, HIPAA, FERPA [177, 178], and other privacy guidelines.

The intentions articulated the issues of most importance to librarians at the time. They include developing a culture of openness in research, building the infrastructure that is extensible and sustainable for archiving and making the data discoverable, developing metadata standards and recognising and supporting the intellectual property rights of researchers.

The principles are widely known among librarians in the United States and in other countries.

3.1.6 Other statements and policies supporting open scientific data

Several statements and policies have emerged promoting the dissemination of scientific data in online spaces following adoption of the Berlin Declaration and the OECD Open Access Principles. In 2009 the Toronto Statement reaffirmed earlier principles relating to the prepublication release of genomic data and recommended these principles be extended to other types of large biological datasets [179]. The Rome Agenda called for scientific data to be released immediately after the publication of journal articles [180]. The Panton Principle for Open Data in Science—developed in 2010—provides guidelines on licencing of open scientific data [181]. In early 2015, the Research Data Alliance released draft principles on the legal interoperability of research data [182]. These initiatives have facilitated broadening the scope and coverage of open access to research data to include prepublished, published and unpublished data—particularly data generated from publicly funded research.

Many attempts to define the principles of open scientific data also incorporate the challenges associated with implementation, thus restraining the scope for data sharing. These include legal, ethical and commercial limitations on data release; early availability and long-term preservation of research data; the management and curation of the data, metadata and software; sharing the costs of developing research data infrastructures; developing incentives and reward structures; facilitating searchability of the data; and respecting the privacy of research subjects. The challenges are clearly articulated in more recent and more comprehensive sets of principles for open scientific data, summarised below and canvassed in Chapters 4–7.

3.2 Key policies of research funders

For several years now, leading funders of research have required grant recipients to share their data with other investigators. However, originally they had no policies on how this should be accomplished. The game has changed completely in recent years, with many funders requiring the recipients of grants to enable open access to research data and, often, requiring the submission of research data management plans at the grant proposal stage. Such policies ensure that data resulting from publicly funded research is retained and can be reused over time—usually 3–10 years.

Research organisations and universities are largely dependent on grant funding. Suddenly, these institutions realised that to enable researchers to successfully compete for grants, they had to provide support in the formulation of data management plans. Libraries, too, have taken up this approach, and researchers are changing their research data management practices as a result. Within the past decade, the policies introduced by research funders appear to have built a momentum for significant organisational and behavioural changes, and these changes are driving the retention and sharing of research data globally.

3.2.1 The United States

The funders of research in the United States are the leaders when it comes to open research data. The National Institutes of Health (NIH) were among the first to

introduce open access deposit of peer-reviewed journal articles in PubMed Central as a condition of receipt of grant funds.⁹ The NIH also:

... expects a data sharing plan for all proposals over \$500,000 per year in direct costs. Some research communities have developed their own policies [183] in which sharing is expected—and executed—for all grants, not just those over the \$500,000 threshold [184].

Awareness of the need to develop data management infrastructure took a leap forward in 2010 when the National Science Foundation (NSF) announced that it, too, would begin requiring data management plans with applications. Proposals submitted to NSF on or after 18 January 2011:

... must include a supplementary document of no more than two pages labelled 'Data Management Plan'. This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results [185].

Importantly, the data management plan is to be included with every application for NSF funding, even if the plan is a statement that 'no detailed plan is needed'. According to the NSF policy:

Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. Investigators and grantees are encouraged to share software and inventions created under the grant or otherwise make them or their products widely available and usable.¹⁰

The US government has taken significant steps to enable the dissemination of scientific outcomes arising from public research. On 22 February 2013, the Office of Science and Technology Policy at the White House issued the memo 'Increasing Access to the Results of Federally Funded Scientific Research'. It directed each federal agency with over US\$100 million in annual research and development expenditure to develop plans to make 'the results of unclassified research arising from public funding publicly accessible to search, retrieve and analyse and to store such results for long-term preservation'.¹¹ The research results include peer-reviewed publications, publication metadata and digitally formatted scientific data. The major shortcoming is that the memo does not mention metadata associated with research data. This omission is unfortunate because, in many cases, scientific data without metadata is unlikely to be reusable.

The memo also directed agencies to ensure that intramural researchers and all extramural researchers receiving federal grants and contracts for scientific research

⁹ The NHS requires that 'an electronic version of all final peer-reviewed journal articles accepted for publication on and after 7 April 2008 be made publicly available no later than 12 months after the date of publication'.

¹⁰ See NSF Award and Administration Guide, Chapter VI—Other Post Award Requirements and Considerations, points 4(b) and (c). Available at: http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/aag_6.jsp#VID4

¹¹ *Ibid*, p. 3.

have data management plans in place along with mechanisms to ensure compliance with the plans. To support the implementation of data management plans, grant proposals may include appropriate costs for data management and access. Further, agencies are to promote the deposit of data in publicly accessible repositories and develop approaches for identifying and providing appropriate attribution to scientific datasets.

The memo builds on the NIH and NSF open data mandates and covers all larger federally funded organisations. Prior to the memorandum, only six federal funders of research had in place policies requiring the retention and sharing of research data—NIH, NSF, the National Aeronautics and Space Administration, the National Oceanic and Atmospheric Organisation, and the National Endowment for the Humanities, Office of Digital Humanities [186].

3.2.2 The European Union

The European Commission was one of the first major research funders to recognise open access to research data. The Commission considers that facilitating broader access to scientific publication and data can improve the quality of research results, foster collaboration, avoid duplication of research effort and improve the transparency of scientific enquiry—including through increased involvement by citizens [187]. Increasing access to the outcomes of publicly funded research lies at the core of the European policies. Underlying this vision is realisation that research outcomes originating from public sources should not require payment with each access or use. Instead, the outcomes should be preserved and made freely available for the benefit of all.

Open access to science falls broadly under three flagship initiatives of the Commission—namely the Digital Agenda for Europe [188], the Innovation Union Policy [189], and the European Research Area Partnership [190]. The Recommendation on Access to and Preservation of Scientific Information [191], published in July 2012, encourages European Union member states to develop policies for open access to scientific results, including research data and information. The Commission further stated that such policies should include concrete objectives and indicators of progress, implementation plans and appropriate funding mechanisms.¹² The Communication of the Commission regarding open access is not binding on European Union member states, and they are free to adopt any policy that best suits the needs of their own scientific communities. Some countries—Germany, Spain and the Netherlands—have legislated open access to scientific publications and data [192].

The European Commission was among the first large funders to test funding arrangements that encourage open access to publicly funded research. In 2008, the Commission launched the Open Access Pilot as part of its Framework Program 7 (later replaced by the Horizon 2020 Pilot) for data underlying publications, including curated data and raw data [21]. The Rules of Participation [193] represent the legal basis for open access to research data funded by the European Commission under Horizon 2020:

With regard to the dissemination of research data, the grant agreement may, in the context of the open access to and the preservation of research data, lay down terms and conditions under which open access to such results shall be provided, in particular in ERC (European Research Council) frontier research and FET (Future and Emerging Technologies) research or in other appropriate areas, and taking into

¹² *Ibid*, Recommendation 3.

*consideration the legitimate interests of the participants and any constraints pertaining to data protection rules, security rules or intellectual property rights. In such cases, the work programme or work plan shall indicate if the dissemination of research data through open access is required.*¹³

These principles are translated into specific requirements in the Model Grant Agreement¹⁴ under the Horizon 2020 Work Programme. The Commission has also developed a user guide that explains the provisions of the Model Grant Agreement to applicants and beneficiaries, including guidance for open scientific data, as follows:

Regarding the digital research data generated in the action, the beneficiaries [participating in the open research data pilot] must:

- a. Deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate—free of charge for any user—the following:
 - i. The data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible
 - ii. Other data, including associated metadata, as specified and within the deadlines laid down in the ‘data management plan’
- b. Provide information—via the repository—about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and, where possible, provide the tools and instruments themselves).¹⁵

The guidelines also define exceptions to data sharing. These include the obligation to protect research results with intellectual property, confidentiality and security obligations and the need to protect personal data and specific cases in which open access might jeopardise the project. If any of these exceptions is applied, then the data research management plan must state the reasons for not giving or restricting access.

3.2.3 European Research Council

The European Research Council (ERC) is a leading funder of research in the sciences and humanities. The ERC regards open access as the most effective way for ensuring that the fruits of the research it funds can be accessed, read and used in further research. On that basis, the ERC:

... considers it essential that primary data, as well as data-related products such as computer codes, is deposited in the relevant databases as soon as possible, preferably immediately after publication and in any case not later than 6 months after the date of publication [194].

The guidelines also list discipline-specific repositories. The recommended repository for life sciences is the Europe PubMed Central [195] (formerly known as

¹³ *Ibid.*

¹⁴ Multi-beneficiary General Model Grant Agreement, Version 1.0 11 December 2013.

¹⁵ Annotated Model Grant Agreement, Version 1.7, 19 December 2014, 215.

UK PubMed Central), and for physical sciences and engineering, the recommendation is to use ArXiv [196].

3.2.4 The United Kingdom

The peak body for research councils in the United Kingdom, Research Councils UK (RCUK, now transitioned into UK Research and Innovation),¹⁶ instituted policies on open access in 2005 and their Common Principles for Open Data [198] that took account of the evolving global policy landscape. These Principles encouraged the practice of making research data openly available, with as few restrictions as possible, in a timely and responsible manner.¹⁷ The Principles further addressed a number of important issues.

Firstly, data management policies and plans should be in accordance with community best practice and relevant standards set by research institutions themselves.¹⁸ The onus for ensuring that legal, ethical and commercial issues are considered lies with research institutions, and these issues should be considered at all stages in the research process.¹⁹

Secondly, published results should always include information on how to access the supporting data. Metadata should be recorded and made openly available.²⁰

Thirdly, the principles allow for the delay in data release to enable the original data collectors to publish the results of their research.²¹

Finally, public funds can be used to support the management and sharing of publicly funded research data.²² At the same time, research organisations are responsible for ensuring there are enough resources allocated to research data management—for example, from research grants. RCUK clarified in 2013 that all costs associated with research data management are eligible expenditure of research grant funds, but the expenditure must be incurred before the end date of the grant [199].

Open data is thus defined as an integral part of doing research, and the costs are front-loaded into that research. This can initially make the conduct of research more expensive, but significant savings are realised down the track through the recycling of research data and improved quality of research outcomes. These principles are important as they address the concerns raised by several organisations and scientists who pointed out that open scientific data should not be an unfunded mandate [200].

Since the release of RCUK Common Principles on Data Policy in 2011, many member funding organisations have mandated the requirement for a data management plan with each new application. Most research funders in the United Kingdom have issued data policies; however, the extent and coverage of these vary greatly [201].

The RCUK policy on open access states:

Peer-reviewed research papers which result from research that is wholly or partially funded by the research councils:

¹⁶ Subsumed into UK Research and Innovation in [197].

¹⁷ *Ibid*, bullet point 2.

¹⁸ *Ibid*, bullet point 3.

¹⁹ *Ibid*, bullet point 5.

²⁰ *Ibid*, bullet point 4.

²¹ *Ibid*, bullet point 6.

²² *Ibid*, bullet point 8.

1. *Must be published in journals which are compliant with Research Council policy on Open Access*
2. *Must include details of the funding that supported the research and a statement on how the underlying research materials—such as data, samples, or models—can be accessed* [202]

Unlike the United States, where institutional approaches to research data management are developing, most research councils in the United Kingdom ‘place the responsibility on individual researchers to provide evidence that data management and sharing issues have been considered’ [203].

However, one research council—the Engineering and Physical Sciences Research Council (EPSRC)—took a different approach. The EPSRC encouraged research organisations to develop their specific approaches to data management, appropriate to their own structures and cultures. At the same time, these approaches were required to align with the EPSRC’s expectations. To that end, EPSRC requested that applicant institutions develop road maps for open data management. These requirements appear to have acted as a catalyst for developing data management policies and support systems in many UK research organisations.

In 2015, RCUK provided publicly funded research institutions and investigators with explanatory text on each of the seven ‘common principles’ first developed in 2005. This guidance was intended to inform the RCUK consultation on a draft Concordat on Open Research Data²³—a broader network of stakeholders and interested parties in open data. The Concordat committed to the seven ‘common principles’ adopted by the RCUK.

3.2.5 Australia

The Australian Government was among the first to invest in the development of research data infrastructure. The Australian National Data Service (ANDS) was established in 2008 to develop an Australian Research Data Commons platform [204]—an Internet-based discovery service designed to provide rich connections between data, projects, researchers and institutions. Funding was also allocated for the development of metadata tools through the ‘Seeding the Commons’ initiative.

Open research data is a priority area for the Data to Decisions Cooperative Research Centre established in July 2014. The centre brings together researchers and industry to contribute to the development of Australia’s big data capability.

The Australian data management framework, which has emerged over time, is based on four principles:

1. The institutional data management framework is in accordance with the Australian Code for the Responsible Conduct of Research and other external legal and regulatory frameworks.
2. The research institution will support all aspects of the data lifecycle, through creation and collection, storage, manipulation, sharing and collaboration, publishing, archiving and reuse.
3. Data management is an essential part of doing good research and supporting the research community of which each researcher is a part.

²³ The Concordat on Open Research Data includes a broader coalition of UK funders and university stakeholders.

4. Effective data management is best achieved through teamwork and collaboration between researchers, research offices, information specialists and technical support staff.

While the principles were originally drafted to outline how responsibilities between research institutions and researchers should be divided, it is now clear that increasing the availability of open scientific data is a collective endeavour. At the same time, accountability for the preparation and curation of such data must be clearly assigned. It is for this reason that research funders, providers and researchers themselves are likely to remain the key stakeholders in this process. The Australian Code for the Responsible Conduct of Research (revised in 2007) remains the principal document guiding Australian research organisations and researchers in data management. The code states:

Each institution must have a policy on the retention of materials and research data. It is important that institutions acknowledge their continuing role in the management of research material and data [205].

The Australian Research Council (ARC) and National Health and Medical Research Council (NHMRC)—two principal funders of national research—mandated open access to peer-reviewed publications in 2012. Starting from 2014, the ARC requires data publication for selected grants. The ARC Centre of Excellence funding agreement:

... strongly encourages ... the depositing of data and any publications arising from a Project in an appropriate subject and/or institutional repository [206].

The NHMRC mandate did not extend to open data until early 2018. These very recent developments are covered in Chapter 8, Section 8.2.

3.2.6 Canada

The principal funders of research in Canada—the Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council and the Social Sciences and Humanities Research Council—all adhere to open access practices in research. Following a long consultation process, the final version of their Tri-Agency Open Access Policy was released in March 2015. With regard to open data, several submissions suggested that all three agencies should practice long-term preservation and digital release. Yet only the CIHR has committed to a policy on open research data at this stage:

Recipients of CIHR funding are required to adhere with the following responsibilities:

1. Deposit bioinformatics and atomic and molecular coordinate data into the appropriate public database (e.g. gene sequences deposited in GenBank) immediately upon publication of research results.
2. Retain original datasets for a minimum of 5 years after the end of the grant (or longer if other policies apply). This applies to all data, whether published or not. The grant recipient's institution and research ethics board may have additional policies and practices regarding the preservation, retention and protection of research data that must be respected.²⁴

²⁴ Article 3.2 of Ref. [207].

This policy applies to all CIHR grants awarded from 1 January 2008 and onwards. An important aspect that the data deposit is required (not just encouraged) for all CIHR grants.

3.3 Selected policies of publishers

Meanwhile, publishers are also having a profound influence, with changes to how they provide scholarly communications. Journal publication is the primary mode of disseminating scientific research. However, recent years have seen the emergence of data journals and of open access data repositories for holding the data associated with journal articles.

The best-known example of the latter is perhaps the Dryad Digital Repository [208], governed by a consortium of scientific members who collaboratively promote data archiving, free access, reusability and citation. Membership of Dryad is open to any stakeholder organisation—including journals, scientific societies, publishers, research institutions and libraries. Dryad initially covered biosciences and ecology studies and, in recent years, has expanded to other disciplines. Many libraries and research organisations now refer to Dryad as a generic data repository and recommend it for deposit in all instances where discipline-specific online repositories do not exist.

As a result of these practices, Dryad is increasingly becoming an interdisciplinary resource covering data from a variety of scientific fields and international sources. Data repositories such as Dryad can provide quicker access to findings in advance of results published in paper journals or e-journals.

The growing significance of data publications has prompted established journals to expand their offerings. In early 2014 the Nature Publishing Group announced a new peer-reviewed open data publication, *Scientific Data*. The journal introduces data descriptors—a combination of traditional content and structured data and information to be curated in-house. Such descriptors may include articles and data from multiple journals. The actual datasets will not be stored in-house but in a recognised discipline data repository²⁵ or, in the absence of such repository, in a more generic data repository such as Dryad. The initial focus of *Scientific Data* is on biomedical, life and environmental sciences—subject matter that appears to overlap with the initial collecting priorities at Dryad. It will be interesting to see how Dryad and *Scientific Data* differentiate themselves and develop into the future.

Another important driver of open research data is the changing policy among traditional journal publishers who increasingly require that underlying data be made available to both peer reviewers and readers. In many cases, the publishers also specify the requirements for sufficient data description so as to facilitate reuse and validation of the research findings. For instance, the policy of *Journal of the Royal Society Interface* states:

To allow others to verify and build on the work published in Royal Society journals it is a condition of publication that authors make available the data and research materials supporting the results in the article. Datasets should be deposited in an appropriate, recognised repository and the associated accession number, link, or digital object identifier (DOI) to the datasets must be included in the methods section of the article. Reference(s) to datasets should also be included in the reference list of the article with DOIs (where available). Where no discipline-specific data repository exists authors should deposit their datasets in a general repository such as Dryad [210].

Similarly, the journal *Nature* has a policy on the availability of data and materials that implies that the data should be described sufficiently to allow for validation and reuse:

²⁵ *Nature* lists publicly-recognised data repositories on its website [209].

*An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. Therefore, a condition of publication in a Nature journal is that authors are required to make materials, data and associated protocols promptly available to readers without undue qualifications.*²⁶

Importantly, *Nature* reserves the right to refuse publication to authors who fail to comply with the journal's requirements on data availability.

This open data policy is far more specific and stringent than similar policies introduced by other publishers. An authoritative study by Vasilevsky et al. published in 2017 evaluated the open data policies of 318 biomedical journals [211]. That investigation found that only 12% of these journals required data sharing as a condition of publication—a policy similar to that of *Nature*.²⁷ Out of the journals surveyed, 23% explicitly encouraged or addressed data sharing, but did not require it as a condition of publication, while 9% required data sharing but made no explicit statement regarding the effect on publication. Additionally, 15% only addressed data sharing for specific subsets of genomic data. Sadly, 32% of all journals did not mention anything about data sharing.²⁸ The study confirmed earlier findings by the same authors that fewer than 50% of journals require data sharing [212].

However, in 2017–2018 many publishers introduced changes to their editorial policies that provide for greater transparency and openness, including statements on expectations for data sharing. Publishers typically choose one of the following approaches for implementing data transparency:

- **Option 1: Duty to disclose.** Published articles must state whether the data upon which they are based is available and must provide information on how to access it. The wording of publisher policies typically includes “sharing upon reasonable request” or “expects data sharing”.
- **Option 2: Mandate to deposit.** Authors of articles must include in a trusted repository for the underlying data for sharing. If any portion of such data cannot be shared, this must be clearly identified while the authors must provide as much of the remaining data as can be reasonably shared. This type of policy typically focuses on creating “open data” or even “open FAIR data”.
- **Option 3: Verification of reproducibility.** Open data must be verified by a third party to establish whether the data enable the replication of findings as represented in the article. This type of data is typically referred to as “peer-reviewed data”.

The introduction of 2017–2018 policies by leading publishers, such as Springer Nature, Elsevier, Taylor and Francis, and Wiley, has appeared to have increased the expectations not only for the digital availability of research data, but also for the credibility and veracity of that data.

3.4 Issues covered in the open data policies

Reflecting on the above analysis of the emergent principles and policies in this chapter, it becomes clear that open scientific data extends open access to scientific

²⁶ *Nature* at point 65.

²⁷ *Ibid*, 18.

²⁸ *Ibid*.

publications. Several issues need consideration when developing policies for open research data, specifically:

1. The 'data' that should be covered by the policy
2. The timeframe for releasing research data into the public domain and who is responsible for the data deposit
3. The period for storing the data in digital archives
4. Whether research data management policies should be required and, if so, whether they should be submitted at the grant proposal stage or later
5. How open access to research data should be provided and under what conditions
6. Whether to recommend specific data repositories or whether to leave the decision with the project participants
7. When data sharing may not be required and whether the reasons for not sharing should be known to the broader research community
8. Whether and how data deposits should be embedded in the rewards and recognition frameworks for researchers and their organisations
9. Whether compliance with the policies should be monitored and, if so, whether penalties should apply
10. How to foster an environment that enables researchers and the public to maximise the value of research data
11. How to encourage the sharing of the best practices and experiences with research data management, including data transparency, code transparency, design and citation standards, and replication policies.

While the above points represent an ideal open data policy, the current policies of research funders and journal publishers are highly fragmented, covering only selected aspects of the data preservation, sharing and reuse process. This gap leaves those aiming to implement open data in a position of experimentation. The gap also makes any comparative analyses difficult. Nevertheless, it is apparent that the open data mandates have created a momentum driving the release of research data in many parts of the world.

At the same time, the policies are more like high-level statements of principles and expectations rather than detailed guidelines for researchers. One particular concern is the unclear meaning of research data in the policies. At best, the list of possible research data outputs included in the policies is incomplete and lacks a level of detail. At worst, the definitions of 'data' provided do not appear to match the notions of data commonly used by the key stakeholders across different scientific disciplines. The inability to clearly acknowledge and articulate the heterogeneous nature of research data is a major shortcoming of the open data mandates.

The above overview of policy statements supporting access to scientific data shows that all major players in the system have shown a commitment to open data. The policies also illustrate, however, that concerns about implementing open scientific data remain and require further attention. And while policies may state clearly what challenges exist, the solutions and best practices are only just starting to emerge.

Nevertheless, open scientific content is increasingly becoming readily available, largely due to policies recently introduced by research funders and publishers.

3.5 Open scientific data in emerging and developing countries

Elsewhere in Europe and Asia, open scientific data practice is already in place, and it is emerging in many Latin American countries. Yet these policies are not readily available in English and therefore are not analysed in this chapter. The awareness of open access has increased rapidly in recent years, with countries including China introducing open access mandates.

3.5.1 China

Chinese research output has increased rapidly—from 48,000 articles in 2003, or 5.6% of the global total, to more than 186,000 articles in 2012, or 13.9% [213]. Of those, more than 100,000, or 55.2% of the global share, involved some funding from the National Natural Science Foundation (NNSF) of China, one of the country's major basic science funding agencies. This administered the equivalent of US\$3.1 billion in its 2014 budget.²⁹ The research output from the Chinese Academy of Sciences (CAS)—which funds and conducts research at more than 100 institutions—is also impressive. CAS scientists published more than 18,000 Science Citation Index³⁰ articles in 2012 and more than 12,000 articles in Chinese journals [214].

On 15 May 2014, these two principal funders of research in China announced an open access policy for publications. Researchers supported by NNSF or CAS should deposit their papers into online repositories and make them publicly accessible within 12 months of publication. The policies are modelled around those introduced by the NIH in the United States and came into effect the same day they were announced.³¹ At this point the open access mandate does not appear to extend to scientific data.

Both CAS and NNSF plan to release more detailed guidelines on implementation. In particular, the NSFC will establish a repository into which researchers can upload papers. This repository is likely to be modelled on PubMed Central developed by the NIH.³² CAS started developing a network of repositories for its institutes 5 years ago and has a central website [215] for searching them. As of December 2013, more than 400,000 articles had been deposited and had generated 14 million downloads.³³

3.5.2 Central and Eastern Europe

Many countries in Central and Eastern Europe have well-developed digital infrastructures, and several countries have increased their R&D expenditure in recent years. Estonia and Slovenia now spend more on R&D than the European Union average. The Czech Republic has reached a level that is close to the average, while Hungary, Lithuania, Latvia, Slovakia, and Romania spend significantly less than the average.³⁴ While these countries do not appear at this stage to have formulated open access policies, the digital agenda promoted by the European Union and the

²⁹ *Ibid.*

³⁰ Science Citation Index is a bibliometric tool offered by Thomson Reuters. The index provides citation information for articles included in the Web of Science database.

³¹ Xialing Zhang at 72.

³² *Ibid.*

³³ Xialing Zhang at 72.

³⁴ R&D expenditure measured as a percentage of Gross Domestic Product. See the OECD Science, Technology and Innovation Indicators [216]; and for non-OECD countries Eurostat [217].

conditions already embedded in European grants are likely to drive the digital sharing of research outcomes originating from these countries in the near future.

3.5.3 African countries

In large parts of Africa, scientific education remains underdeveloped, and funding for science is lacking. At the same time, many African countries have, in recent years, adopted important open access and open government projects and also have committed significant resources to develop relevant infrastructures. The vision for the open access movement in Africa is to spur development and promote the transfer of technologies to the continent.

Kenya recently announced the establishment of a pilot regional data-sharing centre at the Jomo Kenyatta University. The centre aims to accelerate the generation, analysis, management, and archiving of scientific data emanating from Africa. Other significant open data programmes are implemented in Kenya [218], Morocco [219], Tunisia [220], Tanzania [221], Sierra Leone [222], Nigeria [223, 224], and Ghana [225]. In addition, the African Development Bank sponsors the Open Data for Africa Initiative [226] that aims to enhance the statistical capacity of African countries as well as provide the tools necessary to monitor developments, such as progress with implementing the Millennium Development Goals.

It will be interesting to see how open scientific data will be used in innovative ways to promote development across Africa.

Conclusion

The early stages of implementing data stewardship in open science are promising. Key players in the system—research funders, governments, and leading publishers—have made a clear commitment to open scientific data and have developed policies governing it. Such policies are now in place in the developed world and Latin America and are starting to emerge in other countries.

These policies have created a momentum for data curation and are driving the release and sharing of research data globally. Data journals and discipline-specific data repositories have emerged and are becoming more popular. Scientists are increasingly aware of the need to share data and are more readily prepared to work with librarians to develop and implement research data management policies.

Yet challenges remain. The policies for open scientific data explicitly list limitations to data release. This appears to have sent mixed messages to research organisations. Instead of focusing their efforts on finding opportunities for data sharing, many have diverted their resources to ensuring compliance with existing limitations.

In the long term, this stage can be necessary to identify best practices for responsible research data management. In the short term, however, this stage may have delayed data release for other purposes, with major concerns surrounding research data management, particularly the interface between intellectual property and open knowledge, and the sharing of data involving personal information of subjects involved in data collection.

A major shortcoming of the open data policies is the high-level statements of objectives and expectations. They provide little guidance to researchers regarding the preparation of data management plans, the curating, and the sharing of data. One particular concern is the unclear meaning of research data, which leaves many researchers guessing what ‘data’ they need to make available.

These concerns are examined further in the next chapter, which discusses the meaning of open scientific data.

IntechOpen

IntechOpen

Author details

Vera J. Lipton
Zvi Meitar Institute for Legal Implications of Emerging Technologies,
Harry Radzyner Law School, IDC Herzliya, Israel

*Address all correspondence to: vera.lipton@bigpond.com

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. Distributed under the terms of the Creative Commons Attribution - NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited. 