OECD publishing

Please cite this paper as:

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

OECD Science, Technology and Industry Policy Papers No. 25

Making Open Science a Reality

OECD

OECD

MAKING OPEN SCIENCE A REALITY

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

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

This report was declassified by written procedure by the Committee for Scientific and Technological Policy on 20 August 2015.

© OECD/OCDE 2015

Applications for permission to reproduce or translate all or part of this material should be made to: OECD Publications, 2 rue André-Pascal, 75775 Paris, Cedex 16, France; e-mail: rights@oecd.org

Photo credit: Cover © Getty Images International.

TABLE OF CONTENTS

FOREWORD....5
GLOSSARY....7
EXECUTIVE SUMMARY....9
Introduction....9
The rationale for open science....10
Key actors in open science....12
Policy trends in open science....13
Main findings and policy messages....14
REFERENCES....17
CHAPTER ONE THE RATIONALES AND THE IMPACTS OF OPEN SCIENCE: AN OVERVIEW....18
Accessing scientific publications....20
Accessing data....26
"Altmetrics", an alternative way to measure scientific impact....28
REFERENCES....31
CHAPTER TWO OPEN ACCESS TO SCIENTIFIC PUBLICATIONS....36
Defining open access....36
Open access publishing and IP protection....41
Open access publishing and its legal implications....48
REFERENCES....51
CHAPTER THREE OPEN RESEARCH DATA....52
Data-driven scientific research....52
Defining open data....55
Data sharing: challenges and opportunities....57
Data protection frameworks in OECD countries....62

Unsolved legal issues: public-private partnerships and text and data mining.....65

REFERENCES.....68

CHAPTER FOUR THE GOVERNANCE OF OPEN SCIENCE: ACTORS, TRENDS AND POLICIES .71

The key actors.....71

Open science and citizen involvement.....84

Governance of open science: Recent policy trends 86

REFERENCES.....97

BIBLIOGRAPHY 99

FOREWORD

Science is the mother of the digital age. And yet, twenty-two years after CERN placed the World Wide Web software in the public domain, effectively creating the open internet, science itself has struggled not only to "go digital" but also to "go open". This report, *Making open science a reality* reviews the progress in OECD countries in making the results of publicly funded research, namely scientific publications and research data openly accessible to researchers and innovators alike. The report i) reviews the policy rationale behind open science and open data; ii) discusses and presents evidence on the impacts of policies to promote open science and open data; iii) explores the legal barriers and solutions to greater access to research data; iv) provides a description of the key actors involved in open science and their roles; and finally v) assesses progress in OECD and selected non-member countries based on a survey of recent policy trends.

The project was carried out as a part of the activities of the OECD's Working Party on Innovation and Technology Policy (TIP) of the Committee for Scientific and Technology Policy (CSTP). It has been prepared jointly by the OECD Secretariat (Giulia Ajmone Marsan and Mario Cervantes, Directorate for Science, Technology and Innovation) and members of the TIP steering group on Open Science: Alexandre Bourque-Viens (Canada), Päivi Rauste and Pirjo-Leena Forsström (Finland), Wojtek Sylwestrzak, Lukasz Bolikowski and Krzysztof Siewicz (Poland), Dirk Meissner (Russian Federation), Fernando Mérida Martín (Spain), Nick Seaford and Micheal Reda (United Kingdom), and Jerry Sheehan (United States). Lucie Guibault and Thomas Margoni (University of Amsterdam) have prepared a background paper to this report, containing detailed analysis of the legal aspects of open science and open data; this has been used in drafting the sections on the legal aspects of open science in this report. Barbara Ubaldi (OECD Secretariat, Directorate for Public Governance and Territorial Development), Fernando Galindo-Rueda, Brunella Boselli, Claire Jolly and Brigitte Van Beuzekom (OECD Secretariat, Directorate for Science, Technology and Innovation) provided additional input. Salvatore Mele, Vasco Vaz, Bo-Christer Björk and Mikael Laasko provided comments and data. Dominique Guellec, Head of the OECD Science and Technology Policy Division provided overall guidance and comments. Katjusha Boffa prepared this report for publication. In addition to the above-mentioned authors, who also provided the country notes relative to their countries, additional country notes were prepared by:

- Eric Laureys (Belgium)
- Patricia Muñoz and Paula González Frías (Chile)
- Viktor Muuli (Estonia)
- Mark Asch, Alain Colas, Marie-Pascale Lizée, Laure Menetrier, Justin Quemener, Romain Tales and Frédérique Sachwald (France)
- The Federal Ministry of Education and Research (Germany)
- Evi Sachini (Greece)
- Usha Munshi and Devika Madalli (India)
- Claudio Artusio, Juan Carlos De Martin, Federico Cinquepalmi and Giulietta Iorio (Italy)

- Kazuhiro Hayashi (Japan)
- Jeong Hyop Lee and Seokjong Lim (Korea)
- Margarita Ontiveros (Mexico)
- Rene Daane, Marjan van Meerloo and Dries van Loenen (the Netherlands)
- Rune Rambæk Schjølberg and Hanne Monclair (Norway)
- Luisa Henriques and Vasco Vaz (Portugal)
- The Ministry of Science, Industry and Technology (Turkey)
- Jean-Claude Burgelman, Rene Von Schomberg and Ana Nieto (the European Commission)
- Neil Thakur (United States)

The report has also benefited from the inputs and comments of participants in the TIP Open Science Workshops organised in Paris (December 2013), Warsaw (March 2014) and Helsinki (November 2014). This report also takes into account comments received from TIP and CSTP Delegates by written procedure, and the discussion at the CSTP meeting on 21-22 October 2014.

GLOSSARY

Open science **–** There is no formal definition of open science. In this report, the term refers to efforts by researchers, governments, research funding agencies or the scientific community itself to make the primary outputs of publicly funded research results – publications and the research data – publicly accessible in digital format with no or minimal restriction as a means for accelerating research; these efforts are in the interest of enhancing transparency and collaboration, and fostering innovation. The report focuses on three main aspects of open science: open access, open research data, and open collaboration enabled through ICTs. Other aspects of open science – post-publication peer review, open research notebooks, open access to research materials, open source software, citizen science, and research crowdfunding are also part of the architecture of an "open science system".

Open access **–** Unrestricted online access to scientific articles. Access can occur via a number of channels, such as institutional repositories, journal publishers' websites, researchers' webpages, etc.

Gratis open access **–** refers to scientific publications that are free of charge and technical restrictions but not necessarily free of legal restrictions.

Libre open access **–** access to scientific publications that are free of charge and of technical and (at least some) legal and permission restrictions.

Gold open access **–** Open access provided by a publisher. Under gold open access, generally the publishing costs and revenues are recovered through fees.

Green open access **–** The practice of self-archiving the pre-print or the post-print of an article, generally by its author. The costs of green open access are generally covered by institutional funding or a percentage of research grants.

Article processing charge (APC), also known as a publication fee, is a fee which is sometimes charged to authors in order to publish an article in a scholarly or academic journal.

Hybrid open access – Open access provided by subscription-based journals where some articles are available in open access, provided that APCs have been paid.

Public access – In Canada and the United States, increased access to the output of publicly funded research results in digital format is more commonly referred to as public rather than open access because the term open access is typically reserved for access obtained through gold open access. Public access can take a green or gold approach.

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

Research data – are factual records used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. They are collected and produced in a wide range of formats: digital spreadsheets and databases, compilations from surveys, images, or objects. The consultation and usage of research data often involves use of specific computer programmes, software, etc.

Metadata – are detailed descriptions of the data sets and documentation of the workflow needed to access these resources; they are often necessary for the usage of the data itself.

Open government data (OGD) – Open Government Data (OGD) refers to government or public sector data (i.e. any "raw" data produced or commissioned by public sector) made available through open access regimes so that it can be freely used, re-used and distributed by anyone.

EXECUTIVE SUMMARY

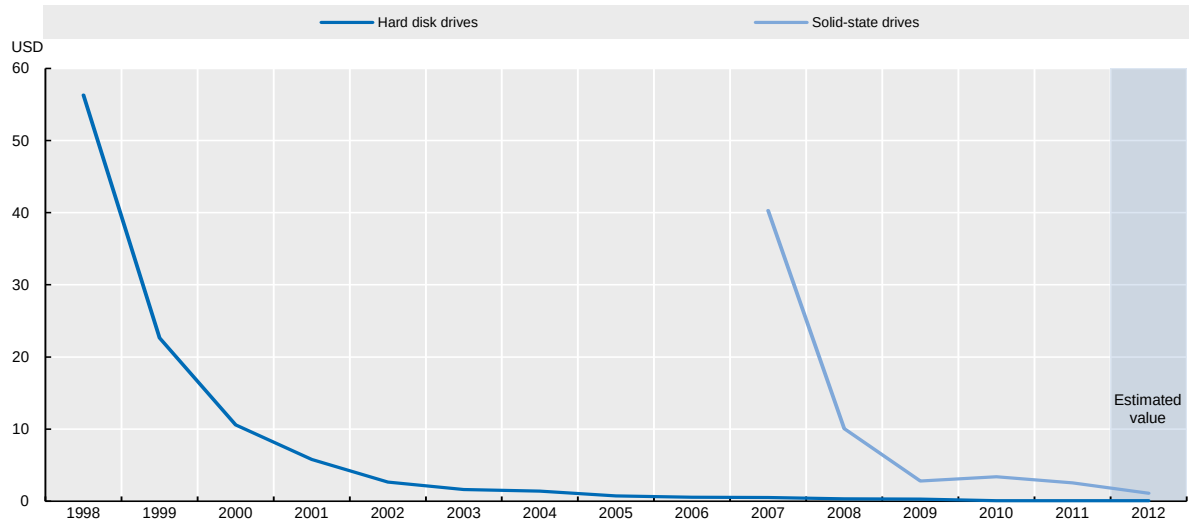
Introduction

Open science commonly refers to efforts to make the output of publicly funded research more widely accessible in digital format to the scientific community, the business sector, or society more generally. Open science is the encounter between the age-old tradition of openness in science and the tools of information and communications technologies (ICTs) that have reshaped the scientific enterprise and require a critical look from policy makers seeking to promote long-term research as well as innovation.

On the one hand, the Internet and online platforms are creating new opportunities to organise and publish the content of research projects, scientific publications and large data sets, so as to make them immediately available to other scientists and researchers as well as potential users in the business community and society in general. On the other hand, ICTs allow the collection of large amounts of data and information that can be the basis of scientific experiments and research, contributing to make science increasingly data-driven (Figure ES.1). Online repositories and archives offer the possibility to store, access, use and reuse research and scientific inputs and outputs (both articles and data sets), and speed the transfer of knowledge among researchers and across scientific fields, opening up new ways of collaborating and new research methods (Force11, 2012). This evolution of science into a more open and data-driven enterprise is often referred to as "open science".

Figure ES.1 Average data storage cost for consumers

1998-2012, per Gbit



Source: OECD (2014), *Measuring the Digital Economy: A New Perspective*, OECD Publishing, Paris.

The term "open science" was coined by economist Paul David (2003) in an attempt to describe the properties of scientific goods generated by the public sector and in opposition to the perceived extension of intellectual property rights into the area of information goods. Economists consider scientific knowledge generated by public research as a public good, which means that everyone can make use of that knowledge at no additional cost once it is made public, generating higher social returns. This thinking is not altogether new. As far back as 1942**,** Robert King Merton, an American sociologist of science, described a set of ideals that characterised modern science and to which scientists are bound. First and foremost is the notion of common ownership of scientific discoveries, according to which the substantive findings of science are seen as a product of social collaboration and are assigned to the community. Scientists' claims to intellectual property are limited to recognition and esteem.[1] The race to be the first to claim recognition – the so-called priority rule – in science has traditionally been a strong incentive for scientists to make their knowledge public.

While this ideal-based system has functioned in part through the current system of peer review and subscription-based scholarly publication, the ICT revolution has shaken, if not the underlying ideals, at least the system of scientific production and diffusion. Open science in the information age espouses the notion that knowledge created from public research has public good characteristics that go beyond the concept of the "commons" developed in the 18th century, insofar as ICT-enabled access broadens the possibilities to enrich the commons and extend it to a broader range of users.

In recent years, open science has become an active area of policy development, both within the OECD area and beyond. Although recognising that open science is a broad concept that encompasses more than open access to research data and publications that takes place at all stages of research (see Glossary), this report aims to provide an analytical overview of recent open science policy trends, by focusing in particular on those initiatives to promote broad access to publicly funded research results, including both scientific publications and research data.

The rationale for open science

The particularities of open science provide the policy and economic rationales for supporting it. Open search tools increase the efficiency of research as well as of its diffusion. Greater access to scientific inputs and outputs can improve the effectiveness and productivity of the scientific and research system, by: reducing duplication costs in collecting, creating, transferring and reusing data and scientific material; allowing more research from the same data; and multiplying opportunities for domestic and global participation in the research process. Scientific advice can also benefit from the greater scrutiny offered by open science, as it allows a more accurate verification of research results. In addition, increased access to research results (in the forms of both publications and data) can foster spillovers not only to scientific systems but also innovation systems more broadly (Box 1.1). With increased access to publications and data, firms and individuals may use and reuse scientific outputs to produce new products and services. Open science also allows the closer involvement and participation of citizens.

[1] See Merton, R.K. (1973), *The Sociology of Science: Theoretical and Empirical Investigations*, University of Chicago Press.

There is growing evidence that open science has an impact on the research enterprise, business and innovation, and society more generally. Recent analysis reveals that enhanced public access to scientific publications and research data increases the visibility of, and spillovers arising from, science and research.

There has been debate in the academic literature as to whether open access publications receive more citations than non-open access publications, which has led to attempting to measure the so-called *open access citation advantage*. Most of the studies conducted on this question do find that open access increases citations. It has also been argued that the open access citation advantage is caused by a quality bias (i.e. researchers tend to publish via open access their best-quality works, and this is why they get more citations); however, there is also evidence that the citation advantage is not caused by the quality bias but by the advantage from users self-selecting what to use and cite, without any constraint related to selective accessibility to subscribers only.

Scientists and academics are not the only groups that can benefit from greater open science efforts. The demand from the business sector and individual citizens to access research results is significant. For example, usage data from PubMedCentral (the online repository of the US National Institutes of Health) show that 25% of the daily unique users are from universities, 17% from companies, 40% are individual citizens and the rest are from government or in other categories (UNESCO, 2012).

Calculating estimates of the economic value of research publications and data is challenging, but these have begun to emerge. Available estimates include those of Houghton and Sheehan (2009), who analyse the effects of increasing accessibility to public sector research outputs in Australia; they conclude that increased accessibility generates a return of approximately AUD 9 billion over 20 years. Houghton, Rasmussen and Sheehan (2010) estimated that a public access policy mandate for US federal research agencies over a transitional period of 30 years may be worth around USD 1.6 billion and up to USD 1.75 billion if no embargo period is in place. Around USD 1 billion would benefit the US economy directly and the remaining amount would translate in economic spillovers to other countries. These figures would be significantly higher than the estimated cost of implementing open access archiving. JISC (2014) conducted a study on the economic impact of three UK data centres (the Economic and Social Data Service, the Archaeology Data Service and the British Atmospheric Data Centre), and estimated that the returns to investment of each of these three centres could be between approximately twofold and tenfold over 30 years.

Additional evidence shows that firms and smaller research institutions face barriers to accessing public research results. A recent study on R&D-intensive small and medium-sized enterprises (SMEs) in Denmark (Houghton, Swan and Brown, 2011) found that 48% of those SMEs consider research outcomes very important for their business activities, and more than two-thirds reported difficulties in accessing research material. A survey on UK SMEs found evidence that the equivalent of 10% to 20% of articles were not easily accessible for the survey respondents (Ware, 2009). Finally, it has been argued that making research data publicly available may promote public understanding of science, evidence-based practices, and citizen-science initiatives (Kowalczyk and Shankar, 2010).

Key actors in open science

Several actors in local, national and global innovation systems are involved in open science efforts:

- *Researchers themselves* have been at the forefront of efforts to promote open science. There are several motivations for researchers, ranging from the cultural values inherent in science (e.g. openness to scrutiny, willingness to engage society) to necessity, i.e., developing a technological infrastructure to allow for collaboration. Researchers also respond to incentives from funding agencies, universities and public research institutes. Tension may nevertheless exist between the competitive "publish or perish" paradigm and the interest in sharing data and collaborating.
- *Government ministries* have developed national strategies for open science, either as stand-alone strategic efforts or as part of broader open government agendas. These agendas help define national- level strategic priorities that can be translated into concrete initiatives by other innovation system actors.
- *Research funding agencies* are key actors in the promotion of open science efforts, as they are responsible for defining the mechanisms and requirements to benefit from grants and funding for research. In many countries in recent years, funding agencies have increasingly adopted rules and mechanisms to promote open science and in some cases mandate it, by including open or public access of funded research outputs as a requirement. In addition to mandatory requirements, funding agencies may promote open science through financial support to cover open access publishing charges or costs associated with the release of data and other research material.
- In a majority of OECD countries, *universities and public research institutes* have some degree of autonomy and are responsible for drawing up their own policies to support open science and implementing the policies of funding councils or agencies. In addition, universities and higher education institutions may play a role in training students and researchers to develop the skills necessary to enable open science practices – from basic skills related to the use of online repositories, to the ones needed to implement data cleaning, curation and management.
- *Libraries, repositories and data centres* are key actors for and fundamental enablers of open science. Libraries have adapted their role and are now active in the preservation, curation, publication and dissemination of digital scientific materials, in the form of publications, data and other research-related content. Libraries and repositories constitute the physical infrastructure that allows scientists to share use and reuse the outcome of their work, and they have been essential in the creation of the open science movement.
- *Private non-profit organisations and foundations* may play a significant role in developing, raising awareness and encouraging an open science culture. They may not only fund open research and introduce requirements in grant agreements, but also develop and facilitate the creation of networks of stakeholders worldwide.
- *Private scientific publishers* offer a broad range of open access publishing (for example via the gold route or publishing in hybrid journals) and related services such as the maintenance of digital repositories and data sets or other scientific material, or the development of text and data mining (TDM**)** tools.
- *Businesses* constitute part of the demand for open access publications and data that they use to develop new products and services. Businesses such as pharmaceutical firms also enable open science through public-private partnerships with universities or their financing of open clinical trials, for example.

Finally, *supra-national entities* play a major role in the definition of international co-ordination agreements or guidelines to address open science issues with an international and global perspective. Inter-governmental organisations (IGOs) play a critical role in promoting inter-governmental co-ordination at international level and in shaping the political agenda, through developing guidelines and principles around specific themes that are subsequently adopted and implemented by member countries and beyond. IGOs such as the OECD, UNESCO, the EU and the World Bank have been active in recent years in promoting open science efforts of member and (in some cases) non-member countries.

Policy trends in open science

Following the 2007 OECD principles and guidelines on access to public research data, OECD countries have made efforts to adapt legal frameworks and implement policy initiatives to encourage greater openness in science (OECD, 2015). At the level of research institutions, implementing measures and policies may take the form of mandatory rules on access to scientific publications or data, incentives for open access publishing, or funding for infrastructure. The measures are thus of three kinds: sticks (mandatory rules), carrots (incentives), and enablers (soft and hard infrastructure).

- *Mandatory rules* are often implemented in the form of requirements in research grant agreements, or in some cases are defined in national strategies or institutional policy frameworks.
- *Incentive mechanisms* may take the form of financial support to cover open access publishing costs or the release of data sets. They may also be in the form of proper acknowledgment of open science efforts of researchers and academics, for instance data set citations or career advancement mechanisms partly based on metrics that take into account open science or data-sharing efforts.
- *Enablers* include for example infrastructure developed to share articles or data; initiatives undertaken to develop an open science culture; amendments to the legal framework to make them increasingly open science-friendly; or development of the skills necessary for researchers to share and reuse the research outputs produced by others. Data management guidelines for universities and public research institutes also can constitute an enabling condition.

Measures belonging to the three types of actions may be implemented together to promote open access, by means of integrated and multi-faceted approaches. Recent policy trends, however, have revealed that the majority of initiatives implemented so far involve mandatory rules for open science and development of the infrastructure to enable open science.

As regards incentives, research funding agencies and governments often provide funding to cover the costs of open access publishing. In contrast, reward mechanisms for researchers involved in open access and open data activities are less common. Reward mechanisms that are currently under discussion include widespread use of data set citation and/or proper acknowledgment of open science and data-sharing efforts in career advancement mechanisms, or grant attribution to research teams.

Legal frameworks that explicitly accommodate open science (i.e., that are open science-friendly) are an additional means of promoting open science. For example, in Germany the national copyright act was modified in 2013 to allow publicly funded scientists and researchers to retain the legal right to upload their publications on line, even if they have transferred their exploitation rights to the publishers, after an embargo period of up to 12 months. The United Kingdom has recently passed a series of amendments to its legal framework for copyright (that came into force in 2014); these include greater freedom of reuse of

copied or recorded material for education and non-commercial research purposes. Australia and Finland are also considering modifications of the existing legal framework around the publication of publicly funded research results, to make the copyright legislation increasingly open science-friendly.

Ultimately however, the key to making open science a reality will be to ensure that the social contract between scientists is strengthened and not weakened. Governments and public research institutions must ensure that open science policies, especially when it comes to open data, allow scientists to continue to compete and be recognised for their contributions if they are to be incentivised to share access to scientific data and results.

Main findings and policy messages

Open science is a means and not an ends. Open science strategies and policies are a means to support better quality science, increased collaboration, and engagement between research and society that can lead to higher social and economic impacts of public research.

Open science is more than open access to publications or data; it includes many aspects and stages of research processes. Although this report focuses primarily on open access to publications and research data, it is important to remember that open science is a broader concept that also includes the interoperability of scientific infrastructure, open and shared research methodologies (such as open applications and informatics code), and machine-friendly tools allowing, for example, text and data mining.

Policies to promote open data are less mature than those to promote open access to scientific publications. While the principle of open access to scientific data is well established in OECD countries, the scope of access varies greatly across countries. This is due to the fact that data sets are not as easily identified and defined as scholarly research articles. Diversity of scientific data and differing traditions and standards in their treatment are also issues. Some of the additional challenges related to data sets include the definition of ownership of large-scale data sets, potentially collected by machines or software providers; privacy; confidentiality; and even national security issues. In addition, certain classes of data, such as medical records, are particularly sensitive due to privacy issues.

Open science policies should be principle-based but adapted to local realities. Open science policies require a diversity of approaches, taking into account the needs of the different actors involved in research projects. For example, if a research project involves business sector partners and commercial interests are present, the requirements for sharing research results may be different from the case in which only public actors are involved. In other cases, privacy or confidentiality concerns may apply to the treatment of certain classes of individual data.

Better incentive mechanisms to promote data-sharing practices among researchers are needed. While all public sector researchers have an interest in maximising the sharing of published research articles, the same is not true for research data sets, especially at the pre-publication stage. In addition, data cleaning and curation (for example, by developing metadata) is a time-consuming activity that is rarely acknowledged in evaluation mechanisms or grant allocation procedures. Most evaluations of universities and researchers are almost entirely based on teaching and bibliometric indicators, attributing little value to the sharing of pre-

publication inputs and post-publication outcomes, such as data and other relevant information. Extending citation mechanisms to data sets can partly address this issue.

Data-related skill development is essential. Researchers' skills needed for sharing articles or data sets openly on line are unevenly distributed. Some disciplines such as computer science or physics may have a longer tradition of uploading research material on repositories and curating and maintaining large data sets. Researchers in other disciplines, however, may need to be trained to develop the necessary skills to make open science happen. At the same time, students and citizens need to acquire the skills to take advantage of, use and reuse data set shared by the research community. Some countries are currently developing data science curricula to address this issue.

Training of and awareness-raising among researchers is important for the development of an open science culture. Recent surveys on the behaviour of scientists reveal that not all researchers are necessarily aware of the possibilities offered by open science. In some countries, different institutions regularly organise workshops and training sessions to make researchers aware of these possibilities.

Repositories and online platforms will not have impact if the information they contain is not of good quality. If repositories are not user-friendly and the data sets they contain have not been properly cleaned and curated, or the metadata have not been sufficiently developed, it may be difficult to maximise their usage.

The long-term preservation costs of openly available research output need to be considered. Open access is not without costs. Many governments and research institutions are currently bearing the costs of offering open access to articles and to data, as well as the costs of storage and the preservation of data sets on line. Given the rapidly increasing amounts of data, public institutions will be challenged to find sustainable funding and business models. Public-private partnerships with private service providers may offer innovative solutions.

Clear legal frameworks for the sharing of publications and reuse of data sets are needed at the national and international levels. A lack of clarity on the interpretation of national and international legal frameworks may prevent the sharing or reuse of research results. In addition, clear guidelines around text and data mining are needed, as this tool will become increasingly used by researchers in the future. Some OECD countries are currently discussing or have recently modified national legal frameworks to make them increasingly open science-friendly.

Consultative approaches that involve all relevant actors for open science are a key component of successful open science strategies. Open science efforts involve different communities and different actors: researchers, governmental institutions, universities and research centres, libraries and data centres, private non-profit organisations, business sector organisations including private academic publishers, supra-national entities, citizens, etc. These actors do not necessarily have the same incentives, goals or expectations. A successful strategy needs to take into account this diversity, and react accordingly.

International collaboration in the area of open science is necessary to address global challenges. International collaboration is becoming more important than ever, as publications and data in electronic form travel across national frontiers. Shared and interoperable infrastructure is necessary to disseminate research results and promote scientific collaboration. Such efforts can help avoid the duplication of effort,

as well as helping share the risks or the associated investments. In addition, BRIC countries Brazil, China and India are also adopting open science policies and data infrastructure roadmaps. International co-ordination and co-operation in this area will become even more important as the global production of knowledge and R&D increasingly shifts towards the emerging economies. Furthermore, tackling global challenges will require greater access to and sharing of national public research data sets – and consequently, greater co-operation at a global level.

Policy makers need to promote openness in science while at the same time preserving competition. Competition is a key aspect of the scientific enterprise: pushing for open access and open data too early may be counterproductive in some cases.

REFERENCES

David, P.A. (2003), "The economic logic of "open science" and the balance between private property rights and the public domain in scientific data and information: A primer", in P. Uhlir and J. Esanu (eds.), *National Research Council on the Role of the Public Domain in Science*, National Academy Press, Washington, DC.

Force11 (2012), "Improving the future of research communications and e-scholarship", Force11 white paper, www.force11.org/white_paper.

Houghton, J. and P. Sheehan (2009), "Estimating the potential impacts of open access to research findings", *Economic Analysis and Policy*, Vol. 29, No. 1, pp. 127-42.

Houghton, J., B. Rasmussen and P. Sheehan (2010), "Economic and social returns on investment in open archiving publicly funded research outputs", Report to the Scholarly Publishing and Academic Resources Coalition (SPARC), Center for Strategic Economic Studies, Victoria University.

Houghton, J., A. Swan and S. Brown (2011), "Access to research and technical information in Denmark", Technical Report, School of Electronics and Computer Science, University of Southampton.

JISC (2014), "The value and impact of data sharing and curation: A synthesis of three recent studies of UK research data centres", JISC, March, http://www.cni.org/news/jisc-report-value-impact-of-data-curation-and-sharing/, accessed 11 June 2015.

Kowalczyk, S. and K. Shankar (2010), "Data sharing in the sciences", *Annual Review of Information Science and Technology*, No. 45, pp. 247-94.

Merton, R.K. (1973), *The Sociology of Science: Theoretical and Empirical Investigations*, University of Chicago Press.

OECD (2015), *Inquiries into Intellectual Property's Economic Impact*, OECD Publishing, Paris.

OECD (2014), *Measuring the Digital Economy: A New Perspective*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264221796-en.

UNESCO (2012), *Policy Guidelines for the Development and Promotion of Open Access*, UNESCO Publishing.

Ware, M. (2009), *Access by UK Small and Medium-sized Enterprises to Professional and Academic Literature*, Publishing Research Consortium, Bristol.

Chapter One

THE RATIONALES AND THE IMPACTS OF OPEN SCIENCE: AN OVERVIEW

The various rationales for policies on open science and open data (Box 1.1) naturally imply that the criteria for assessing the impacts are equally multiple. On the one hand, greater access to scientific inputs and outputs may improve the effectiveness and the productivity of the scientific and research system by reducing duplication costs in collecting, creating, transferring and reusing data and scientific material; allowing more research from the same data; and multiplying opportunities for domestic and global participation in the research process.

On the other hand, increased access to research results (in the forms of both publications and data) can not only foster spillovers to scientific systems, but also boost innovation systems more broadly. With unrestricted access to publications and data, firms and individuals may use and reuse scientific outputs to produce new products and services (see Boxes 1.2 and 4.9).

There is evidence that both research and innovation system actors may experience difficulties in accessing scientific material that is not openly available. Several surveys report access difficulties in the academic community in the United States and Europe. For instance, according to the Committee for Economic Development, 15% of US and Canadian scholars from all disciplines reported their level of access not to be satisfactory (CED, 2012). Ware and Monkman (2008) found that only 66% of scientists in Europe and the Middle East reported having good or excellent access (85% in the United States). And the numbers outside those regions are even lower. Barriers to access to scientific material for researchers due to the high cost of subscriptions are also reported by the surveys of Rowlands and Nicholas (2005) and Sparks (2005).

Box 1.1 Rationales for open science and open data for research and innovation

The following factors are often associated with openness in science and research:

- *Improving efficiency in science* – Open science efforts can increase the effectiveness and productivity of the research system, by 1) reducing duplication and the costs of creating, transferring and reusing data; 2) allowing more research from the same data; 3) multiplying opportunities for domestic and global participation in the research process.
- *Increasing transparency and quality* in the research validation process, by allowing a greater extent of replication and validation of scientific results.
- *Speeding the transfer of knowledge* – Open science can reduce delays in the re-use of the results of scientific research including articles and data sets and promote a swifter path from research to innovation.
- *Increasing knowledge spillovers to the economy* – Increasing access to the results of publicly funded research can foster spillovers and boost innovation across the economy as well as increase awareness and conscious choices among consumers.
- *Addressing global challenges more effectively* – Global challenges require co-ordinated international actions.

Open science and open data approaches can promote collaborative efforts and faster knowledge transfer for a better understanding of challenges such as climate change or the ageing population, and could help identify solutions.

- *Promoting citizens' engagement* in science and research – Open science and open data initiatives may promote awareness and trust in science among citizens. In some cases, greater citizen engagement may lead to active participation in scientific experiments and data collection.

Source: OECD (2013a), Background paper for the TIP workshop on Open Science and Open Data, unpublished, DSTI/STP/TIP(2013)13.

Developing countries especially may benefit from open access (OA) to scientific material. Chan, Kirsop and Arunachalam (2005) note that according to a World Health Organization survey, in countries with an annual GNP per capita of less than USD 1 000, around 56% of medical institutions have no subscriptions to journals; in countries with GNP per capita of between USD 1 000 and USD 3 000, the percentage of medical institutions with no subscriptions was lower, but still as high as 34%. This is why initiatives to provide developing countries with access to scientific material have been developed. For example, the Research4Life programme is a public-private partnership among three United Nations agencies, two universities and major commercial publishers that enables eligible libraries and their users to access peer-reviewed international scientific journals, books and databases for free or for a small fee (Royal Society, 2012). In some disciplines, open access journals have been created directly in developing countries, such as the African Journal of Health Sciences.

Scientists and academics are not the only groups that can benefit from greater open science efforts. Demand from the business sector and individual citizens for access research results in the form of data and publications is significant. For example, usage data from PubMedCentral show that 25% of the daily unique users are from universities, 17% from companies, 40% are individual citizens, and the rest are from government or in other categories (UNESCO, 2012). A recent study of R&D-intensive SMEs in Denmark (Houghton, Swan and Brown, 2011) found that 48% of those SMEs consider research outcomes very important for their business activities, and more than two-thirds reported difficulties in accessing research material. Ware (2009) conducted a survey on UK small and medium-sized enterprises and found evidence that the equivalent of 10% to 20% of articles were not easily accessible for his survey respondents. Finally, it has been argued that making research data publicly available may promote public understanding of science, evidence-based practices and citizen-science initiatives (Kowalczyk and Shankar, 2010).

Box 1.2 The opportunities arising from text and data mining (TDM)

Text and data mining (TDM) refers to an ensemble of computer science techniques to analyse and extract knowledge and information from large digital data sets (i.e. big data), by looking for trends and patterns unnoticeable to human eyes. TDM is increasingly used by researchers in all fields, from historians who scan historical documents and archives, to medical experts who find common patterns in medical records. TDM is also a well-established technique in fields such as astronomy and genetics. TDM methods and techniques are widely used both in the public and private sectors. TDM algorithms investigate large-scale data sets containing not only figures and numbers but also other types of digital records, including text, images and audio files. TDM enables the use of common techniques, makes connections between unconnected fields of research, and represents a major opportunity for the development of innovation.

Its use has important repercussions in the academic community. With the growing amount of published (and unpublished) academic articles (an estimated 50 million as of 2010), it is becoming impossible for scientists and researchers to manually access, read and analyse publications. TDM provides the potential for accessing, scanning

and analysing publications by means of machines. The publishing industry is developing services to make scientific journal databases increasingly interoperable and to standardise terminology in order to make it easier for researchers to apply TDM techniques (Clark, 2013).

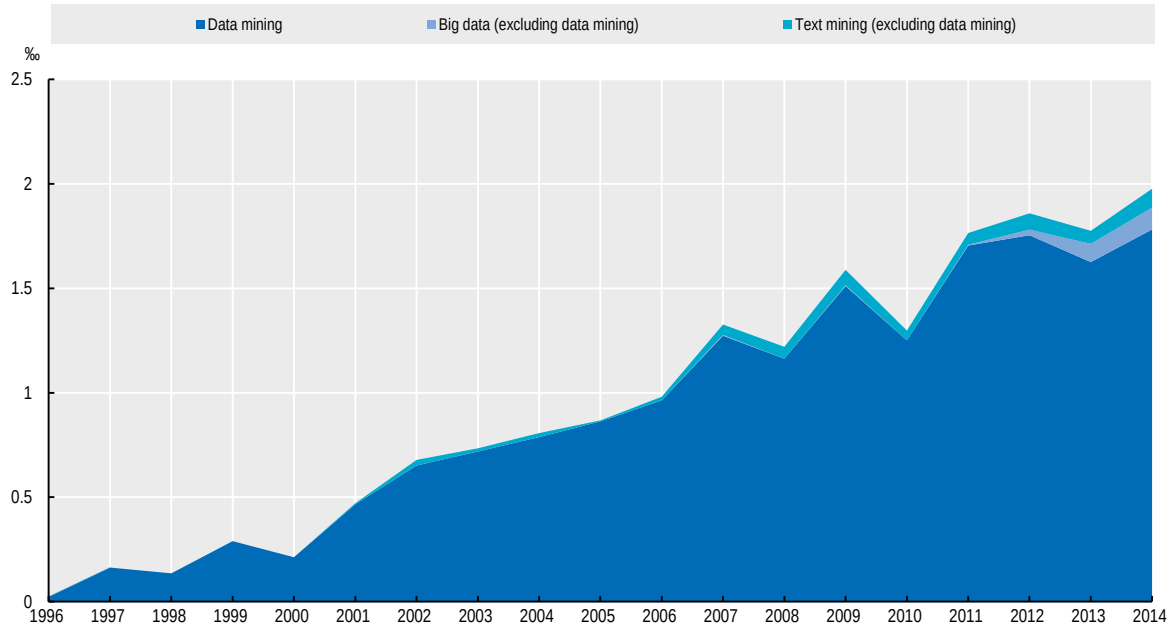
Research on these techniques has advanced considerably in recent years (Figure 1.1). The number of academic articles published on the subject of TDM since the beginning of the 1990s reveals that the United States has so far produced 46.6% of the publications dealing with TDM, followed by the United Kingdom (11.1%), Taiwan (8.8%), Canada (5.7%) and China (4.6%).

Whether current copyright regimes are promoting or hindering TDM is an open question. According to a recent JISC report on the value and benefit of text mining (JISC, 2012), licensing agreements represent a key barrier to the use of text mining techniques in the higher education and research communities in the United Kingdom. Recent OECD analysis has highlighted how the context in which IP frameworks operate has been changing substantially. In this evolving context, the way copyright laws address TDM is not always clear in all jurisdictions (OECD, 2015a). According to the same report, there is some (disputed) evidence that researchers in certain jurisdictions (such as the European Union and Brazil) are inhibited from engaging in TDM due to fears of infringing copyright in the process.

Sources: Clark, J. (2013), *Text Mining and Scholarly Publishing*, Publishing Research Consortium; European Commission (2014), "Standardisation in the area of innovation and technological development, notably in the field of text and data mining", Report from the Expert Group; Filippov, S. (2014), "Mapping tech and data mining in academic and research communities in Europe", Lisbon Council, 16/2014; OECD (2015a), *Inquiries into Intellectual Property's Economic Impact*, OECD Publishing, Paris; JISC (2012), *The Value and Benefits of Text Mining*, JISC, www.jisc.ac.uk/sites/default/files/value-text-mining.pdf.

Figure 1.1 TDM-related scientific articles

1995-2014, per thousand article



Source: OECD (2014), *Measuring the Digital Economy: A New Perspective,* OECD Publishing, Paris.

Accessing scientific publications

Recent analysis shows that over the past decade, open access articles have steadily increased their relative share of all scholarly journal articles. Different estimates of open access uptake are available depending on the different definitions of open access and the samples used in the analysis, as well as the time at which the analysis was conducted. Open access uptake may also depend on different open access

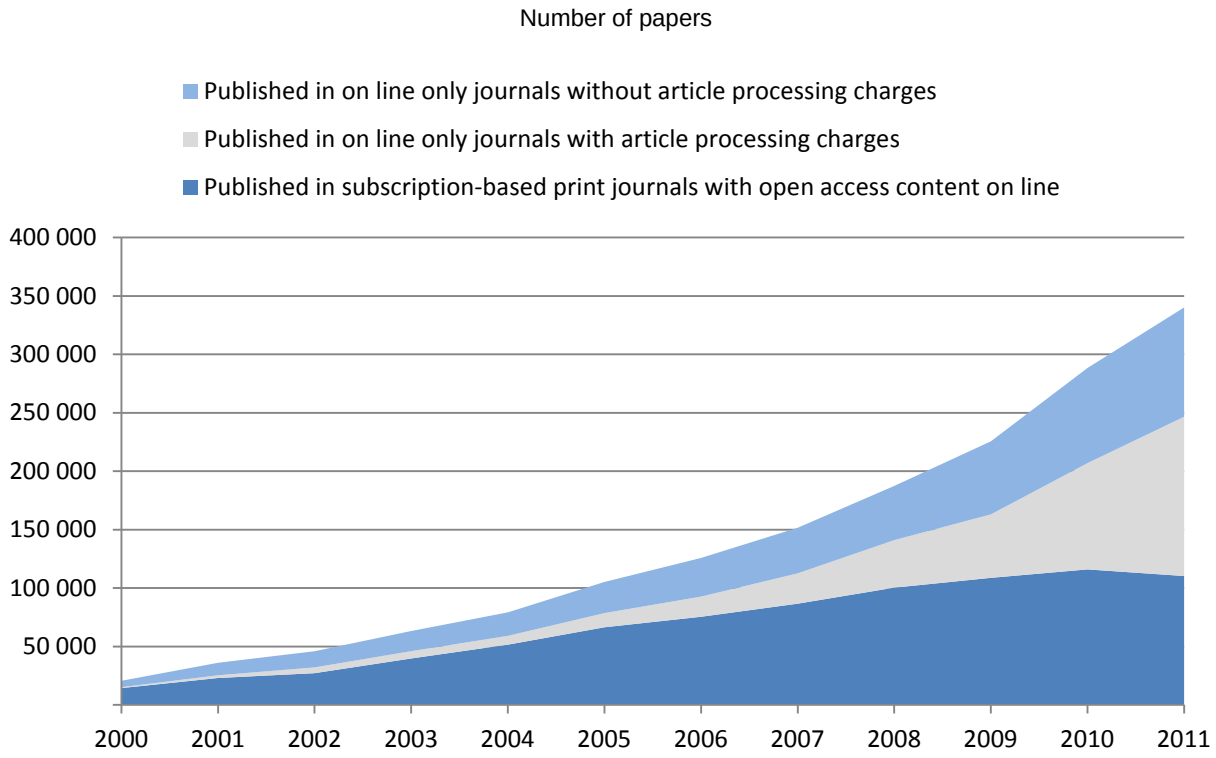
paths (Figure 1.2). A recent study by Archambault et al. (2014) found that as of the beginning of 2014, over 50% of the scientific papers published between 2007 and 2012 can be accessed and downloaded for free on line.

According to Laakso and Björk (2012), about 17% of scientific articles published in 2011 and indexed in Scopus (the most comprehensive article-level index of scientific articles) were available through journal publishers (i.e. gold open access). Most articles were immediately available (12%) whereas the remaining 5% were available 12 month after publication. Hybrid open access articles accounted for 0.7% of the total published articles in 2011. Open access articles involving APCs accounted for 49% of all gold open access articles (Laakso and Björk, 2012). Preliminary evidence seems to suggest that article processing charges do not strongly correlate with journal impact (Björk and Solomon, 2014), especially in the case of hybrid open access (Romeu et al., 2014).

Estimating green open access uptake is more complicated, as researchers archive articles not only on official repositories but also on personal webpages or on other digital infrastructure. Several estimates have been developed: more conservative estimates, such as those by Björk et al. (2013), suggest that the share of green open access articles accounted for approximately 12% of all recently published peer reviewed literature, at the time they conducted the analysis. Other estimates (Gargouri et al., 2012) put that figure at slightly above 20% as of 2011.

Lewis (2012) suggests that gold open access (i.e. when the authors publish in scientific journals openly available on line, commonly referred to as open access journals) could account for 50% of the scholarly journal articles between 2012 and 2017 and 90% of all articles between 2020 and 2025. However, Miguel, Chichilla-Rodrígues and de Moya-Anegón (2011) show that the percentage of green road journals greatly surpasses the percentage of gold road publications. In addition, green open access (i.e. when the author self-archives the article in an online repository) was recently argued to be the most effective and affordable means for funders, institutions and other stakeholders (Houghton and Swan, 2013).

Figure 1.2 Gold open access uptake

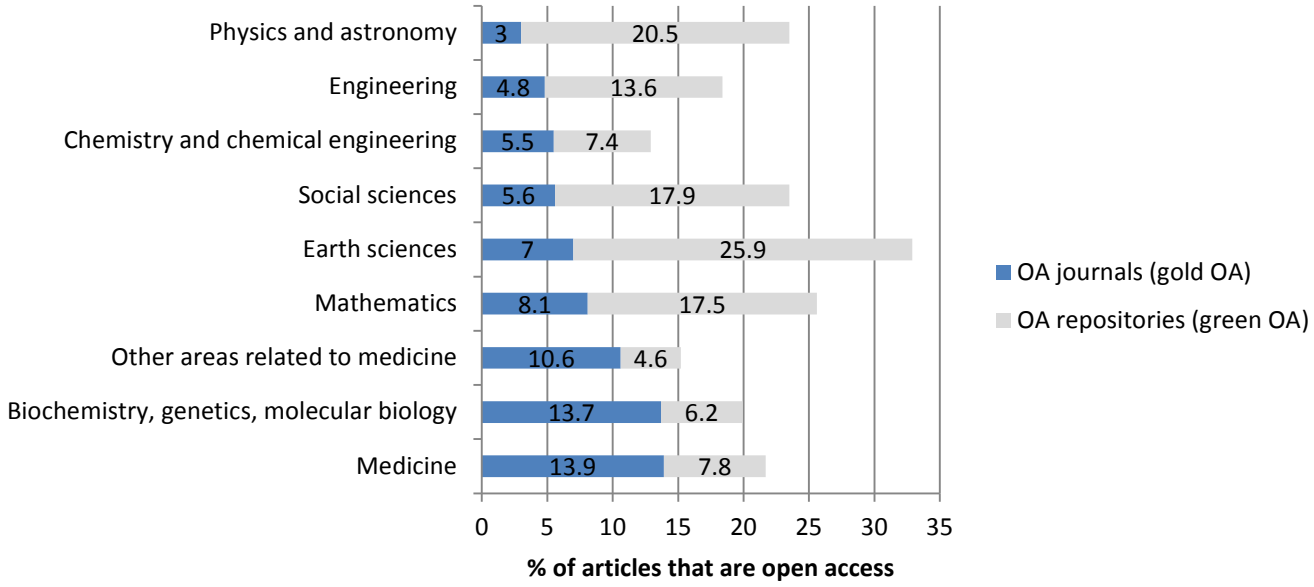


Source: Laakso, M. and B.-C. Björk (2012), "Anatomy of open access publishing: A study of longitudinal development and internal structure", *BMC Medicine,* 10, p. 124.

The amount of material that is publicly accessible varies considerably from discipline to discipline, often due to different cultures of sharing in different domains (UNESCO, 2012, Figure 1.3). For example, open science behaviour in the field of high-energy physics (HEP) dates back decades, to when scholars sent pre-prints (manuscripts of their publications that had not yet appeared in peer-reviewed journals) to their peers around the world (Goldshmidt-Clermont, 2002**;** Heuer, Holtkamps and Mele, 2008; Aymar, 2009).

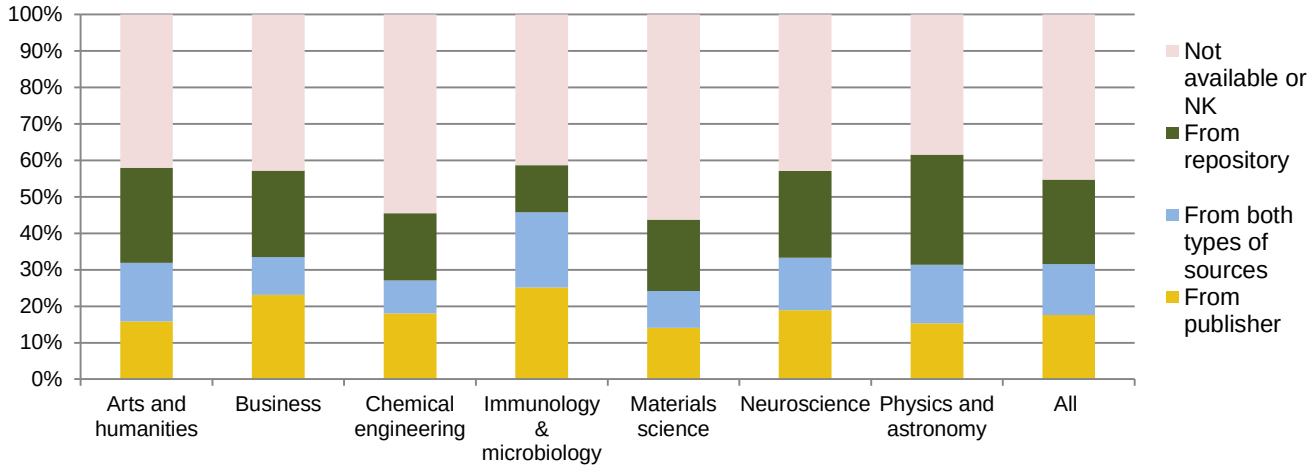
Björk et al. (2010) and UNESCO (2012) indicate that green open access is characterised by substantial variation by discipline. According to these studies, the uptake of green open access is higher in physics and astronomy; earth and environmental sciences; mathematics; and social sciences, arts and humanities, than in medicine, chemistry or biology and genetics (Figure 1.3a). They also estimate that two of the most well-established repositories – PubMedCentral (PMC) for biomedicine and arXiv (see Box 4.5) for physics and mathematics – together contain 38% of all green copies and 94% of all copies in subject repositories. As is reasonable to expect, PMC dominates in the life sciences and arXiv in physics and mathematics. Moreover, according to Laakso (2014), there is considerable variation in the length of embargo periods in different disciplines (Figure 1.4). A recent OECD survey shows that open access has progressed but that variation by disciplines remains significant (Figure 1.3b).

Figure 1.3a Open access varies by discipline



Source: UNESCO (2012), *Policy Guidelines for the Development and Promotion of Open Access*, UNESCO Publishing, and Björk et al. (2010), "Open Access to the scientific journal literature: Situation 2009", *PloS ONE*, Vol. 5, No. 6.

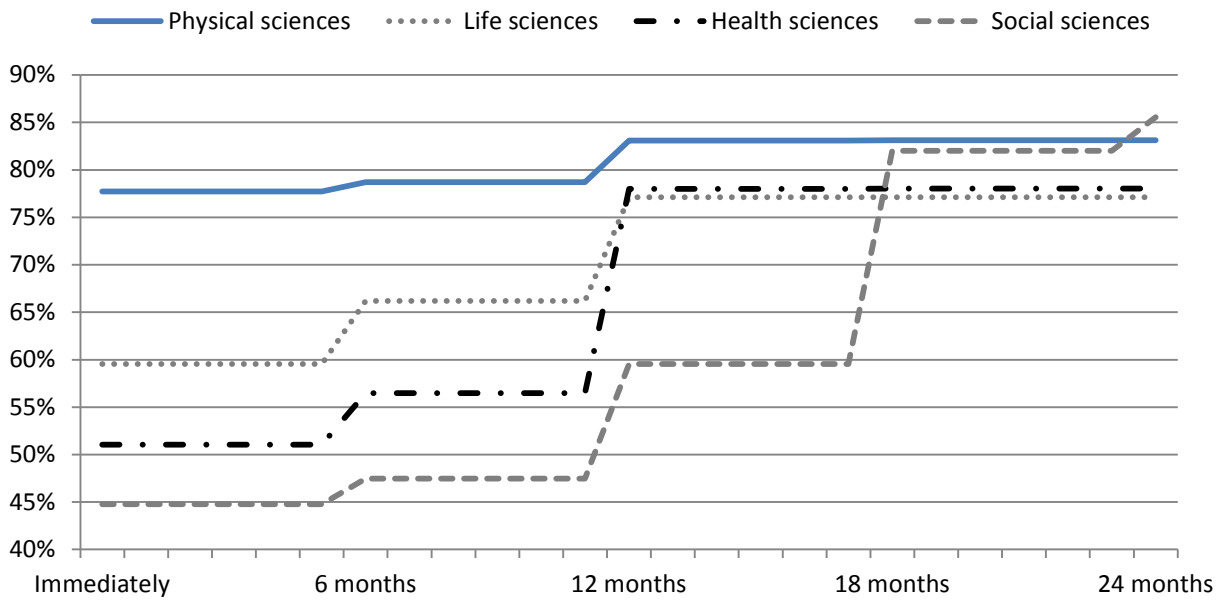
Figure 1.3b



Source: Preliminary analysis of OECD NESTI Pilot Survey of Scientific Authors 2014-15. Note: NK = not known.

Figure 1.4 Embargo length of different disciplines

% of available articles over time



Source: Laakso, M. (2014), Green open access policies of scholarly journal publishers: A study of what, when, and where self-archiving is allowed, *Scientometrics*, Vol. 99, pp. 475-494.

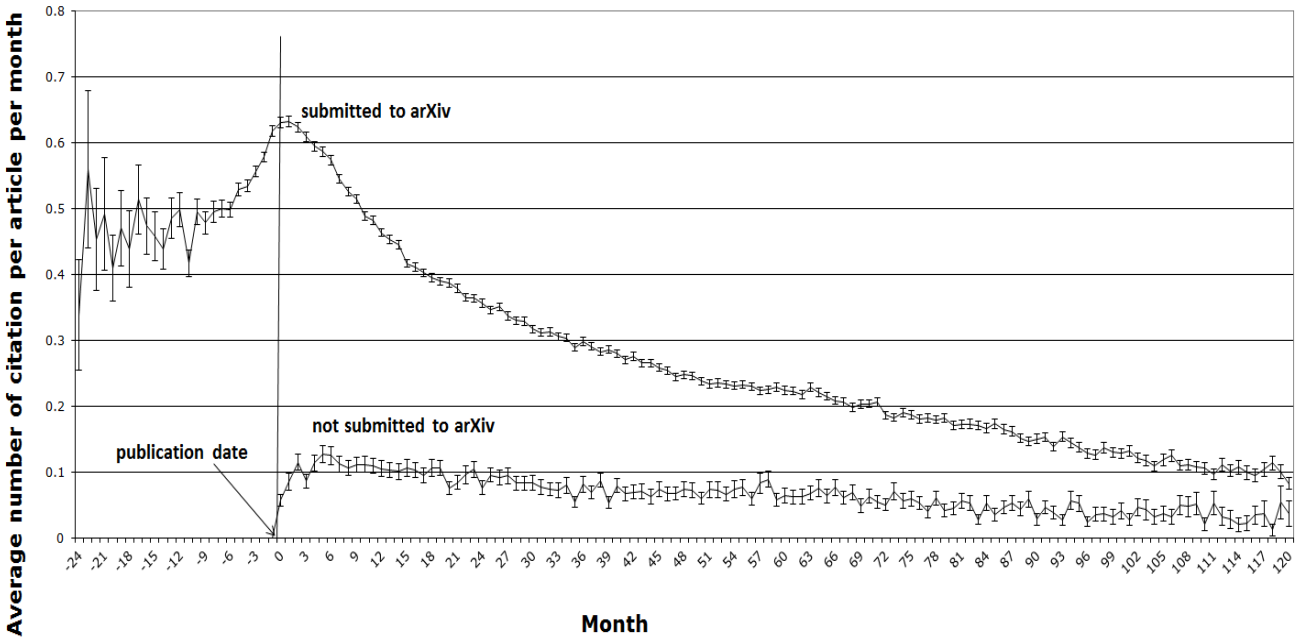
There has been much debate in academic literature as to whether open access publications receive more citations than non-open access publications, which has led to attempts to measure the so-called *open access citation advantage*. Most of the studies conducted on this topic tend to demonstrate that open access increases citation impact. However, there is no general consensus on the intensity of that increase (UNESCO, 2012; Swan, 2010; Wagner, 2010; OpCit Project, 2012; McCabe and Snyder, 2014). Although a minority, some other studies did not show any citation advantage (see for example Davis et al., 2008; Fradsen, 2009; Lansingh and Carter, 2009).

It has also been argued that the open access citation advantage is caused by a quality bias: researchers tend to publish via open access the best quality works, and this is why they get more citations. However, Gargouri et al. (2010) find evidence that the citation advantage is not caused by the quality bias but by the quality advantage from users self-selecting what to use and cite, without any constraint related to selective accessibility for subscribers only.

Gentil-Beccot, Mele and Brooks (2009) found that free and immediate online dissemination of pre-prints creates a considerable citation advantage in high-energy physics (HEP). This is also caused by the fact that in the field of HEP, researchers have the tendency to cite pre-print versions of the paper before publication (Figure 1.5). In addition, the analysis of Internet clickstreams in the leading digital libraries reveals that high-energy physics scientists have the tendency to prefer to download articles from repositories rather than journal websites (HEP scientists are between four and eight times more likely to download an article from arXiv rather than its final published version on a journal website).

Figure 1.5 The open access citation advantage in the high-energy physics

Analysis based on articles published in the Journal of High energy Physics and Physical Review D



Source: Gentil-Beccot, A., S. Mele and T.C. Brooks (2009), Citing and Reading Behaviors in High-Energy Physics: How a Community Stopped Worrying about Journals and Learned to Love Repositories.

Since the US National Institutes of Health (NIH) implemented its mandatory public access policy for publications resulting from its funded research, there has been a dramatic increase in the number of articles available on PubMed Central. As of late 2014, PMC contained more than 3.2 million full-text articles in biomedicine and related fields. The number of retrieved articles had doubled between 2011 and 2014; in 2011, approximately 500 000 unique visitors were accessing PMC on a typical weekday, downloading 1 million articles (CED, 2012). By 2014 the number of unique visitors was approaching 1 million per day, and the number of retrieved articles surpassing 2 million. Hardisty and Haaga (2008) conducted research on medical articles and found that open access articles were read twice as often by mental health practitioners. In addition, reading the open access article was associated with the practitioner recommending a more cutting-edge treatment.

The SOAP project (Study of Open Access Publishing) has conducted a large-scale survey of the attitudes of researchers on open access publishing. According to a summary of the results (Dallmeier-Tiessen et al., 2011), researchers surveyed showed a positive attitude towards open access publishing; the share was even higher in the humanities (90% positive responses) than in hard sciences (around 80%). The benefit for the scientific community as a whole was considered higher than for individual researchers. Most of the respondents identified funding as the main barrier to publishing via open access, followed by the quality of open access journals. Funding was perceived as the major barrier unevenly across disciplines: in the biological sciences, agriculture sciences and medicine-related sciences, funding barriers were perceived as being higher than in business and administration, astronomy and space, or social sciences. Therefore, according to the results of this project, open access policies are more likely to have effects in those disciplines where open access practices have not yet been the norm, rather than in

those fields where almost everything is already openly available on line. For more information on other recent surveys on open science and the behaviour of scientists, see Box 4.12.

The calculation of estimates for the economic value of research articles only and the related contribution to economic development is more problematic. Available estimates include that of Houghton and Sheehan (2009), who analyse the effects of increasing accessibility to public sector research outputs in Australia and estimate that increased accessibility generates a return of AUD 9 billion over 20 years. Houghton, Rasmussen and Sheehan (2010) estimated that a public access policy mandate for US federal research agencies over a transitional period of 30 years may be worth around USD 1.6 billion and up to USD 1.75 billion if no embargo period is in place. Around USD 1 billion would benefit the US economy directly, and the remaining amount would translate in economic spillovers to other countries. According to the authors, these figures would be significantly higher than the estimated cost of implementing open access archiving.

Accessing data

Sharing data has always been considered a crucial activity for scientific research and widely accepted by the scientific community[2] (Fienberg, Martin and Straf, 1985). There is some evidence that, as regards open access to scientific publication, sharing data can increase the citation rate of scientific papers (Piwowar, Day and Fridsma, 2007; Piwowar and Vision, 2013) and foster good scientific behaviour (Mooney, 2011). Sharing data allows the use and reuse of data from other researchers and individuals (Groves, 2009); it would also protect against faulty behaviours and fraud in science and research, and may contribute to improve data collection and management (Grieneisen and Zhang, 2012). For all these reasons, data sharing practices are often regarded positively by the research community (Cragin et al., 2010).

Data sharing not only allows verification of scientific results but also re-analysis of data for different purposes from the ones originally conceived. This not only enhances the utilisation of data, but also promotes competition of ideas and research (Gardner et al., 2003) and fosters collaboration (Brase et al., 2009; Piwowar and Chapman, 2008). For example, Murray et al. (2009) found that open access to research material raises the incentive for additional research by encouraging the establishment of new research directions.

Data sharing also reduces duplication of effort from different researchers attempting to collect the same data sets (Kowalczyk and Shankar 2010). Lakhani et al. (2007) found that the disclosure of problem information to a large group of outside solvers is an effective means of solving scientific problems. In addition, the disclosure of problem information facilitates problem-solving at the boundary of or outside their fields of expertise, thanks to the transfer of knowledge from one field to another. Williams (2010) used a natural experiment related to the effort to decode the human genome. She found that articles based on the analysis of openly available genome sequences led to 30% more articles than those based on the analysis of sequences protected by IPRs. The advantage in having publications and commercialisations generated by open sequences was notable. As for scientific publications, data sharing is especially important for researchers in developing countries who have fewer possibilities to undertake expensive and time-consuming data collection efforts (Arzberger et al., 2004).

[2] Many of the references listed in this section are derived from Costas et al., 2013.

In 2007, the US National Institutes of Health (NIH) introduced the Genome-Wide Association Studies (GWAS) policy and the associated Database of Genotypes and Phenotypes (dbGaP). As of December 2013, the NIH had received more than 17 500 requests for dbGaP data from 2 221 researchers, resulting in 924 publications. Twenty-five per cent of those publications appeared top-tier journals, including *PLoS One* and *Nature Genetics*. Research originated from dbGaP data has had a considerable impact on scientific discoveries: it enabled new discoveries in Alzheimer's and psychiatric-related research. Following these successful results, the 2014 NIH Genomic Data Sharing Policy will extend the GWAS policy to data collected in other types of genomics research (Paltoo et al. 2014).

Several studies have attempted to estimate the impact of greater access to data on the economy in general. A recent analysis of UK organisations (Royal Society, 2012; CEBR, 2012) estimated that data were worth approximately GBP 25 billion to UK private and public sector organisations in 2011. The estimate is the cumulative result of GBP 17.4 billion gained in business efficiency, GBP 2.8 billion derived by business innovation and GBP 4.8 billion gained from business creation.

In the United States, data released by the US National Weather Service are estimated to contribute to the development of the private sector meteorology market in an amount corresponding to approximately USD 1.5 billion (Spiegler, 2007). In 2008, the NASA Landsat satellite imagery of the Earth's surface environment became freely available on the Internet (see Box 1.3 for an overview of satellite data availability). The usage of this database increased from 19 000 scenes per year (when scenes were sold for USD 600 each) to 2.1 million scenes per year. Leading Silicon Valley companies such as Google (in particular Google Earth) use these images, and the open release is estimated to have generated direct benefits of more than USD 100 million per year to the US economy. According to a recent estimate of the US Open Data initiative (data.gov), open data has the potential to generate more than USD 3 trillion per year in additional value in sectors such as finance, consumer products, health, energy and education OECD (2015b).

An ongoing study of the US GovLab Academy at New York University, an online community that uses technology and innovation to solve public domain problems, is attempting to understand how US companies use open government data, through the Open Data 500 project. The project is analysing US-based companies (including international companies with a major presence in the United States) using open government data, a critical resource for their business. Most of the companies in the study belong to the technology, financial and business/legal service industries. According to the study, the most widely used data originates from the Department of Commerce, followed by the Department of Health and Human Services and the Securities and Exchange Commission.

The European Commission Open data initiatives are expected to generate a yearly income of EUR 140 billion (EC, 2012). In addition, OECD (2013b) estimated that the public sector information (PSI)-related market for the OECD area could be around USD 500 billion plus an additional USD 200 billion if barriers to use were removed, skills enhanced, and data infrastructure improved.

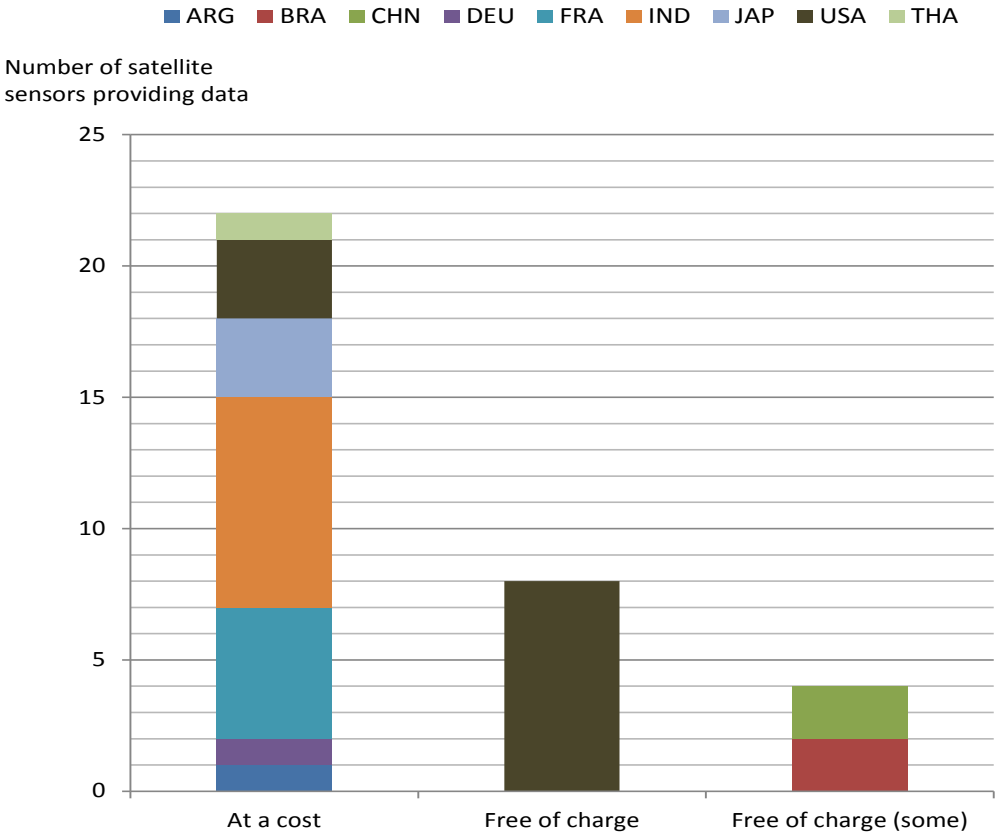
JISC (2014) conducted a study on the economic impact of three UK data centres (the Economic and Social Data Service, the Archaeology Data Centre and the British Atmospheric Data Centre), and estimated that the returns to investment of each of these three centres could be worth between approximately twofold and tenfold over 30 years.

Box 1.3 Satellite earth observation data policies

Satellite data have become indispensable for many scientific models and for numerous applications, including climate forecasting and dedicated products to support crop producers and urban planners few. However, data distribution policies for sensors carried on board satellites vary widely between countries. When examining data policies regulating 35 sensors used frequently in land surface imaging (carried on board 29 optical satellites from 10 different space-related agencies), 57% of the data coming from these sensors may be used at a cost (see figure immediately below). The costs of satellite imagery depend on several factors, including the level of details available and the freshness of the data.

Figure 1.6 Data policies for 35 satellite land surface imaging sensors

Data distribution policy of space-related organisations responsible for the satellite's sensors



Note: The space-related organisations included are: Centre National d'Etudes Spatiales (CNES), France; Comisión Nacional de Actividades Espaciales (CONAE), Argentina; Deutsches Zentrum für Luft- und Raumfahrt (DLR), Germany; Indian Space Research Organisation (ISRO), India; Japan Aerospace Exploration Agency (JAXA); National Aeronautics and Space Administration/US Geological Survey (NASA/USGS), United States; China Academy of Space Technology (CAST), China; Geo-Informatics and Space Technology Development Agency (GISTDA), Thailand; and Instituto Nacional de Pesquisas Espaciais (INPE), Brazil.

Source: OECD Space Forum, www.oe.cd/spaceforum.org

"Altmetrics", an alternative way to measure scientific impact

Traditionally, the quality of scientific articles has been assessed by academic peer-reviewers, and their impact has been measured via the prestige of the journal the article is published in and citation counts.

These metrics are generally used for evaluations of individual researchers, academic teams or institutions, and also to help compile international rankings of scientific impact.

Information and communication technologies, however, are accelerating the speed at which scientific results are diffused. The use of new online scholarly (and non-scholarly) tools to disseminate results offers the possibility to develop and employ new metrics to capture different types of impacts of scientific outputs.

These new or alternative metrics are often referred to as *altmetrics* (Priem et al., 2010). Proponents of alternative metrics consider that traditional peer-review fails to limit the volume of published research, and therefore to adequately filter and assess the quality of scientific output. They also argue that citation-counting measures are useful but not sufficient to determine the impact of research. For example, a general critique that is often made to basic citation counts is that they neglect the impact of research articles outside academia, and ignore the reason or the context of citations (Taraborelli, 2008; Neylon and Wu, 2009). For example, firm-level data analysis shows that start-ups in the green technologies sector owning patents containing more citations to non-patent literature (i.e. scientific publications) are more likely to be funded by venture capitalists (Criscuolo and Menon, forthcoming).

Alternative metrics, which go beyond citation counts, consider parameters such as the number of discussions around scientific papers on the press, scientific blogs, social networks such as Facebook or Academia.edu, micro-blogging like Twitter, user-edited reference such as Wikipedia, and social videos such as YouTube and Vimeo (Priem and Hemminger, 2010). Alternative metrics can take into account a variety of ways to share and disseminate scientific findings (Bornmann, 2014). These include, for example:

- the number of article viewers, as captured by PDF downloads
- the number of times a research article is saved (for example in scientific articles archiving tools such as CiteULike[3] or Mendeley[4])
- the number of discussions on line around a paper, for example on science blogs or in journals as well as on social networks or micro-blogging platforms
- the number of times an article is recommended in social media or editorials and press articles
- the number of times the article is cited by non-technical academic literature, such as Wikipedia.

Wouter and Costas (2012) identified four advantages that alternative metrics have, compared to traditional bibliometric indicators:

- *Broadness* – Alternative metrics capture impact beyond the technical scientific community.

[3] www.citeulike.org/.

[4] www.mendeley.com/.

- *Diversity* – Alternative metrics capture a diversity of impacts that science and research have on multiple communities.

- *Speed* – Alternative metrics (at least some) allow capture impact immediately after the research output has been released; academic citations instead often takes place years after a paper has been produced.

- *Openness* – It is in principle easy to obtain alternative metrics data.

Social media generally allow scientific authors to reach a wider audience than the academic community. In addition, blogs and social networks allow researchers to receive comments and participate in discussions both with other researchers and the non-technical audience. For example, Groth and Gurney (2010) analyse chemistry blogs on the science blog aggregator Researchblogging.org, by means of keyword and citation similarity maps. As could be expected, they find that scientific discourse on the web is more immediate and contextually relevant, and has a higher degree of non-technical discussion, than academic literature. Analysis of online clickstreams of scientific papers can reveal scientific or research relationships across fields (OECD, 2013c).

A recent analysis (Costas, Zahedi and Wouters, forthcoming) shows that altmetrics are available only for around 15% of the articles published since 2011, although the percentage of publications with altmetrics scores has been rapidly increasing. In 2012 for example, the share of publications with altmetrics records had increased to 20%. This confirms that altmetrics are reliable indicators only for recent scientific outputs. The study also finds that citations and altmetrics are positively but weakly correlated, thus supporting the claim that citation and altmetrics measure different types of scientific impact. In addition, altmetric activity varies according to scientific field (Zahedi, Costas and Wouters, 2013). Publications in the social sciences and humanities exhibit a higher share of altmetric records, whereas mathematics, computer science, engineering and natural sciences are the fields with the lowest recorded altmetric activity. This indicates that the usefulness and power of altmetrics may vary with the scientific field.

Although altmetrics have evident benefits, the academic community (for example Thelwall et al., 2013) acknowledges that many of these alternative metrics deserve further investigation to clearly assess which kind of impact they are measuring and revealing. In addition, a major weakness is the fact that they may be easily modified (for example via social media spam) in order to affect rankings. Methodologies to control these phenomena exist however, and validation techniques may be improved in the future.

REFERENCES

Archambault, E. et al. (2014), "Proportion of open access papers published in peer-review journals at the European and world levels – 1996-2013", European Commission study prepared by Science Metrix, http://science-metrix.com/files/science-metrix/publications/d_1.8_sm_ec_dg-rtd_proportion_oa_1996-2013_v11p.pdf.

Arzberger, P. et al. (2004), "Promoting access to public research data for scientific, economic and social development", *Data Science Journal*, Vol. 3, November, pp. 1777-78.

Aymar, R. (2009), "Scholarly communication in high-energy physics: Past, present and future innovations", *European Review*, No. 1, pp. 33-51.

Björk, B.-C. et al. (2013), "Anatomy of green open access", *Journal of the American Society for Information Science and Technology*, Vol. 65, No. 2, pp. 237-50.

Björk, B.-C. and D. Solomon (2014), "Article processing charges in OA journals – Relationship between price and quality", *Scientometrics*, Vol. 103, No. 2, pp. 373-85.

Björk, B.-C. et al. (2010), "Open access to the scientific journal literature: Situation 2009", *PLOS ONE*, Vol. 5, No. 6, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0011273, 23 June, accessed 10 June 2015.

Bornmann, L. (2014), "Measuring the broader impact of research: The potential of altmetrics", ResearchGate, www.researchgate.net/publication/263506984_Measuring_the_broader_impact_of_research_The_potential_of_altmetrics, accessed 10 June 2015.

Brase, J. et al. (2009), "Approach for a joint global registration agency for research data", *Information Services and Use*, Vol. 29, No. 1, pp. 13-27.

CEBR (2012), "Data equity: Unlocking the value of big data", Centre for Economic and Business Research, London, www.sas.com/offices/europe/uk/downloads/data-equity-cebr.pdf, accessed 10 June 2015.

CED (2012), *The Future of Taxpayer-Funded Research: Who Will Control Access to the Results?*, Committee for Economic Development, Washington, DC.

Chan, L., B. Kirsop and S. Arunachalam (2005), "Open access archiving: The fast track to building research capacity in developing countries", *SciDevNet*, November.

Clark, J. (2013), *Text Mining and Scholarly Publishing*, Publishing Research Consortium.

Costas, R., Z. Zahedi and P. Wouters (2014), "Do 'altmetrics' correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective", http://arxiv.org/abs/1401.4321, accessed 10 June 2015.

Cragin, M.H. et al. (2010), "Data sharing, small science and institutional repositories", *Philosophical Transactions A of the Royal Society*, Vol. 368, No. 1926, pp. 4023-38, http://rsta.royalsocietypublishing.org/content/368/1926/4023, accessed 10 June 2015.

Criscuolo, C. and C. Menon (forthcoming), "Do patents matter for venture capital investments? Evidence from the green sector", OECD Science, Technologi and Industry Working Paper Series.

Dallmeier-Tiessen, S. et al. (2011), "Highlights from the SOAP project survey: What scientists think about open access publishing", http://arxiv.org/abs/1101.5260, accessed 10 June 2015.

Davis, P.M. et al. (2008), "Open access publishing, article downloads, and citations: Randomised controlled trial", *BMJ*, 2008;337:a568, www.bmj.com/content/337/bmj.a568, accessed 10 June 2015.

EC (2010), "Riding the wave: How Europe can gain from the rising tide of scientific data", Final report by the High-level Expert Group on Scientific Data, European Commission, October, http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf, accessed 10 June 2015.

Fienberg, S.E., M.E. Martin and M.L. Straf (1985), *Sharing Research Data*, National Academy Press, Washington, DC.

Filippov, S. (2014), "Mapping tech and data mining in academic and research communities in Europe", *The Lisbon Council*, Issue 16/2014.

Fradsen, T.F. (2009), "The integration of open access journals in the scholarly communication system: Three science fields", *Information Processing and Management*, Vol. 45, No. 1, pp. 131-41.

Gardner, D. et al. (2003), "Towards effective and rewarding data sharing", *Neuroinformatics*, Vol. 1, No. 3, pp. 289-95.

Gargouri, Y. et al. (2012), "Green and gold open access percentages and growth, by discipline", Paper presented at the 17th International Conference on Science and Technology Indicators (STI), Montreal, Canada.

Gargouri, Y. et al. (2010), "Self-selected or mandated, open access increases citation impact for higher quality research", PLOS ONE, Vol. 5, No. 10, Art. No. e13636, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0013636, accessed 11 June 2015.

Gentil-Beccot, A., S. Mele and T.C. Brooks (2010), "Citing and reading behaviours in high-energy physics", *Scientometrics*, Vol. 84, Issue 2, August, http://link.springer.com/article/10.1007%2Fs11192-009-0111-1, accessed 11 June 2015.

Goldschmidt-Clermont, L. (2002), "Communication patterns in high-energy physics", *High-Energy Physics Libraries Webzine*, Issue 6, March.

Grieneisen, M.L. and M. Zhang (2012), "A comprehensive survey of retracted articles from the scholarly literature", PLOS ONE, Vol. 7, No. 10, Art. No. e44118.

Groth, P. and T. Gurney (2010), "Studying scientific discourse on the web using bibliometrics: A chemistry blogging case study", in *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*, April 2010.

Groves, T. (2010), "The wider concept of data sharing: View from the BMJ", *Biostatistics*, Vol. 11, No. 3, pp. 391-92.

Hardisty, D.J. and D.A.F. Haaga (2008), "Diffusion of treatment research: Does open access matter?", *Journal of Clinical Psychology*, Vol. 67, No. 7.

Heuer, R.D., A. Holtkamp and S. Mele (2008), "Innovation in scholarly communication: Vision and projects from high-energy physics", *Information Services and Use*, Vol. 28, No. 2, pp. 83-96.

Houghton, J. and P. Sheehan (2009), "Estimating the potential impacts of open access to research findings", *Economic Analysis and Policy*, Vol. 29, No. 1, pp. 127-42.

Houghton, J. and A. Swan (2013), "Planting the green seeds for a golden harvest: Comments and clarification on going for gold", *D-Lib Magazine*, Vol. 19, No. 1/2.

Houghton, J., B. Rasmussen and P. Sheehan (2010), "Economic and social returns on investment in open archiving publicly funded research outputs", Report to the Scholarly Publishing and Academic Resources Coalition (SPARC), Center for Strategic Economic Studies, Victoria University.

Houghton, J., A. Swan and S. Brown (2011), "Access to research and technical information in Denmark", Technical Report, School of Electronics and Computer Science, University of Southampton.

JISC (2014), "The value and impact of data sharing and curation: A synthesis of three recent studies of UK research data centres", JISC, March, http://www.cni.org/news/jisc-report-value-impact-of-data-curation-and-sharing/, accessed 11 June 2015.

JISC (2012), "The value and benefits of text mining", JISC, www.jisc.ac.uk/sites/default/files/value-text-mining.pdf, accessed 11 June 2015.

Kowalczyk, S. and Shankar K. (2010), "Data sharing in the sciences", *Annual Review of Information Science and Technology*, Vol. 45, pp. 247-94.

Laakso, M. (2014), "Green open access policies of scholarly journal publishers: A study of what, when, and where self-archiving is allowed", *Scientometrics*, Vol. 99, No. 2, pp. 475-94, http://dx.doi.org/10.1007/s11192-013-1205-3, accessed 10 June 2015.

Laakso, M. and B.-C. Björk (2012), "Anatomy of open access publishing: A study of longitudinal development and internal structure", *BMC Medicine*, Vol. 10, pp. 124, http://www.biomedcentral.com/1741-7015/10/124, accessed 11 June 2015.

Lakhani, K.R. et al. (2007), "The value of openness in scientific problem solving", HBS Working Paper No. 07-050, Harvard Business School, http://hbswk.hbs.edu/item/5612.html, accessed 11 June 2015.

Lansingh, V.C. and M.J. Carter (2009), "Does open access in ophthalmology affect how articles are subsequently cited in research", *Ophthalmology*, Vol. 116, No. 8, pp. 1425-31.

Lewis, D.W. (2012), "The inevitability of open access", *College and Research Libraries*, Vol. 73, No. 5, pp. 493-506.

McCabe, M. and C.M. Snyder (2014), "Identifying the effect of open access on citations using a panel of science journals", *Economic Inquiry*, Vol. 42, No. 4, pp. 1284-1300.

Miguel, S., Z. Chichilla-Rodrígues and F. de Moya-Anegón (2011), "Open access and scopus: A new approach to scientific visibility from the standpoint of access", *Journal of the American Society for Information Science and Technology*, Vol. 62, No. 6, pp. 1130-45.

Mooney, H. and M.P. Newton (2012), "The anatomy of a data citation: Discovery, reuse, and credit", *Journal of Librarianship and Scholarly Communication*, Vol. 1, No. 1, Art. No. eP1035, http://dx.doi.org/10.7710/2162-3309.1035, accessed 10 June 2015.

Murray, F.P. et al. (2009), "Of mice and academics: Examining the effect of openness on innovation", NBER Working Paper No. 14819, National Bureau of Economic Research, www.nber.org/papers/w14819, accessed 11 June 2015.

Neylon, C. and S. Wu (2009), "Article-level metrics and the evolution of scientific impact", PLOS Biology, 2009, Vol. 7, No. 11, Art. No. e1000242, http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1000242, accessed 11 June 2015.

OECD (2015), *Inquiries into Intellectual Property's Economic Impact*, OECD Publishing, Paris.

OECD (2014), *Measuring the Digital Economy: A New Perspective*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264221796-en.

OECD (2013a), Background paper for the TIP workshop on Open Science and Open Data, unpublished, DSTI/STP/TIP(2013)13.

OECD (2013b), "Public sector information: A review of the Recommendation", Working Party on the Information Economy, DSTI/ICCP/IE(2012)2/REV2,

OECD (2013c), "Knowledge Networks and Markets", *OECD Science, Technology and Industry Policy Papers*, No. 7, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k44wzw9q5zv-en, accessed 11 June 2015.

OpCit Project (2012), Open Citation (OPCIT) Project.

Paltoo, D.N. et al. (2014), "Data use under the NIH GWAS Data Sharing Policy and future directions", *Nature Genetics*, Vol. 46, No. 9, September.

Piwowar, H.A. and W.W. Chapman (2008), "Identifying data sharing in biomedical literature", *AMIA Annual Symposium Proceeding Archive*, pp. 596-600.

Piwowar, H. and T.J. Vision (2013), "Data reuse and the open data citation advantage", *PeerJ*, 1:e175, http://dx.doi.org/10.7717/peerj.175.

Piwowar, H., R.S. Day and D.B. Frisma (2007), "Sharing detailed research data is associated with increased citation rate", PLOS ONE, Vol. 2, No. 3.

Priem, J. and B.M. Hemminger (2010), "Scientometrics 2.0: Toward new metrics of scholarly impact on the social web", *First Monday*, Vol. 15, No. 7, 5 July.

Priem, J. et al., (2010), "Altmetrics: A manifesto", www.altmetrics.org/manifesto, accessed 11 June 2015.

Romeu, C. et al. (2014), "The SCOAP3 initiative and the open access article-processing-charge market: Global partnership and competition improve value in the dissemination of science", http://cds.cern.ch/record/1735210/files/?ln=it, accessed 11 June 2015.

Rowlands, I. and D. Nicholas (2005), "Scholarly communication in the digital environment: The 2005 survey of journal author behaviour and attitudes", *Aslib Proceedings*, Vol. 57, Issue 6, pp. 481-97.

Royal Society (2012), "Final report: Science as an open enterprise", Royal Society Science Policy Centre Report 02/12, https://royalsociety.org/policy/projects/science-public-enterprise/Report/, accessed 11 June 2015.

Sparks, S. (2005), *JISC Disciplinary Differences Report*, Rightscom, London.

Spiegler, D.B. (2007), "The private sector in meteorology: An update", *AMS Journals Online*, American Meteorological Society, http://journals.ametsoc.org/doi/pdf/10.1175/BAMS-88-8-1272, accessed 11 June 2015.

Swan, A. (2010), "The open access citation advantage: Studies and results to date", Research on Institutional Repositories (IRs), SelectedWorks, http://works.bepress.com/ir_research/31/, accessed 11 June 2015.

Taraborelli, D. (2008), "Soft peer review: Social software and distributed scientific evaluation", *Proceedings of the 8th International Conference on the Design of Cooperative Systems* (COOP'08), Carry-Le-Rouet, 20-23 May.

Thelwall, M. et al. (2013), "Do altmetrics work? Twitter and ten other social web services", PLOS ONE, Vol. 8, No. 5, Art. No. e64841, http://dx.doi.org/10.1371/journal.pone.0064841.

UNESCO (2012), *Policy Guidelines for the Development and Promotion of Open Access*, UNESCO Publishing.

Wagner, B. (2010), "Open Access Citation Advantage: An Annotated Bibliography", *Issues in Science and Technology Librarianship*, No. 60.

Ware, M. (2009), *Access by UK Small and Medium-Sized Enterprises to Professional and Academic Literature*, Publishing Research Consortium, Bristol.

Ware, M. and M. Monkman (2008), *Peer Review in Scholarly Journals: Perspective of the Scholarly Community – An International Study*, Publishing Research Consortium, Bristol.

Williams, H. (2010), "Intellectual property rights and innovation: Evidence from the human genome", NBER Working Paper 16123, National Bureau of Economic Research, July, http://www.nber.org/papers/w16213, accessed 11 June 2015.

Wouters, P. and R. Costas (2012), *Users, Narcissism and Control: Tracking the Impact of Scholarly Publications in the 21st Century*, SURF Foundation, Utrecht.

Zahedi, Z., R. Costas and P. Wouters (2013), "How well developed are altmetrics? Cross disciplinary analysis of the presence of alternative metrics in scientific publications", *Scientometrics*, Vol. 101, Issue 2, http://dx.doi.org/10.1007/s11192-014-1264-0.

Chapter Two

OPEN ACCESS TO SCIENTIFIC PUBLICATIONS

Defining open access

Although the term "open access" was first formally defined at a meeting in Budapest in early December 2001, it was preceded by other initiatives (on data for example, the Bermuda Principles were established in 1996 to enable the rapid and public release of genome data).[5] Out of that meeting came the so-called Budapest Open Access Initiative,[6] in which "open access" was defined as the

"free availability of scientific literature on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited."

The Budapest Open Access Initiative was followed by the Bethesda Statement,[7] which arose from a one-day meeting of scientists, funding agencies, librarians, scientific societies and publishers, held in April 2003. In October of the same year, the Max Planck Society in Germany convened a meeting on "open access to knowledge in the sciences and humanities". This meeting widened the discussion to include the humanities and produced the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (October 2003).[8]

Open access contributions may include original scientific research results, such as articles and monographs, as well as raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material. On the basis of the Berlin[9] and Budapest statements

---

[5] This chapter is drawn from inputs drafted by Krzysztof Siewicz, and Guibault, L. and T. Margoni (2014), "Legal aspects of open science and open data", Background paper for the OECD CSTP/TIP project on open science, Instituut voor Informatierecht, Universiteit van Amsterdam.

[6] Available at: http://www.opensocietyfoundations.org/openaccess/.

[7] Available at: www.earlham.edu/~peters/fos/bethesda.htm/.

[8] Available at: http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/.

[9] The Berlin Declaration defines a contribution that qualifies as open access:

1. "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 license 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 (community standards, will continue to provide the mechanism for enforcement of proper attribution and responsible use of the published work, as they do now), as well as the right to make small numbers of printed copies for their personal use."

and initiatives, the three following essential characteristics of open access emerge: *free accessibility*, *further distribution*, and *proper archiving* (Open Society Institute, 2012).

The recurring theme in all such documents is that "open access" is understood as free (gratis) availability of research material (publications, sometimes also research data) on the Internet without technical restrictions. Some declarations stress the importance of removing the legal restrictions as well (Budapest, Bethesda, Berlin), including the ability of legally unencumbered reuse within the definition of open access.

Gold, green and hybrid open access

A number of different implementations of open access digital publishing coexist; the prevailing forms are named *gold, green, and hybrid open access models*. The colour-based terms, initially popular in the United Kingdom, eventually gained worldwide acceptance.

Under the *gold model*, authors submit articles to open access journals – that is, journals that directly provide free open access to the articles they contain on line. Open access journals are usually licensed under one of the six core Creative Commons (CC) licences (see Boxes 2.3 and 2.4). Among the more successful open access journals and databases are the Public Library of Science (PLOS),[10] Biomed Central[11] and the open access alternative offered by Springer Open Choice Publishing.[12] The publishing cost and revenue in the gold model is usually recovered through APCs, which is a publication fee the author's institution or research funder has to pay. Alternatively, an open access journal can charge subscription fees for printed versions, and make only the electronic version openly accessible. Finally, a gold open access journal can rely on other means of funding (such as advertising or being sponsored by foundations) without charging either the authors or the readers.

A specific type of gold open access journal is the so-called hybrid journal, where an otherwise subscription-based journal makes specific articles available through open access, provided its APCs have been funded by the authors or their institutions. Hybrid journals have the advantage of increasing the possible venues for authors to publish via open access, as an increasing number of subscription-based journals allow this type of open access publishing. However, according to some, this model may involve paying twice for the same content, once in the form of APCs, and then again with payment by the journal's subscribers.

Open access publishing is rapidly evolving, and there are currently several open questions regarding best practices when it comes to APCs and their sustainability. According to Johnson (2015), both institutions and publishers need to constantly adapt processes and systems as more and more articles are

2. "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 organization that seeks to enable open access, unrestricted distribution, interoperability, and long-term archiving."

[10] See: www.plos.org/about/openaccess.html.

[11] See: www.biomedcentral.com/info/about/copyright.

[12] See: www.springer.com/sgw/cda/frontpage/0,11855,5-40359-12-161193-0,00.html.

published through gold open access. In addition, "the complex relationship among APC pricing, licensing, and embargo periods remain a subject for debate" (Johnson, 2015).

A different implementation – often considered to be complementary to the gold model while in reality possibly overlapping with it – is called the *green model*, and involves the author self-archiving the pre-print or post-print of their article. The authors provide access to their own published articles by making their own e-prints free for all. Self-archiving refers to providing open access to a publication by depositing it on the Internet, usually in a repository or through the author's webpage. Open access self-archiving is *not* self-publishing; it is not about online publishing without quality control (peer review); and it is not intended for writings for which the author wishes to be paid, such as books or magazine/newspaper articles. In specific cases the green copy of an article can be archived by the publisher instead of the author.

While the green model does meet the main OA requirements – namely free access, the possibility of copying, using and distributing the work, and archiving – the fact that the publications are first published through traditional channels means that authors retain only certain rights on their publication/data. Self-archived articles are usually accompanied by the text of a licence telling users what they can and cannot do with the article. By contrast, releasing research results using the gold model generally ensures broader and immediate access, clearer reuse possibilities, visibility, and "fundability" of research output on the Internet (Guibault, 2011). Insofar as gold open access involves APCs and is available only to certain journals, gold open access may limit researchers' choice of publisher.

Open access publishing models come in different shades of gold and green, and carry different advantages and disadvantages (Table 2.1). In the majority of countries that responded to the OECD survey (see Country Notes), green open access policies predominate, although many countries and institutions do provide funding to cover the costs of gold open access. The European Commission's Horizon 2020 policy for example allows authors to choose from both green and gold open access channels. The APC costs incurred in open access publishing are eligible for reimbursement from Horizon 2020 grants. CERN, the European Organization for Nuclear Research, launched the SCOAP[3] initiative to set APCs through a tendering process in January 2014 (Box 2.1). Even when the gold open access channel is used, some funding agencies and institutions require that the article be deposited in an open access repository.

Table 2.1 Different shades of open access publishing

	Description	Extra features	Advantages	Disadvantages
Gold open access (for profit)	The article is immediately available at the time of publication and, in most cases, the publication costs are covered through APCs paid by the author or the funder	Articles published in journals in which all articles are accessible Immediate open access may be given to the article only, or in some cases additional related material such as data sets,	Gold OA typically has full reuse rights under Creative Commons (CC-BY) Immediate access to the article with no embargo periods Publishers are increasingly offering innovative services	APCs costs need to be covered by funders or research Limited choice of publishing venue

		figures, images or videos	around gold open access publishing (example: Sage Open, Springer Open Choice) Some publishers such as *PLOS* offer fee waivers to authors without institutional funding	
Gold open access **(not for profit)**	The article is immediately available at the time of publication	Articles published in journals in which all articles are accessible Immediate open access may be given to the article only, or in some cases additional related material such as data sets, figures, images or videos	Gold OA typically has full reuse rights under Creative Commons (CC-BY) Immediate access to the article with no embargo periods Lower costs for authors	Limited choice of publishing venue
Hybrid OA	A hybrid open access journal is a subscription journal in which some of the articles are open access. This status typically requires the payment of an APC or publication fee to the publisher If a payment for OA is received, this is offset in due course, according to publisher rules	The cost of APCs varies considerably depending on the journal Immediate open access may be given to the article only or in some cases additional related material such as data sets, figures, images or videos	Authors wanting to publish in an open-access journal, are not limited to the relatively small number of "full" open-access journals; they can publish in hybrid OA journals of the main publishers Hybrid journals are low risk for publishers to set up, because they still receive subscription income	Limited (although growing) number of journals where it is possible to publish The high price of hybrid APCs has led to low uptake of the hybrid open access option
Green Open Access (pre-print versions)	Pre-print version of articles (i.e. prior to submission to a journal for peer review) which are accessible online, typically at personal	Pre-print versions of the article commonly available online include working papers and/or unpublished	The article can be uploaded in multiple venues: from institutional or disciplinary repositories to	Green OA does not typically have full reuse rights under Creative Commons licence (CC-BY)

	or institutional webpages, or institutional or subject repositories	version of an article	personal websites No extra costs for authors (no APCs need to be paid) Authors have complete freedom in the choice of publishing venue	The version of the article deposited on line has not been subjected to peer review Maintenance costs of repositories
Green Open Access (accepted author manuscript)	Versions of articles (i.e. after undergoing peer review and incorporating any revisions required for acceptance by a journal) which are accessible online, typically at personal or institutional webpages or institutional or subject repositories	Journal-specific embargos may apply	No extra costs for authors (no APCs need to be paid) Authors have compete freedom in the choice of publishing venue The version of the paper available on line has been subjected to peer review	Green open access of the accepted manuscript often involves an embargo period that varies considerably (generally up to 24 months) Green OA does not typically have full reuse rights under Creative Commons licence (CC-BY) Maintenance costs of repositories

Both gold and green open access models are currently being promoted by governments, funding agencies, universities and research centres, as well as by other open science stakeholders in OECD member countries and beyond. While the green model is the default model for basic open access in the majority of OECD countries, variants of the gold model have emerged to respond to author preferences to publish in leading journals and attempts by publishers to develop new services to make for-profit models competitive. Some countries have launched initiatives to promote the emergence of innovative business models around open access (Box 2.2).

Box 2.1 The SCOAP^3 initiative: an international open access partnership in high-energy physics

The SCOAP^3 (Sponsoring Consortium for Open Access Publishing in Particle Physics) initiative is an international partnership to make scholarly literature in the field of high-energy physics (HEP) open access. SCOAP^3 is an initiative developed by CERN, the European Organization for Nuclear Research, in partnership with organisations in 37 countries. SCOAP^3 has been operating since January 2014, and is estimated to cover around 4 000 articles per year. SCOAP^3 introduced competition among publishers in the APCs market. After a tendering process started in June 2012, CERN decided to grant publishers a three-year contract (over 2014-16) to offer open access publishing of articles in the HEP field. The budget allocated to SCOAP was EUR 10 million. Publishers participating in the tendering process had to specify the requested APCs, which Creative Commons licence they would adopt, and which format they would grant SCOAP^3 for further dissemination (i.e XML, PDF, etc.).

Only journals offering the best value for money could be retained in the tendering process. Value was defined through

quality criteria such as the journal impact factor (a measure of the prestige of an individual journal) and the services provided. Through the tendering process, 12 journals from 12 different publishers were retained to participate in SCOAP^3: they represent the vast majority of HEP scientific literature. Only one of the retained journals decided not to sign a contract and participate in SCOAP^3.

Some initial assessment of the SCOAP^3 experiment shows that, thanks to the open competitive process, the initiative manages to obtain better "value for money" with respect to APCs. SCOAP^3 is an interesting example of how international partnerships among different open science actors (the publishing industry, libraries, national funding agencies and international organisations) may develop innovative business models.

Source: Romeu, C. et al. (2014), "The SCOAP3 initiative and the open access article-processing-charge market: Global partnership and competition improve value in the dissemination of science", available at http://cds.cern.ch/record/1735210/files/SCOAP3-APC.pdf.

Box 2.2 Supporting alternative business models for open access, the case of DFG initiatives

The "Infrastructure for Electronic Publications and the Digital Communication of Science", funded and coordinated by the German Research Foundation (DFG), aims to fund pilot projects and model-type projects that stand out for technical and/or organisational innovations or for the development, testing and refinement of innovative business models in the area of electronic open-access publications. The objective is optimal creation, open provision and distribution of genuinely digital publications of scientific papers and the guarantee of their long-term availability. Project results must be made freely available and accessible to third parties for use in other contexts.

The programme covers a wide range of topics. It encompasses the development of tools for electronic publishing, the independent creation of electronic publications by researchers, the construction and development of networked publications repositories, the promotion of the open-access model in various scientific communities, the organisation of collaborative models with the publishing industry, and ensuring the long-term availability of genuinely digital content.

Source:
http://www.dfg.de/en/research_funding/programmes/infrastructure/lis/funding_opportunities/infrastructure_electronic_digital_publications/index.html

Open access publishing and IP protection[13]

Copyright and other intellectual property rights play a decisive role in the way scientific output is being disseminated and used by the scientific community as they underpin the relevant licensing practices. Copyright law and other relevant intellectual property rights support, impede, or are neutral towards the implementation of open access principles for the dissemination of scientific results.

Although implementation of open access principles is based on contractual arrangements between authors, publishers and universities, the framework set by the copyright regime is a determinant in how those arrangements are to take form. The way copyright law defines the scope of rights and recognises limitations and exceptions on these rights serves as the backbone to the licensing agreements.

[13] This section is largely based on Guibault L. and T. Margoni (2014), Legal aspects of open science and open data, Background paper for the OECD CSTP/TIP project on open science, Instituut voor Informatierecht, Universiteit van Amsterdam.

The main way to determine how the IP regime in general and copyright law in particular stand towards open access in science is to look at the scope of rights granted to scientific results and possible exceptions to these rights in different jurisdictions and the manner in which both rights and exceptions are defined form the basis for the exploitation of rights of scientific results, either by the researcher, their employing institution or a publisher (Guibault, 2011). The stronger the rights (see Box 2.3), the stronger the possibility to exercise those rights through licensing (see Boxes 2.4 and 2.5), either by placing restrictions on use following the traditional model or by promoting broad unrestricted reuse following open access principles. Other elements – such as the principle of originality (e.g. the question of what is or is not protected by an IP right), the ownership of rights and the duration of protection – are important factors in assessing a regime's inclination towards the open access model, but are less decisive.

Box 2.3 Different types of openness: libre vs. gratis open access

Due to the different approaches towards legal restrictions of reuse, and the ambiguity of the words "open" (and "free"), it is often proposed to distinguish between *gratis and libre open access*. The term *gratis open access* is used to denote public availability of scientific publications (and sometimes also of research data) without payment or technical restrictions. The term *libre open access* encompasses the former, with an additional explicit requirement that the material not be subject to legal restrictions.

The gratis/libre dichotomy naturally leads to the following question: how many restrictions have to be removed in order to qualify material as "libre open access" material? There are many different answers to this question. For example, some argue that "there is more than one kind or degree of libre open access",[14] and seem to treat as "libre open access" any situation that removes at least some restrictions; hence, materials under CC NC or CC ND licences (see Box 2.4) would be libre open access. (See Box 2.4 for a description of the different licence models.) An opposite approach may be derived from the understanding of "freedom" proposed in the works of Richard Stallman with regard to free software. Under Stallman's Free Software Definition,[15] Free (libre) Software exists when a user is able to exercise all four freedoms defined therein: i) freedom to run the programme; 2) freedom to study the programme; 3) freedom to redistribute; 4) freedom to distribute copies of modified versions, which practically cover the full scope of the monopoly granted by copyright law. The understanding of "freedom" as proposed by Stallman with regard to software has been adapted to other intangibles in the Definition of Free Cultural Works[16] (DFCW).

Even the freedom defined by Stallman with regard to software and transposed to cultural works in the DFCW does not mean that there are no restrictions at all. For example, making a resource free under these definitions does not mean that it may be used in a manner that constitutes a breach of moral rights or privacy. Certain user obligations are explicitly allowed, such as elaborate attribution obligations (BY clauses – see Boxes 2.4 and 2.5) or copyleft clauses, which prohibit restriction on the freedom of others (GPL or CC BY-SA, see Boxes 2.4 and 2.5).

Source: Text provided by Krzysztof Siewicz.

Box 2.4 Creative commons (CC) licensing models

Although anyone may draft their own licences that satisfy free or gratis criteria, in practice model clauses are usually used. With regard to works other than software, the most popular model clauses are stewarded by the Creative

[14] http://legacy.earlham.edu/~peters/fos/newsletter/08-02-08.htm#gratis-libre.

[15] See www.gnu.org/philosophy/free-sw.html.

[16] http://freedomdefined.org/Definition.

Commons non-profit organisation. Creative Commons offers six basic model clauses, two of which satisfy the above criteria of a free licence: CC BY and CC BY-SA.

Each of the six different CC model clauses contains a different set of user obligations. All of them require attribution (BY). The other clauses are:

1. NC (limitation of use to non-commercial uses)
2. SA (requirement that derivatives are licensed under the same licence as the original)
3. ND (limitation of use to the original only; no derivatives).

Examples of licenses (www.creativecommons.org):

- **CC BY:** "This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials"

CC BY-SA: "This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to "copyleft" free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects"

- **CC BY-NC:** "This license lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms"
- **CC BY-ND**: "This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you"

The resulting number of six model clauses follows from the fact that BY (the attribution) is present in all of them, and SA (the requirement that derivatives are licensed under the same licence as the original) and ND (limitation of use to the original only with no derivatives) are mutually exclusive. Under each of the above symbols, there is a clause drafted in legal language, which specifies the exact scope of the accompanying obligation.[17] The Creative Commons organisation itself is not a party to the licences, and does not hold any record of the licences granted under the model clauses.

CC model clauses are periodically reviewed, which leads to revised versions of all six clauses, marked with numbers similar to the numbering of software versions (1.0, 2.0, 2.5, 3.0, 4.0). The current version is 4.0, which addresses certain issues discovered in the course of operation of the previous version. Up until 4.0, CC licences were also "ported" into national jurisdictions, leading to yet one more distinction: between the universal, unported version, and the versions more accommodated to a particular national law (e.g. CC BY 3.0 Unported vs. CC BY 3.0 PL). Starting with 4.0, CC has ended this practice, and currently CC model clauses are drafted in such a way that the universal wording covers the widest possible approaches found in national laws.

In 2013, a new Creative Commons licence, the Creative Commons 3.0 Intergovernmental Organisation (IGO) License, was developed thanks to the joint effort of several international organisations led by the collaboration between the World Intellectual Property Organisation (WIPO) and the OECD. This licence makes it easier for IGOs to share their studies and report data sets and other material on line. The licence is similar to other CC licences, but it also includes a provision related to mediation or arbitration mechanisms for resolutions of disputes involving IGOs.

Source: www.creativecommons.org and text provided by Krzysztof Siewicz

[17] Full legal texts of all CC model clauses may be found at http://creativecommons.org/licenses.

Box 2.5 Licensing models and options

When discussing licensing models and options, one should be heedful of the distinction between 1) gratis and 2) libre open access (see Box 2.3). It is also possible to introduce a parallel distinction of legal relations, into: 1) relations between the author and the distributor (publisher, repository, etc.); and 2) relations between the distributor (publisher, repository, the author) and the user. This leads to a 2x2 matrix of 4 different situations:

1. Gratis OA in relations between the author and the distributor

Since the gratis open access does not need to involve the granting of any licence to the end-user, the legal relationship between the author and the distributor can follow the standard pattern of a publishing contract. It is only necessary that the contract clearly allow the distributor to make the publication available on the publicly available Internet.

If parties so desire, the contract may include a variety of optional clauses, such as for example a clear obligation for the distributor to provide for open access of the material, an embargo period, etc.

2. Libre OA in relations between the author and the distributor

Libre open access involves the granting of a free licence (CC BY, CC BY-SA) to the end-user, so it implies that the licensor must obtain the necessary scope of rights, or is a copyright holder himself/herself. The scope of acquired rights has to be equal to or more than the scope of rights granted in the CC licence, but the rights do not have to be transferred to the future CC licensor. A CC licence may be granted as a sub-licence from a person who holds a licence from the copyright holder. Hence, the author may retain full copyrights and merely authorise a distributor to sublicense the work under a CC licence. There is a wide variety of possible legal arrangements. In particular, it is possible that the distributor obtains from the author the same CC licence as anyone else. A more traditional option is also possible: the author transfers copyrights to the publisher, who grants CC licences to end-users. A contract between the author and the distributor may include many side clauses, as mentioned above – a clear obligation to provide open access, an embargo period, etc.

3. Gratis OA in relations between the distributor and the user

Under the gratis open access, there is no licence necessary for the end-user, since there is no need to remove any copyright restrictions in order to qualify as gratis open access. But it may be possible that gratis open access material is accompanied by some licence, such as a CC NCor CC ND licence. In such a case, the remarks in Point 4 below apply.

4. Libre OA in relations between the distributor and the user

Under the libre open access, the end-user is granted a free licence, which removes copyright from the material restrictions that are in conflict with user freedoms. The licence may be regarded as free if it allows for an unlimited, gratis, non-exclusive use of a work and its derivatives. It may, however, contain certain obligations that do not affect the core of the freedom, such as attribution or copyleft (obligations not to restrict the freedoms of others with regard to the work and its derivatives).

Source: Text provided by Krzysztof Siewicz.

Another key characteristic of copyright law is found in the balance it strikes between uses reserved for rights-holders and uses that are free – or in other words, uses for which no authorisation is required. Two typologies can be found: uses that are not covered by copyright law, and uses that are covered but exempted from authorisation, although sometimes a “fair compensation” is required. The second category,

"exceptions and limitations to copyright" generally permit the use of a copyrighted work without the permission of the author or copyright owner. Such exceptions and limitations may be provided by statue or case law, including uses covered by "fair use" or "fair dealing" provisions. The policy rationales for exceptions and limitations vary according to national law. They include the protection of constitutional and/or fundamental rights, the regulation of industry practice and competition, the dissemination of knowledge, or market failure considerations. Examples of this second category can be seen in cases such as quotation;[18] illustration for teaching;[19] certain articles on current economic, political, or religious topics;[20] and the reproduction of works for the purpose of reporting current events.[21] Once more, given the minimum level of protection approach of the conventions, signatories are free to enact other exceptions and limitations, as long as these apply only to certain special cases that do not conflict with a normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the author (Senftleben, 2004; Gervais 2005; Ricketson 2003).[22]

Although no generalisation can be made, some of the countries that actively encourage compliance with open access principles for the publication of publicly funded research results seem to steer the copyright reform in a more flexible and research-friendly direction. The United Kingdom is a good example of this: while the Research Council has adopted a "Golden Road" policy, mandating researchers to publish results under a Creative Commons Attribution Licence 4.0, the legislator has also proceeded with the adoption of new exceptions on copyright, including a specific exception for text and data mining. The German research council may not have officially opted to issue an open access mandate to its grant recipients, but the legislator did modify the copyright act to make it easier for authors to comply with the contractual arrangements with publishers.

The legal framework in Europe

In 2012, the European Commission published its Communication to the European Parliament and the Council entitled "Towards better access to scientific information: Boosting the benefits of public investments in research".[23] As the Commission observes, "discussions of the scientific dissemination system have traditionally focused on access to scientific publications – journals and monographs. However, it is becoming increasingly important to improve access to research data (experimental results, observations and computer-generated information), which form the basis for the quantitative analysis underpinning many scientific publications".[24]

---

[18] See Art. 10 (1) of the Berne Convention.

[19] See Art. 10 (2) of the Berne Convention.

[20] See Art. 10bis (1) of the Berne Convention.

[21] See Art. 10bis (2) of the Berne Convention.

[22] See Art. 9(2) of the Berne Convention. This article is the first appearance in an international agreement of the so-called Three-step Test. Art. 9(2) was introduced in the 1967 Stockholm revision of the Convention, and in its Berne formulation applies only to the right of reproduction. Successively, it was introduced in all the other major international agreements.

[23] Brussels, 17.7.2012 COM(2012) 401 final.

[24] Id., p. 3.s.

The Communication marks an official new step on the road to open access to publicly funded research results in science and the humanities in Europe. Scientific publications are no longer the only elements of an open access policy: research results upon which publications are based must now also be made available to the public. To implement this policy, the European Commission set up a pilot initiative on open access to peer-reviewed research articles in its Seventh Research Framework Programme (FP7), otherwise known as the OpenAire project (see Box 4.6), to ensure that the results of the research it funds are disseminated as widely and effectively as possible so as to guarantee maximum exploitation and impact in the world of researchers and beyond.

The European Open Access Policy is not binding on the EU Member States, which are free to adopt the policy that best suits the needs of their own scientific community. This leads to a mosaic of open access policies across Europe, ranging from the mandatory golden road for publications and data put in place by the Research Councils of the United Kingdom (RCUK), to the preference for gold open access in the Netherlands, to the green road for publications in Germany. In the recent years, for example, the national research councils of the UK and the Netherlands have issued policy statements according to which research grants will be awarded only provided that the applicants commit to publishing their results, both publications and data, under open access conditions.

The legal framework in the United States

According to the US Government directive issued by the Office of Science and Technology Policy (Public Access Directive), all federal agencies with more than USD 100 million per year in research and development expenditure are required to develop plans to make the published results of federally funded research freely available to the public within one year of publication.[25] Additionally, the Fair Access to Science and Technology Research Act (FASTR) was introduced in the US Congress at the beginning of 2013[26]. If passed, such a bill would back up the goals of the Directive with the more robust structure of a legislative tool. The bill is similar to the Directive, with small but significant differences in terms of the number and types of agencies covered, the maximum embargo period (six months versus one year), and the reference to publications (both) or also other research data (Directive).[27] At the time of writing, legislations mandating public access policies had been passed for the US Department of Labor, the Department of Education and the Department of Health and Human Services, which includes the National Institutes of Health.

The recent amendments to copyright laws in Germany…

In Germany, a recent addition to the Copyright Act deals directly with the issue of licensing scientific publications created thanks to public funding, and reads:

---

[25] See www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research with direct links to the Directive.

[26] The bill was reintroduced to the new Congress in early 2015.

[27] The text of the bill is available at: http://doyle.house.gov/sites/dxoyle.house.gov/files/documents/2013%2002%2014%20DOYLE%20FASTR%20FINAL.pdf.

> "The author of a scientific contribution which is the result of a research activity publicly funded by at least fifty percent and which has appeared in a collection which is published periodically at least twice per year has the right, even if he has granted the publisher or editor an exclusive right of use, to make the contribution available to the public in the accepted manuscript version upon expiry of 12 months after first publication, unless this serves a commercial purpose. The source of the first publication shall be indicated. Any deviating agreement to the detriment of the author shall be ineffective.[28]"

This provision is intended to allow the author of a scientific work that is generated in the context of (at least 50%) publicly funded research and published in a periodical collection (at least biannual), to make the accepted version of the manuscript publicly available for non-commercial purposes after an embargo period of 12 months. The right to republish cannot be limited by contractual agreements, which means that even if the author has licensed all exclusive rights to a publisher, that author will still be entitled to the right of republication (Hilty et al., 2013; Moscon, 2013).

… and the United Kingdom

The United Kingdom has recently implemented[29] a number of amendments to the national copyright framework, which are expected to facilitate the conduct of scientific research and analysis. Of particular interest is the new section 29A, which reads as follows:

> "29A Copies for text and data analysis for non-commercial research
>
> 1. The making of a copy of a work by a person who has lawful access to the work does not infringe copyright in the work provided that—
> (a) the copy is made in order that a person who has lawful access to the work may carry out a computational analysis of anything recorded in the work for the sole purpose of research for a non-commercial purpose, and
> (b) the copy is accompanied by a sufficient acknowledgement (unless this would be impossible for reasons of practicality or otherwise).
> 2. Where a copy of a work has been made under this section, copyright in the work is infringed if—
> (a) the copy is transferred to any other person, except where the transfer is authorised by the copyright owner, or
> (b) the copy is used for any purpose other than that mentioned in subsection (1)(a), except where the use is authorised by the copyright owner.
> 3. If a copy made under this section is subsequently dealt with—
> (a) it is to be treated as an infringing copy for the purposes of that dealing, and
> (b) if that dealing infringes copyright, it is to be treated as an infringing copy for all subsequent purposes.

[28] See Art. 38(4) of the German Copyright Act.

[29] These provisions refer to personal copies for private use, quotation and parody, entered into force in 2014.

4. In subsection (3) "dealt with" means sold or let for hire, or offered or exposed for sale or hire.
5. To the extent that a term of a contract purports to prevent or restrict the making of a copy which, by virtue of this section, would not infringe copyright, that term is unenforceable."

The provision basically clarifies that making a copy of a work for the purpose of text and data mining (TDM) is not an infringement of the copyright in the work provided that this is made for the sole purpose of research for non-commercial purposes. The provision makes clear that any contractual agreement that has the effect of limiting the possibility of making copies under this provision is unenforceable. The exception does not cover the *sui generis* database right (SGDR, see following sections), however. It is the opinion of the UK Government that the SGDR's fair dealing exception for non-commercial scientific uses offers a parallel defence adequate to the present needs.[30]

Open access publishing and its legal implications

The author's choice of a given publication path can be driven by several factors, including their expectation of the impact this may have. While open access may increase the accessibility of the work, researchers have to contend with the fact that established journals built on traditional publishing models tend to have a long-standing reputation and established reviewer networks. Acceptance by a journal confers upon the author and their work some of the implied reputation associated with the articles that have been previously published in the journal, regardless of how may citations the document receives relative to the "norm" for that particular title. To the extent that several incentives, such as promotion decisions and research grants, hinge on such metrics, the decision to publish on an open access basis cannot be treated independently of factors such as quality and relevance. This needs to be taken into account in the appraisal of different models and policies.

Scientific authors willing to publish through open access channels need to consider the legal implication of such a choice. Open access publishing (through both the green and the gold models) requires a collaborative effort involving different stakeholders – notably scientific authors, research institutions, funding agencies and publishers. Scientific authors in particular need to follow a multi-step procedure:

- Identify the scope of rights necessary in order to provide open access.
- Acquire these rights – or, as happens more often, the author has to prevent disposing of them in the first place.
- Undertake certain activities in order to enable open access to the work published in a journal (or accepted for publication in a journal).

The scope of rights necessary for open access depends on the type of open access. There are more rights necessary for the libre, since it implies the grant of a free licence (CC BY or CC BY-SA – see Boxes 2.2 and 2.3), and no one may grant fewer rights than he or she holds. For the gratis open access, it suffices

[30] See the official opinion of the UK Government in the document titled "Technical Review of Draft Legislation on Copyright Exceptions: Government Response", at 13 ("The Government's view is that this existing exception will permit the extraction of whole works if required for text and data mining through the provision for "fair dealing with a substantial part").

that the author has only the right to place the work on the Internet. Other rights may be transferred to a publisher or a third party.

All rights necessary for open access are held by the author alone only in simple cases, which generally apply to the typical situation of publicly funded research (e.g. the author is the only author, the publication does not take place in the course of employment, and there are no contracts concluded with regard to the publication). In case of joint works, works made in the course of employment, derivative works (e.g. translations of other publications), as well as when a contract with the publisher has already been concluded, usually it will be necessary to obtain the respective party's consent for open access. The exact legal basis for such consent and the appropriate form, etc. will be different depending on circumstances and the applicable law. Whenever there is a need of consent, it should be informed and explicit – especially in case of the libre open access, which involves the grant of a free licence.

Most complicated cases include situations where works are already subject to a contract. Due to the contractual freedom, there are practically no limits as to the combinations of legal relationships resulting from such contracts. This means that individual legal advice may be necessary in order to determine whether the contract leaves the author with sufficient rights, and if not, how to arrange for them. Therefore, the simplest solution is to avoid contracts that may have such a limiting effect, and use non-exclusive licences only (a copyright owner may grant an unlimited number of different non-exclusive licences, including free licences). Also, open access should be explicitly negotiated for with the other parties beforehand. While there are some examples of publishers who refuse to negotiate their standard-form contracts for copyright transfer, such negotiations are often possible and practicable, and a simple change of such a contract into a non-exclusive licence will not usually require sophisticated legal advice.

In some cases, publishers already do not limit open access in their standard-form contracts, or even explicitly provide for it, as they implement open access in gold or green ways themselves. Standard-form contracts however may involve a copyright transfer with a licence back to the author allowing for only limited open access options (e.g. only gratis but not libre; after the embargo period; not the final published version but only the author's accepted manuscript; etc.). A general overview of publishers' open access policies may be found in the SHERPA/ROMEO database.[31]

Currently, for an individual author who wishes to make his/her publication open access, the procedure can be cumbersome; individual negotiations, for example, can be a burden on the author. Institutions such as employers (e.g. universities), as well as funding agencies, have a wide variety of legal tools to help individual authors with their relations with publishers. First, institutions may legally bind the author to follow a certain open access policy. Second, institutions often act as publishers themselves, which means that they can offer an immediate solution for authors who do not pursue any higher-evaluated journals. Third, apart from legal obligations, institutions may provide authors with organised support in their relations with the publishers, in both legal and technical senses.

One of the lessons from observing successful implementation of open access policies at the institutional level is that in order to be robust, policy should go beyond general support for open access.

[31] www.sherpa.ac.uk/romeo/.

The policy should also examine what type of open access (gratis/libre, green/gold – see Box 2.3 and Table 2.1) is pursued, and whether it is required or simply recommended.

Open peer review

The main purpose of scholarly peer review is to maintain the high quality of published research results, which is highly valued by researchers (Ware and Monkman, 2008). To facilitate free and unbiased expression of reviewers' opinions, peer review has traditionally been either single-blind (the reviewer knows the identities of the authors, but remains anonymous to them) or double-blind (both the reviewer and the authors remain anonymous). The double-blind peer review requires extra effort on the authors' side to prepare an anonymised version of their manuscript or grant proposal (no names, references to earlier works, etc.)

However, traditional peer review has several shortcomings. Firstly, this model gives few incentives to reviewers: they are not credited when they spend their time and energy on writing reviews. Secondly, the process is not fully transparent. Some critics argue that more widespread access to the data may increase the chances of avoiding the publication of articles containing incorrect results or conclusions. Post-publication peer review is viewed as an alternative, although it too is not without its challenges.[32]

Several studies raise concerns about the quality of the research results published in scientific journals (see for example Ioannidis 2005 and cited references). Although it is difficult to estimate the number of published scientific articles containing incorrect conclusions, the number of retractions may provide information on the problems associated with traditional peer-review verification of scientific results. Grieneisen and Zhang (2012) surveyed 42 of the largest bibliographic databases for major scholarly fields and publisher websites. They found that the number of retractions has increased considerably after 2001. Retractions happen more in fields such as medicine, life sciences and chemistry than in fields such as mathematics, physics, engineering and the social sciences. According to the study, the main cause of retraction is publishing misconduct (such as plagiarism, authorship or copyright issues), followed by incorrect use of data or interpretation and research misconduct (such as the use of fraudulent or fabricated data).

Partly to address the above-mentioned issues, open peer review models are emerging, in many cases to complement traditional peer reviewing models. For example, F1000Research, an open access journal in the life sciences, has adopted an open refereeing model, in which reviewers' and authors' responses are available publicly.

[32] See for example http://www.timeshighereducation.co.uk/news/can-post-publication-peer-review-endure/2016895.article.

REFERENCES

Gervais, D. (2005), "Towards a new core international copyright norm: The Reverse Three-Step Test", *Marquette Intellectual Property Law Review*, Vol. 9, p. 1.

Grieneisen, M.L. and M. Zhang (2012), "A comprehensive survey of retracted articles from the scholarly literature", PLOS ONE, Vol. 7, No. 10, Art. No. e44118.

Guibault, L. (2011), "Owning the right to open up access to scientific publications", in L. Guibault and C. Angelopoulos (eds.), *Open Content Licences: From Theory to Practice*, Amsterdam University Press, Amsterdam, pp. 137-167.

Guibault, L. and T. Margoni (2014), "Legal aspects of open science and open data", Background paper for the OECD CSTP/TIP project on open science, Instituut voor Informatierecht, University of Amsterdam.

Hilty, R. et al. (2013), "Zum Referentenentwurf eines Gesetzes zur Einführung einer Regelung zur Nutzung verwaister Werke und weiterer Änderungen des Urheberrechtsgesetzes sowie des Urheberrechts-wahrnehmungsgesetz", Stellungnahme des Max-Planck-Instituts für Immaterialgüter- und Wettbewerbsrecht zur Anfrage des Bundesministeriums der Justiz vom 20 February 2013, www.ip.mpg.de/files/pdf2/Stellungnahme-BMJ-UrhG_2013-3-15-def1.pdf.

Ioannidis, J. P. A. (2005), "Why Most Published Research Findings are False", *PLOS Medicine*, Vol. 2, p. 696.

Johnson, R. (2015), "Making open access work for authors, institutions and publishers", Report on an Open Access Roundtable hosted by Copyright Clearance Center, Inc., www.copyright.com/content/dam/cc3/marketing/documents/pdfs/Report-Making-Open-Access-Work.pdf, accessed 16 June 2015.

Moscon, V. (2013), "Open access to scientific articles: Comparing Italian with German law", Kluwer Copyright Blog, http://kluwercopyrightblog.com/2013/12/03/open-access-to-scientific-articles-comparing-italian-with-german-law, accessed 16 June 2015.

Open Society Institute (2005), *Open Access Publishing and Scholarly Societies – A Guide*, OSI, New York, p. 6; see also Suber, 2012.

Ricketson, S. (2003), "WIPO study on limitations and exceptions of copyright and related rights in the digital environment", World Intellectual Property Organization, www.wipo.int/edocs/mdocs/copyright/en/sccr_9/sccr_9_7.pdf, accessed 16 June 2015.

Romeu, C. et al. (2014), "The SCOAP3 initiative and the open access article-processing-charge market: Global partnership and competition improve value in the dissemination of science", http://cds.cern.ch/record/1735210/files/?ln=it, accessed 11 June 2015. Senftleben, M. (2004), *Copyright, Limitations and the Three-Step Test*, Kluwer Law International, London.

Suber, P. (2012), *Open Access*, MIT Press.

Ware, M. and M. Monkman (2008), *Peer Review in Scholarly Journals: Perspective of the Scholarly Community – An International Study*, Publishing Research Consortium, Bristol.

Chapter Three

OPEN RESEARCH DATA

Data-driven scientific research

Data and measurement have always been fundamental to science. The advent of new instruments and methods of data-intensive exploration has prompted some to suggest the arrival of "data-intensive scientific discovery", which builds on the traditional uses of empirical description, theoretical models and simulation of complex phenomena (BIAC, 2011). This could have major implications for how discovery occurs in all scientific fields (Hey and Trefethen, 2003; Jirotka et al., 2006; Anderson, 2004; Bell, Hey and Szalay, 2004). Data-driven science allows the development of scientific experiments as well as computer-based algorithmic simulations, even in those fields that traditionally were less data-intensive than others. Frischmann (2012) and OECD (2015) suggest looking at data as *infrastructure* (see Box 3.1).

Box 3.1 Data as infrastructure

Most data (not all) can in principle be considered as infrastructural resources, as they are "shared means to many ends" that satisfy all three criteria of infrastructure resources highlighted by Frischmann (2012). These criteria include:

- *Data are non-rivalrous goods* – (Non-)rivalry, or (non-)rivalrousness of consumption describes the degree to which the consumption of a resource affects the potential of the resource to meet the demands of others. Data are a non-rivalrous good that can be consumed in principal an unlimited number of times. This property is at the source of significant spillovers that provide the major theoretical link to total factor productivity growth, according to a number of scholars including Corrado et al. (2009). While it is widely accepted that social welfare is maximised when a pure rivalrous good is consumed by the person who values it the most, and that the market mechanism is generally the most efficient means for rationing such goods and for allocating resources needed to produce such goods, this is not always true for non-rivalrous goods (Frischmann, 2012). Social welfare is not maximised when the good is consumed only by the person who values it the most, but by everyone who values it. Maximising access to the non-rivalry good will in theory maximise social welfare, as every additional private benefit comes at no additional cost.

- *Data are capital goods* – Data are not a consumption good, or an intermediate good. In most cases, data can be classified as capital goods. The UN (2008) System of National Accounts (SNA) defines a consumption good or service as "one that is used […] for the direct satisfaction of individual needs or wants or the collective needs of members of the community". In contrast, intermediate goods and capital goods are used as inputs to produce other goods. Capital goods, according to the OECD, are "goods, other than material inputs and fuel, used for the production of other goods and/or services". In contrast to intermediate goods, such as raw materials (e.g. oil), capital goods are not used up, exhausted, or otherwise transformed when used as input to produce other goods. In most cases data are usually used as an input for goods or services, and this is in particular true for large volumes of data, which are means rather than ends themselves. Furthermore, data are also not an intermediate good as they are not exhausted when used, given their *non-rivalrous nature*. As with most capital goods, data can depreciate in particular when it becomes less relevant for its particular intended purpose it is intended to be used.

- *Data are general-purpose inputs* – As Frischmann (2012) explains, "infrastructure resources enable many systems (markets and nonmarkets) to function and satisfy demand derived from many different types of users". They are not inputs that have been optimised for a special limited purpose, but "they provide basic, multipurpose functionality". Data may often be collected for a particular purpose, and in the case of personal

data the ex ante specification of the purpose. However, there is theoretically *no limitation on what purposes data can be used for*, and in fact many of the benefits of data sharing arise from the reuse of data in ways that were or could not be anticipated when the data were collected. In addition, the reuse of data created in one domain may lead to further insights when applied in another. This is apparent in the case of public sector data, where data sets used originally for administrative purposes are reused by entrepreneurs to create new services that were never foreseen when the data were originally created.

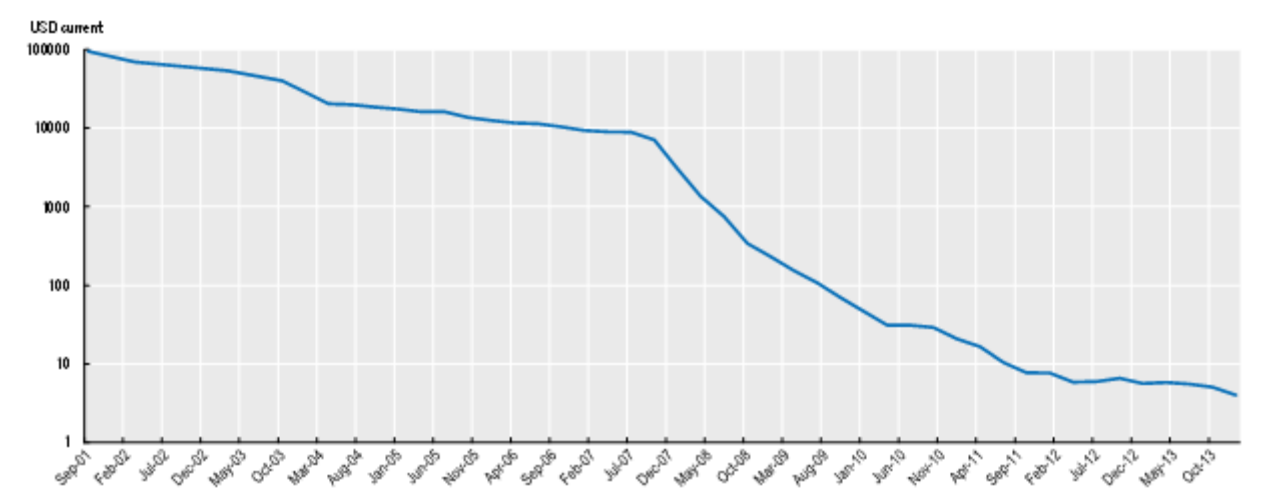
Source: OECD (2015), *Data-driven Innovation for Growth and Well-being*, OECD Publishing, Paris; Frischmann, B.M. (2012), *Infrastructure: The Social Value of Shared Resources*, Oxford University Press.

Data analytics tools (such as machine learning or pattern recognition techniques) are increasingly used by scientists to gain knowledge of phenomena and to test or validate models. Large-scale data sets allow computer-based experiments and simulations, even in those fields where traditional lab experiments were impossible or too difficult to organise. With sufficiently large data sets, machines can detect complex patterns and relationships that are invisible to researchers (Anderson, 2008; Bollier, 2010). In addition, data science and algorithmic-based experiments and research represent *per se* an opportunity for innovation and scientific discovery: fields such as computer or data science are currently exploiting big data as an opportunity to develop new and more efficient algorithms for data analytics, to be used by researchers active in different disciplines and fields (in both the public and private sector).

New instruments such as super colliders or telescopes, but also the Internet as a data collection tool, have been instrumental in new developments in science, as they have changed the scale and granularity of the data being collected. The Digital Sky Survey, for example, which started in 2000, collected more data through its telescope in its first week than had been amassed in the history of astronomy (*The Economist*, 2010), and the new SKA (square kilometre array) radio telescope could generate up to 1 petabyte of data every 20 seconds (EC, 2010). Furthermore, the increasing power of data analytics has made it possible to extract insights from these very large data sets reasonably quickly. In genetics, for instance, DNA gene sequencing machines based on big data analytics can now read about 26 billion characters of the human genetic code in seconds. This goes hand in hand with the considerable fall in cost of DNA sequencing over the past five years (Figure 3.1).

Figure 3.1 Cost of genome sequencing

2011-14, cost per genome, logarithmic scale



Source: OECD (2014), *Measuring the Digital Economy: A New Perspective*, OECD Publishing, Paris.

These new developments, scaled across all scientific instruments and across all scientific fields, indicate the potential for new scientific developments, and raise new issues for science policy. These issues range from the skills that scientists and researchers must master to the need for a framework for data repositories that adheres to international standards for the preservation of data; sets common storage protocols and metadata; protects the integrity of the data; establishes rules for different levels of access; and defines common rules that facilitate the combining of data sets and improve interoperability (OSTP, 2010).

Diversity of scientific data

Scientific research data vary enormously – in type and volume, as well as in use and long-term value. Four types of research data are particularly important in research:

Observational data come from telescopes, satellites, sensor networks, surveys, and other instruments that record historical information or one-time phenomena (such as astronomical data from the Sloan Digital Sky Survey, SDSS). This category also includes social science research (such as demographic surveys). In many cases these data cannot be replicated and should be retained.

Experimental data may be captured from high-throughput machines (such as accelerators), through clinical trials and biomedical and pharmaceutical testing, or through other controlled experiments. Preservation is particularly important for experimental data where it is not feasible or ethical to replicate data gathering. This includes some data dealing with human subjects and endangered species (Winickof, Saha, Graff, 2009).

Computational data are generated from large-scale computational simulations. Although such data can be regenerated by rerunning the simulation, there are two reasons why computational data may need to

be preserved over the medium term (three or more years). First, the data may be used as the basis for substantive and subsequent analysis, visualisation, or data mining. Second, time on a computer for additional computations may not be available within a short time frame. This is a common occurrence for very large-scale computations that run on supercomputers shared by the research community, such as those found at the US Department of Energy national laboratories and National Science Foundation (NSF) centres.

Reference data sets are highly curated data that are often in high demand by multiple scientific communities. Such data are created for purposes that range from mapping the human genome and documenting proteins to amassing longitudinal data on economic and social status. The Worldwide Protein Data Bank and Panel Study of Income Dynamics are such reference data sets. With all these data, there is often a need to preserve ancillary materials, such as calibrations of instruments, parameters of experiments, and lab notebooks. While most large research data collections are produced and used by researchers, they are also valuable for public policy. Public policy needs go well beyond the demands of research, and become a matter of urgent public priority when it comes to having information about climate, seismology, oceanography, clinical trials and social science research surveys, endangered species, indigenous sites, archaeological sites, and sensitive security matters.

Defining open data

As with open access, there have been various attempts to elaborate definitions of open data. An interesting common theme of all these approaches is that they all stress that strong reuse rights are important for data to be open.[33] In a nutshell, open data are data that can be used by anyone without technical or legal restrictions. The use encompasses both access and reuse. Whether such openness exists from the legal perspective depends on the applicability of possible legal restrictions (or otherwise, whether the restrictions are removed by a free licence).

The Open Knowledge Foundation,[34] for example, distinguishes between: *legal openness* – that is, the possibility to legally get the data, reuse them, build on them and share them; and *technical openness* – that is, there should be no technical barriers to using the data. Generally, open data refers to data unrestrictedly available, and it should be characterised by: i) availability and access: people can obtain the data; ii) reuse and re-distribution: people can reuse and share the data; iii) universal participation: that is, anyone can use the data.

In 2004, the ministers of science and technology of OECD countries met in Paris to discuss guidelines on access to research data. The meeting was followed in 2007 by the adoption of the OECD Principles and Guidelines for Access to Research Data from Public Funding (see Box 3.2) to maximise the benefits arising from publicly funded research. The Principles and Guidelines acknowledge the importance of open

[33] This section draws largely on inputs drafted by Krzysztof Siewicz as well as Guibault, L. and T. Margoni (2014), "Legal aspects of open science and open data", Background paper for the OECD CSTP/TIP project on open science, Instituut voor Informatierecht, Universiteit van Amsterdam.

[34] See (e.g.) James, L. (2013) "Defining Open Data", http://blog.okfn.org/2013/10/03/defining-open-data/; The Open Data Institute, http://theodi.org/guides/what-open-data; Gray, J. (2010), "Launch of the Panton Principles for Open Data in Science" and "Is it Open Data?", Web Service, http://blog.okfn.org/2010/02/19/launch-of-the-panton-principles-for-open-data-in-science/.

access to research data and materials, but they also recognise the need for conformity with national legal frameworks, such as copyright laws and intellectual property protection.

Box 3.2 The OECD Principles and Guidelines for Access to Research Data from Public Funding

In 2004, ministers of science and technology of OECD countries met in Paris and discussed the need for international guidelines on access to research data. At that meeting a Declaration on Access to Research Data from Public Funding was adopted by OECD countries. Following the meeting, the OECD's Committee for Scientific and Technological Policy launched a project to develop a set of principles and guidelines. The principles and guidelines that resulted from this project were approved by the OECD's Committee for Scientific and Technology Policy in October 2006, and then endorsed by the OECD Council. The CSTP subsequently reviewed progress by countries in the implementation of the principles in 2009 [DSTI/STP(2009)3]. The Principles can be summarised as follows:

- *Openness* – Open access to research data from public funding should be easy, timely, user-friendly and preferably Internet-based.
- *Flexibility* – Flexibility requires taking into account the rapid and often unpredictable changes in ICTs, the characteristics of different research fields, and the diversity of research systems, legal frameworks and cultures of each member country.
- *Transparency* – Information on research data and data-producing organisations, documentation on the data and conditions attached to the use of data should be internationally available in a transparent way, ideally through the Internet.
- *Legal conformity* – Data access arrangements should respect the legal rights and legitimate interests of all stakeholders in the public enterprise. Restrictions on access may be for reason of: national security, privacy and confidentiality, trade secrets and intellectual property rights, protection of rare, threatened or endangered species, legal processes.
- *Protection of intellectual property* – Data access arrangements should consider the applicability of copyright and other intellectual property laws that may be relevant to publicly funded research databases (as in the case of public-private partnerships).
- *Formal responsibility* – Access arrangements should promote the development of rules and regulations regarding the responsibilities of the various parties involved; should be developed in consultation with representatives of all affected parties; should be responsive to factors such as the characteristics of the data, e.g. their potential value for research purpose. Data management plans and long-term sustainability should also be considered.
- *Professionalism* – Institutional arrangements for the management of research data should be based on the relevant professional standards and values embodied in the codes of conduct of the scientific communities involved.
- *Interoperability* – Access arrangements should consider the relevant international data documentation standards.
- *Quality* – The value and utility of data depend to a large extent on the quality of the data themselves. Particular attention should be paid to ensuring compliance with explicit quality standards.
- *Security* – Attention should be devoted to supporting the use of techniques and instruments to guarantee the integrity and security of research data.
- *Efficiency* – One of the central goals of promoting data access and sharing is to improve efficiency of publicly funded scientific research so as to avoid expensive and unnecessary duplication of effort. This also involves cost and benefit analysis to define data retention protocols; the engagement of data management specialist organisations; and the development of new reward structures for researchers and database producers.
- *Accountability* – The performance of data access arrangements should be subjected to periodic evaluation by user groups, responsible institutions and research funding agencies.
- *Sustainability* – Due consideration should be given to the sustainability of access to publicly funded research data as a key element of the research infrastructure.

Source: OECD (2007), OECD Principles and Guidelines for Access to Research Data from Public Funding.

Data sharing: challenges and opportunities

Research data, and data more in general, are in most cases intangible assets involving different actors and stakeholders along the different phases of data creation, compilation and reuse. Several actors can claim ownership of the same data sets. Data and especially research data can be created or collected by an individual, then used by another party, and subsequently the information can be compiled and clean by others. A third party funding or commissioning some or all of the activities above can claim ownership of the data (Loshin, 2002). To add an additional layer of complexity, in some fields of science many of these tasks can be performed by machines or automated action.

An essential element for the usefulness of data sharing efforts is the quality of the publicly released data. The OECD (2011) Quality Framework and Guidelines for OECD Statistical Activities identifies seven key aspects of data quality (see Box 3.3). The ODE project (Opportunity for Data Exchange, a European Commission-funded research project) developed a "Data Pyramid" (Figure 3.2) that allows visualisation of the different phases of data curation: from raw data contained in files in personal computers to processed data linked to publications containing the detailed metadata or information around the data itself.

Box 3.3 The OECD Quality Framework and Guidelines for OECD Statistical Activities

The OECD Quality Framework and Guidelines for OECD Statistical Activities identifies the following dimensions:

1. *Relevance* – "is characterised by the degree to which the data serves to address the purposes for which they are sought by users. It depends upon both the coverage of the required topics and the use of appropriate concepts".

2. *Accuracy* – is "the degree to which the data correctly estimate or describe the quantities or characteristics they are designed to measure".

3. *Credibility* – "The credibility of data products refers to the confidence that users place in those products based simply on their image of the data producer, i.e. the brand image. Confidence by users is built over time. One important aspect is trust in the objectivity of the data".

4. *Timeliness* – "reflects the length of time between their availability and the event or phenomenon they describe, but considered in the context of the time period that permits the information to be of value and still acted upon.

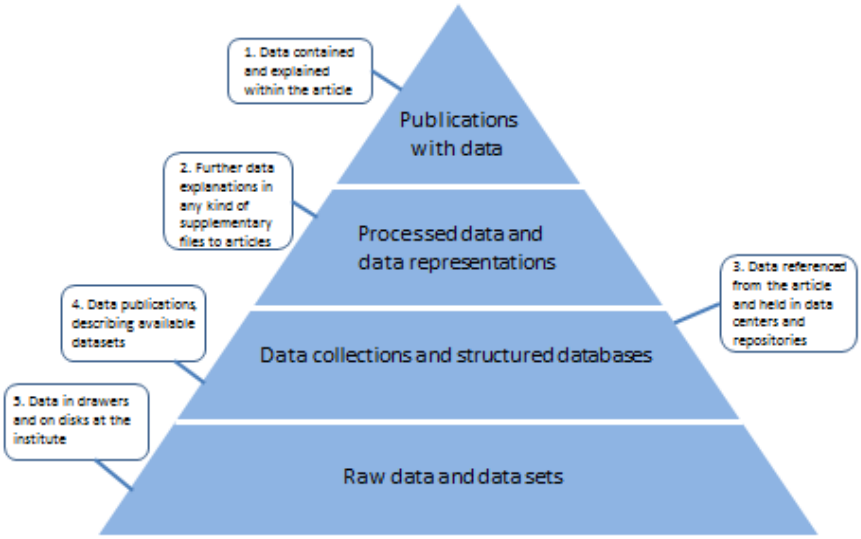
5. *Accessibility* – "reflects how readily the data can be located and accessed".

6. *Interpretability* – "reflects the ease with which the user may understand and properly use and analyse the data". The availability of metadata plays an important role here, as they provide for example "the definitions of concepts, target populations, variables and terminology, underlying the data, and information describing the limitations of the data, if any".

7. *Coherence* – "reflects the degree to which they are logically connected and mutually consistent. Coherence implies that the same term should not be used without explanation for different concepts or data items; that different terms should not be used without explanation for the same concept or data item; and that variations in methodology that might affect data values should not be made without explanation. Coherence in its loosest sense implies the data are 'at least reconcilable'".

Source: OECD (2011), "Quality Framework and Guidelines for OECD Statistical Activities", 17 January, http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=std/qfs%282011%291.

Figure 3.2 Data Pyramid



Note: This work is licensed under the Creative Commons Attribution 3.0 Unported Licence. To view a copy of this licence, visit http://creativecommons.org/licenses/by/3.0/ .The colours of the figure have been altered for editing purposes.

Source: Reilly, S. et al. (2011), Report on Integration of Data and Publications, ODE, Opportunities for Data Exchange, available at: http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/11/ODE-ReportOnIntegrationOfDataAndPublications-1_1.pdf.

In many scientific communities there is as yet no standard data quality assessment protocol as it exists for scientific publications (Brase et al., 2009). As highlighted for example in Royal Society, 2012 and Dallmeier-Tiessen et al., 2011, data have little value if their quality does not meet minimum quality criteria. “Good quality data” implies being not only accessible (for example available on the Internet), but also intelligible, assessable, trustworthy and, of course, reusable. In this respect the development of detailed data-sharing information and metadata is essential for the further use of the same data from multiple teams of researchers.

However, scientists and researchers do not necessarily have the incentives or the skills to perform these tasks, since proper curation and dissemination of data sets is costly and time-consuming and can be even considered as another type of scientific output (Uhlir, 2012). A possible solution to disincentives is data citation: the possibility for researchers to be acknowledged for their work of data collection and curation through mechanisms similar to the one already in place for citations of academic articles (Mooney and Newton, 2012; CODATA-ICSTI, 2013). Data citation however is not necessarily a standardised or widely accepted concept in the academic community. Some scientists see it as being limited to citation to scientific articles; funding agencies in some cases question the idea of recognising individuals as data authors; and traditional bibliometric indicators are not yet taking into account non-article citations (Costas et al., 2013). In addition, there are technical barriers restricting the development of data citation and related metrics: these include incompatibility in machines and software, data file structures, data storage and management (Groves, 2010). A number of organisations are actively engaged in overcoming these barriers (see Box 3.4).

According to the EU FP7 research project ODE (Kotarski et al., 2012), data citation has some unique features, owing to the particular properties of data sets. For instance, data sets may be of very different sizes and it is not always clear which specific elements inside the data sets scholars are referring to – or in the case of updates to the data sets, which version to cite. According to the ODE project, some of the good practices/challenges related to data citations are:

- Citation of the data set, with identifier, should be listed in the reference/bibliography to enable to track and develop citation metrics.
- Publishers need to provide guidance for authors and referees on data citation.
- There is no clear agreement on the accuracy and longevity requirements for data sets to be considered citable or cited.
- There is lack of clarity and agreement on what authorship of a data set means.
- Researchers need to promote awareness in their communities of the benefit of data citation, and follow agreed data citation guidelines.

Other possible vehicles to publish data sets are *data journals*, that is collections of scientific articles specialised in publishing data papers. Data papers are articles with the primary purpose of describing data sets rather than reporting scientific investigation and analysis. Data papers contain fact and descriptions about data. Their goal is to be a citable source of information on data that brings credit to the scholars who produced and described the database; to disclose detailed information on a data set; and to bring the existence of the data to the attention of the scientific community (Chavan and Penev, 2011). Data journals may target broader scientific areas as well as specific domains, such as earth system science (geoscience).

Box 3.4 Organisations involved in promoting data citation

DataCite (www.datacite.org) is an international non-profit organisation. Established in London in 2009, DataCite aims to promote access to research data through the Internet; to support open data archiving; and to allow the verification of scientific results and the reuse of data for further studies. To facilitate data release, DataCite helps researchers with the unique identification and attribution of data sets for citation purposes, and supports journal publishers in establishing linkages between published articles and data sets. In addition, the organisation supports data centres by providing identifiers for data sets and defining workflows and standards for data publication.

ORCID (www.orcid.org) is a non-profit, community-driven organisation that aims to create and maintain a registry of unique researcher identifiers and to link research activities and outputs on the basis of these identifiers. Researcher identifiers can be linked not only to scientific articles but also to other forms of research outputs, including equipment, experiments, patents and data sets. ORCID provides two main functions: i) a registry to obtain a unique identifier for researchers and to manage activities; ii) application programme interfaces that support system-to-system communication and authentication. ORCID codes are available by means of open source licence.

Figshare (www.figshare.com) is an online digital repository of research data (digital data, figures, images, videos, etc.). The company allows data citation and has partnered with ORCID. (See Box 4.9 for more information on Figshare.)

The *Dryad Digital Repository* (www.datadryad.org) is an online repository containing the data underlying scientific publications. Data in Dryad are assigned a unique object identifier (DOI) to allow data citations. Dryad is governed by a non-profit membership organisation. Membership is open to stakeholder organisations such as journals, scientific societies, publishers, research institutions and funding organisations.

ResearcherID (www.researcherID.org) allows unique identification of scientific authors; it was created in 2008 by Thomson Reuters. Researchers can log into the system and link their "researcherID" to their own articles to correct

identification or spelling mistakes. ResearcherID has partnered with ORCID to enable data citations.

Source : www.datacite.org, www.orcid.org, www.figshare.com, www.datadryad.org, www.researcherID.com.

Although open science has positive effects on the scientific enterprise itself, innovation, and society more generally, there exist a number of legitimate reasons to limit the openness of science, especially around data, that go beyond technical issues and involve not only the research community but also society more generally. These include for instance issues related to the privacy of individuals or organisations, or national security (OECD 2015). Data gathered in the course of research often contain personal information (e.g. medical records), so that in opening such data the rights of data subjects must be respected (Lane et al., 2014). This does not mean that the data cannot be opened, but it does call for implementing protective procedures (see Box 3.5). One such procedure is anonymisation, which may however lead to the inapplicability of the whole personal data protection regime. Additionally, not all anonymisation techniques are effective (Narayanan and Shmatikov 2008). Some countries are now promoting "privacy by design" for their health data, through the use of privacy-enhancing technologies to meet both health care data use and privacy protection needs (OECD, 2013a).

Box 3.5 Practical means for preventing information discovery

Data analytics extract information from data by revealing the context in which the data are embedded, and their organisation and structure. There exist a number of practical means for preventing or significantly increasing the cost of extracting the information embedded in the data through data analytics, though they may adversely affect data utility. Examples follow.

Data reduction – Data reduction can be considered the strongest means for preventing information extraction, because where no data are collected, no information can be extracted. Data subjects can withhold or decline to provide data. Data controllers can practice data minimisation. As Pfitzmann and Hansen (2010) have highlighted, data minimisation "is the only generic strategy (misinformation or disinformation aside) to enable unlinkability, since all correct personal data provide some linkability".

Cryptography – Cryptography is a practice that "embodies principles, means, and methods for the transformation of data in order to hide its information content, establish its authenticity, prevent its undetected modification, prevent its repudiation, and/or prevent its unauthorised use" (OECD, 1997). It is a key technological means of providing security for data in information and communications systems. Cryptography can be used to protect the confidentiality of data, such as financial or personal data, whether that data are in storage or in transit. Cryptography can also be used to verify the integrity of data by revealing whether data have been altered and identifying the person or device that sent them.

De-identification – covers a range of practices ranging from anonymisation to pseudonymisation. These practices share a common aim of preventing the extraction of identifying attributes (i.e. re-identification), or at least significantly increasing the costs of re-identification. *Anonymisation* is a process in which an entity's identifying information is excluded or masked so that the entity's identity cannot be, or becomes too costly to be, reconstructed (Pfitzmann and Hansen, 2010; Mivule, 2013). Some research suggests that when linked with other data, most anonymised data can be de-anonymised; that is, the identifying information can be reconstructed (Narayanan and Shmatikov, 2007; Ohm, 2009). Many applications, however, require some kind of identifier, and having complete anonymity would prevent any useful two-way communication or transaction. *Pseudonymisation* is therefore used, whereby the most identifying attributes (*i.e.* identifiers) within a data record are replaced by unique artificial identifiers (i.e. pseudonyms).

Unlinkability and functional separation – Unlinkability results from processes to ensure that data processors cannot distinguish whether items of interest are related or not (Pfitzmann and Hansen, 2010). According to ISO (Pfitzmann and Hansen, 2010), unlinkability "ensures that a user may make multiple uses of resources or services without others being able to link these uses together". De-identification is a means to enable unlinkability, but cannot guarantee unlinkability. Other technical means include functional separation and distribution (decentralisation).

Noise addition and disinformation – The addition of "noise" to a data set can allow analysis based on the complete data set to remain significant while masking sensitive data attributes. Finding the right balance that protects privacy while minimising the costs to data utility is a challenge (Mivule, 2013). Disinformation is false or inaccurate information spread intentionally to mislead. Noise addition techniques are considered promising for helping protect privacy and confidentiality in databases, while keeping all data sets statistically close to the originals**.** Work on "Differential Privacy" is one example (Dwork and Roth, 2014).

Source: OECD (2015), *Data-Driven Innovation for Growth and Well-Being*, OECD Publishing, Paris.

The OECD Global Science Forum has identified a series of challenges associated with data-driven science and data sharing (Box 3.6). According to the Forum, the barriers may relate to legal or technical issues or the lack of skills, both within the scientific community and outside, to perform data management and data-sharing tasks.

Box 3.6 The nine challenges identified by the OECD Global Science Forum

The OECD Global Science Forum has recently identified a number of challenges related to data-driven and evidence-based research.

Challenge 1 – Massive amounts of digital data are being generated at an unprecedented scale, thanks partly to the advent of ICTs. The reliability, statistical validity and generalisability of new forms of data are not yet fully understood.

Challenge 2 – While administrative, survey and census data are widely collected by national statistical agencies and government departments, micro-data records are available to a much lesser extent.

Challenge 3 – New forms of personal data, such as social networking data, are increasingly created and collected. The use of those data may generate risks to individual privacy.

Challenge 4 – Legal, cultural, language and proprietary rights of access barriers hinder cross-national collaboration and international data exploitation, especially in the social sciences.

Challenge 5 – Global research agendas require increasingly interdisciplinary and international co-ordination.

Challenge 6 – Collaboration and experience sharing across countries in the development of comparable data resources is necessary to fully exploit the potential of data sets.

Challenge 7 – Researchers often lack the resources or the skills to make sure that the data they use, gather and produce are available for reuse.

Challenge 8 – National investments in skills and infrastructure related to data creation and curation are essential to avoid risk of data loss or degradation.

Challenge 9 – Researchers need to have the right set of incentives to ensure effective data sharing.

Source: Adapted from OECD (2013b), "New data for understanding the human condition – International perspectives", OECD Global Science Forum Report on Data and Research Infrastructure for the Social Sciences.

For all these reasons (namely the challenges around the ownership of data sets, privacy, confidentiality and security issues, the different degrees of quality of data sets, the lack of skills and incentives in the researchers' community**,** etc.), promoting open data is certainly less straightforward than promoting open access to scientific publication. The level of maturity of policy initiatives in OECD countries and beyond reflects this challenge: although the policy landscape is evolving rapidly, many more policies and initiatives have been recently developed to promote open access than to promote open data.

Data protection frameworks in OECD countries[35]

The expansion of open access policies to publicly funded research data raises a number of legal and policy issues that are often distinct from those concerning the publication of scientific articles and monographs. Since open access of research data – unlike publications – is a relatively new policy objective, less attention has been paid to the specific features of research data.

Internationally, the protection afforded to databases (as collections of data or other elements) is established – or confirmed – by both Art. 10(2) of the TRIPS Agreements and in the almost identical Art. 5 of WCT:

> "Compilations of data or other material, whether in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creation shall be protected as such..." (TRIPS Agreements, Art. 10[2])

Whereas scientific publications virtually always attract copyright protection under the copyright laws of OECD member countries, the individual research data and the data sets containing them may not so easily fall under the copyright regime. A number of OECD countries as well as the European Union, however, adopted legal frameworks that contain implications for research data. In other countries, notably the United States, data sets have no special IP protection.

Databases represent a particular subject matter that is protected by copyright under certain circumstances, but in some areas – for example in the European Union, Japan, Korea – is also protected by a so-called *sui generis* database right (SGDR). This additional layer of protection is found in some countries and is afforded to databases regardless of the intellectual creation (i.e. "selection or arrangement") that may or may not be present. What is protected instead is the investment in making the database, i.e. in the obtaining, verification or presentation of the data. This type of right is typical of the EU Database Directive and of the laws of a number of other countries, and will be dealt with below. It should be borne in mind that while the protection afforded to original databases focuses on the arrangement or selection without extending to the content of the database, the SGDR offers a protection against the copy of substantial parts of the database – that is to say it extends, at least to some extent, to the data themselves.

The complexity of the rights status of research data in Europe and other jurisdictions arguably has the potential to adversely affect the reuse opportunities of collections of scientific data, given the difficulty – both for research institutions making the database available and for prospective re-users – in determining each time whether a certain database is covered by a *sui generis* right and in which measure re-utilisation and extraction can take place freely. Whether the use of compilations or databases for purposes of research and private study in general, and text and data mining in particular, is covered by any relevant exception to copyright or to the database right is uncertain. The use of Creative Commons licences 4.0 (Boxes 2.3 and 2.4) may alleviate the uncertainty, by clearly stating what can and cannot be done with the licensed material.

[35] This section is largely based on Guibault L. and T. Margoni (2014), Legal aspects of open science and open data, Background paper for the OECD CSTP/TIP project on open science, Instituut voor Informatierecht, Universiteit van Amsterdam.

The legal framework in the European Union

In Europe, to be eligible for copyright protection, collections of data, tables and compilations must show a sufficient degree of originality in their selection or arrangement – that is to say, through this selection or arrangement the author was able to express their free and creative choices (Synodinou, 2012).

Whether collections of scientific research data meet the criterion of originality is a question of fact, to be determined on a case-by-case basis. However, if the selection or arrangement of the contents of a scientific database is dictated by technical factors or imperatives of accuracy and exhaustiveness, then the author can exercise little to no creativity or discretion in the choice, sequence or combination of data in the collection. Scientific databases are therefore in most cases not likely to meet the threshold for copyright protection.

The Information Society Directive contains an exception on copyright that might be applicable in some cases. Article 5(3)(a) of this Directive allows Member States to provide for exceptions in the case of "use for the sole purpose of illustration for teaching or scientific research, as long as the source, including the author's name, is indicated, unless this turns out to be impossible, and to the extent justified by the non-commercial purpose to be achieved". This exception is optional; Member States may decide whether to implement it or not. As a result, Member States have different rules and regulations in this context, and some countries recognise no research exception at all (such as the Netherlands and Spain). As a result, the research exception is generally vague and unevenly implemented at national level, which may put some researchers at a disadvantage (Triaille 2013).

Collections of scientific works, data, or other materials arranged in a systematic or methodical way and individually accessible, electronically or by other means, may be protected under the European *sui generis* database right (SGDR). Through Article 7 of the Database Directive, as implemented in the legislation of the Member States, the maker of a database showing a substantial investment (assessed qualitatively and/or quantitatively) in either the obtaining, verification or presentation of its contents has the exclusive right to prevent the extraction and/or re-utilisation of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database. Like copyright protection, the *sui generis* database right arises automatically, without any formal requirement, at the time of completion of the database or its disclosure to the public.

Where the "obtaining, verification or presentation" of research data sets does manifest the substantial investment necessary to qualify for protection, *sui generis* confers two transferable rights on the maker of a database: the right of extraction and the right of re-utilisation of substantial parts of the database, which are respectively defined as follows: "(a) 'extraction' shall mean the permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form; (b) 're-utilization' shall mean any form of making available to the public all or a substantial part of the contents of a database by the distribution of copies, by renting, by on-line or other forms of transmission".

The protection under the *sui generis* right lasts for 15 years from the first of January of the year following the date on which the database was completed. The term of protection for a database may start anew under two conditions, both dealing with the term "substantial". The first one is a substantial modification of the contents of the database, evaluated either qualitatively or quantitatively, which can consist of additions, deletions or alterations (including rearrangement of the contents). Secondly, this

substantial modification must represent a substantial investment, evaluated qualitatively or quantitatively. This is one of the most controversial provisions of the Directive since, according to some, it apparently offers grounds for a perpetual protection of the databases (Reichman and Okediji, 2012).

The Netherlands

The Netherlands is so far the only Member State to have explicitly regulated exercise of the *sui generis* rights by public sector bodies. Article 8 of the Dutch Database Act denies a public authority the right to exercise its exclusive database rights unless the right is reserved explicitly by a general mention in an act, order or ordinance, or in a specific case by notification on the database itself or while the database is made available to the public.

Japan

Database-related provisions were introduced in Japanese Copyright Law for the first time in 1986. The Japanese legislator considered that separate protection from that afforded to compilations under copyright law should be afforded to (electronic) databases, and decided to introduce provisions specifically drafted for electronic databases into the Copyright Law. Based on this distinction between compilations and databases, it was thought that databases that should be protected under the new provisions were computer-searchable databases. At the same time, because creativity of their data arrangement does not need to be protected, electronic databases are excluded from the definition of "compilations." Art.2(1)(xter) of the Law defines the term "database" as "an aggregate of information such as articles, numerals or diagrams, which is systematically constructed so that such information can be searched for with the aid of a computer".

Art. 30(4) of the Japanese Copyright Act 1970,[36] introduced in 2012,allows a publicly disclosed work to be used as needed for the development of technology and in experiments to test audio or visual recording devices.[37] Another amendment, of 2009,[38] introduced – alongside other limitations – an exception aimed at boosting the Japanese Internet economy (Tamura, 2009), an exception specifically designed to permit TDM. Article 47 *septies* of the Copyright Act[39] contains an explicit provision to allow text mining:

> "For the purpose of information analysis ('information analysis' means to extract information, concerned with languages, sounds, images or other elements constituting such information, from many works or other such information, and to make a comparison, a classification or other statistical analysis of such information; the same shall apply hereinafter in this Article) by using a computer, it shall be permissible to make recording on a memory, or to make adaptation (including a recording of a derivative work created by such adaptation), of a work, to the extent deemed necessary. However, an exception is made of database works which that are made for the use by a person who makes an information analysis."

---

[36] See Copyright Act Law N. 48 of 1970.

[37] See Art. 30(4) introduced by Law No. 43 2012, cited by T. Doi in Geller and Bently, cit., at 1[2][c].

[38] See Law No. 53 of 2009.

[39] Japan Copyright Act: http://www.cric.or.jp/english/clj/cl2.html.

A report issued by the Subdivision on Copyright of the Council for Cultural Affairs in January 2009 presents the following examples of information analysis: 1) website information analysis and language analysis, in which the use of a specific language or character string is analysed and statistically processed, and 2) sound analysis and video/image analysis, in which the meaning of the sound wave, video, character string, etc., comprising a certain sound, video, image, etc., is analysed. Although the types of works subject to this provision are not limited, the reverse engineering[40] of computer programming falls outside the scope of this exception: reverse engineering cannot be regarded as "information analysis" because no statistical analysis is conducted.

Recently, Japan has seen the introduction of new services that enable users to search and analyse other users' comments on the Internet, including blogs, review sites and social media. The establishment of said Article 47 is one of the factors that promoted the emergence of those new services (Iida et al., 2011).

Korea

In Korea, the Copyright Act in Chapter IV protects compilations as original works of authorship, if they are creative in selection, arrangement or composition of their contents. Databases are defined as compilations of which their contents are arranged or composed in a way that anyone can individually access or search such contents.[41] Thus original databases are protected by copyright, as they meet such conditions. Non-original databases fall under statutory subject matter of protection in accordance with the new Chapter IV introduced in 2003. The database producer who makes a considerable investment in human or material resources for the production of a database, or renewal, verification, or supplementation of their contents has the rights of reproduction, distribution, broadcasting or interactive transmission.[42] A foreign national can be a beneficiary of sui generis protection on the condition of reciprocity, if he or she is protected in accordance with a treaty to which the Republic of Korea has acceded.[43] The rights of the database producer are limited more broadly than copyrights, as the limitations and exceptions to copyright are applicable mutatis mutandis to the rights of the database producer on the one hand, and the use of the whole or substantial portion of the database is permissible for educational, academic or research purposes, or for reporting current events on the other hand.[44] The term of protection is renewable 5 years.[45]

Unsolved legal issues: public-private partnerships and text and data mining

The core of open access principles aims to make research publications and data available for reuse by any user. In some cases that may create challenges – notably, if research results are not funded entirely by public money. This raises a series of questions, such as whether the funding of research by public-private

[40] Reverse engineering is the process of extracting information from taking apart and object and understand how it works to duplicate or improve that object. This practice, traditionally used in traditional industries, is not frequently used in computer hardware and software.

[41] See Korea Copyright Act Art. 2 Items 18 and 19.

[42] See Korea Copyright Act Art. 2 Item 20 and Art. 93.

[43] See Korea Copyright Act Art. 91.

[44] See Korea Copyright Act Art. 94.

[45] See Korea Copyright Act Art. 95.

partnerships affects ownership of the data and the licensing conditions applied, or how text and data mining (TDM) is affected by copyright law.

Public-private partnerships

Ownership issues may be at stake in cases of public-private partnerships. The funding of a research project through external sources, whether public or private, usually leads to the application of different rules of ownership. Generally at least three parties are involved: the author, the university or the research organisation, and the sponsoring or commissioning party. Depending on the law, the internal policy of the institution or the bargaining position of the respective parties, the copyright ownership may be transferred either to the university or to the external entity.

The issue of the ownership of rights is particularly important in the context of public-private partnerships because it can greatly influence the manner in which research output will be disseminated. The private party will typically attempt to protect their commercial interests. This means that, depending on the option chosen, some restrictions may be placed on the distribution and reuse of the publications and data, for example by limiting commercial reuse.

Although the possibility exists under the Creative Commons licensing system to restrict use for commercial purposes, the distinction between commercial and non-commercial use in the Creative Commons licences raises questions not only in the scientific publishing sector, but also in several other sectors of the copyright industry, as it may leave too much room for interpretation.

Text and data mining (TDM)

Text and data mining (see Box 1.2) is a popular technique used in science and other disciplines to analyse and extract new insights and knowledge from the exponentially increasing store of digital data. TDM is likely to become more important as researchers acquire the skills and the technology to address and investigate data sets of increasing size, complexity and diversity in all media: text, numbers, images, audio files and all other forms. However, current legal frameworks regarding the scope of protection of works and databases can potentially create obstacles to the TDM activities for research purposes.

Currently, under copyright law, database protection law in the EU and specific provisions in intellectual property law, scientific publishers can claim a right to grant or refuse the mining of their works. However, while publishers tend to block software that automatically mines text and data, they allow TDM through the licenses they offer subscribers to journals. Still, some proponents of open access consider that a system resting solely on licences might be insufficient to allow TDM on a large scale. In some cases, transaction costs may be too high for parties to negotiate a licence. Moreover, transaction costs may rise if researchers have to reconcile the terms and conditions of non-standard or non-interoperable licenses.

The publishing industry, and with a view to promoting self-regulation with regard to TDM, has responded by easing access to TDM permissions; the licenses that publishers' offer with institutional subscriptions increasingly include provisions for TDM access. Publishers are also developing automated services, such as "Cross Ref TDM" that allows researchers to access material for TDM with no requirement to make a direct request to publishers.

It is against this background of a tension between open access to research articles and data, TDM technologies and intellectual property rights that the United Kingdom has recently introduced a specific exception in the Copyright, Designs and Patent Act to allow TDM activities for non-commercial research to take place without the rights holder's prior authorisation under the conditions stated in the law. Japan also has an exemption, albeit, a more narrow one, in its copyright law for the purpose of TDM (see above discussion on Japan). At this point it is unclear whether other countries will follow suit with similar exemptions or whether the newer licensing models proposed by a self-regulating publishing industry will be sufficient to allow TDM for research purposes at a breadth and scale necessary for a scientific system that is increasingly data-driven. Much will likely depend on the evidence of TDM use for research purposes.

REFERENCES

Anderson, C. (2008), "The end of theory: The data deluge makes the scientific method obsolete", *Wired*, 23 June, www.wired.com/science/discoveries/magazine/16-07/pb_theory/.

Anderson, W.L. (2004), "Some challenges and issues in managing, and preserving access to, long-lived collections of digital scientific and technical data", *Data Science Journal*, Vol. 3, 30 December, pp. 191-202, retrieved from www.jstage.jst.go.jp/article/dsj/3/0/191/_pdf.

Bell, G., T. Hey and A. Szalay (2009), "Beyond the data deluge", *Science*, Vol. 323, No. 5919, 6 March, pp. 1297-1298, retrieved from www.cloudinnovation.com.au/Bell_Hey_Szalay_Science_March_2009.pdf.

BIAC (2011), "BIAC thought starter: A strategic vision for OECD work on science, technology and industry", Business and Industry Advisory Committee to the OECD, 12 October.

Blue Ribbon Task Force on Sustainable Digital Preservation and Access (2010), "Sustainable economics for a digital planet: Ensuring long term access to digital information", February, http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf, accessed 17 June 2015.

Bollier, D. (2010), *The Promise and Peril of Big Data*, The Aspen Institute, Washington, DC.

Brase, J. et al. (2009), "Approach for a joint global registration agency for research data", *Information Services and Use*, Vol. 29, No. 1, pp. 13-27.

CODATA-ICSTI (Committee on Data for Science and Technology - International Council for Scientific and Technical Information) Task Group on Data Citation Standards and Practices (2013), "Out of cite, out of mind: The current state of practice, policy and technology for the citation of data", *Data Science Journal*, Volume 12, pp. 1-75.

Costas, R. et al. (2013), "The Value of Research Data – Metrics for datasets from a cultural and technical point of view", A Knowledge Exchange Report, www.knowledge-exchange.info/datametrics, accessed 11 June 2015.

Dallmeier-Tiessen, S. et al. (2011), "Highlights from the SOAP project survey: What scientists think about open access publishing", http://arxiv.org/abs/1101.5260, accessed 10 June 2015.

Dwork, C. and A. Roth (2014), "The algorithmic foundations of differential privacy", *Foundations and Trends in Theoretical Computer Science*, Vol. 9, Nos. 2-4, pp. 211-407, http://dx.doi.org/10.1561/0400000042.

EC (2010), "Riding the wave: How Europe can gain from the rising tide of scientific data", Final report by the High-level Expert Group on Scientific Data, European Commission, October, http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf, accessed 10 June 2015.

Frischmann, B.M. (2012), *Infrastructure: The Social Value of Shared Resources*, Oxford University Press.

Geist, M. (2013), "Fairness found: How Canada quietly shifted from fair dealing to fair use", in M. Geist (ed.), *The Copyright Pentalogy*, University of Ottawa Press, Ottawa, pp. 157-186.

Groves, T. (2010), "The wider concept of data sharing: View from the BMJ", *Biostatistics*, Vol. 11, No. 3, pp. 391-92.

Guibault, L. and T. Margoni (2014), "Legal aspects of open science and open data", Background paper for the OECD CSTP/TIP project on open science, Instituut voor Informatierecht, University of Amsterdam.

Hey, T. and A. Trefethen (2003), "The data deluge: An e-science perspective", in F. Berman, G.C. Fox and A.J.G. Hey (eds.), *Grid Computing: Making the Global Infrastructure a Reality*, John Wiley & Sons, Ltd., Chichester, England, pp. 809-24, retrieved from http://eprints.ecs.soton.ac.uk/7648/1/The_Data_Deluge.pdf.

Iida, K. et al. (2011), "Question Q216B exceptions to copyright protection and the permitted uses of copyright works in the hi-tech and digital sectors", the Japanese Group of AIPPI (Association Internationale pour la Protection de la propriété Industrielle), p. 9.

Jirotka, M. et al. (2006), "Collaboration in e-research", *Computer Supported Cooperative Work (CSCW)*, Special Issue, Vol. 15, No. 4, pp. 251-55, http://dx.doi.org/10.1007/s10606-006-9028-x.

Kotarski, R. et al. (2012), "Report on best practices for citability of data and on evolving roles in scholarly communication", *Opportunities for Data Exchange*, www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/08/ODE-ReportBestPracticesCitabilityDataEvolvingRolesScholarlyCommunication.pdf.

Lane, J. et al. (eds.) (2014), *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, Cambridge University Press.

Loshin, D. (2002), "Knowledge integrity: Data ownership", 8 June, www.datawarehouse.com/article/?articleid=3052.

Mivule, K. (2013), "Utilizing noise addition for data privacy: An overview", *Proceedings of the International Conference on Information and Knowledge Engineering (IKE 2012)*, pp. 65-71, http://arxiv.org/pdf/1309.3958.pdf, accessed 17 June 2015.

Mooney, H. and M.P. Newton (2012), "The anatomy of a data citation: Discovery, reuse, and credit", *Journal of Librarianship and Scholarly Communication*, Vol. 1, No. 1, Art. No. eP1035, http://dx.doi.org/10.7710/2162-3309.1035, accessed 10 June 2015.

Narayanan, A. and V. Shmatikov (2008), "Robust de-anonymization of large sparse datasets", *SP 08 Proceedings of the 2008 IEEE Symposium on Security and Privacy*, IEEE Computer Society, Washington, DC.

OECD (2015), *Data-Driven Innovation for Growth and Well-Being*, OECD Publishing, Paris.

OECD (2014), *Measuring the Digital Economy: A New Perspective*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264221796-en.

OECD (2013a), "Strengthening health information infrastructure for health care quality governance", in *OECD Health Policy Studies Series 2013*, OECD Publishing, Paris.

OECD (2013b), "New data for understanding the human condition: International Perspectives", OECD Global Science Forum Report on Data and Research Infrastructure for the Social Sciences, OECD Publishing, Paris, www.oecd.org/sti/sci-tech/new-data-for-understanding-the-human-condition.pdf, accessed 17 June 2015.

OECD (2011), "Quality framework and guidelines for OECD statistical activities", OECD, Paris, 17 January, http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=std/qfs%282011%291.

OECD (2007), *OECD Principles and Guidelines for Access to Research Data from Public Funding*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264034020-en-fr.

OECD (1997), *Recommendation of the Council concerning Guidelines for Cryptography Policy*, 27 March, C(97)62/FINAL, OECD, Paris, www.oecd.org/internet/ieconomy/guidelinesforcryptographypolicy.htm.

Ohm, P. (2009), "The rise and fall of invasive ISP surveillance", *University of Illinois Law Review*, 2009 Volume, No. 5, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1261344, accessed 17 June 2015.

Pfitzmann, A. and M. Hansen (2010), "A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management", 10 August, http://dud.inf.tu-dresden.de/Anon_Terminology.shtml.

Reichman, J.H. and R.L. Okediji (2012), "When copyright law and science collide: Empowering digitally integrated research methods on a global scale", *Minnesota Law Review*, Vol. 96, p. 1362-1480.

Reilly, S. et al. (2011), "Report on integration of data and publications", ODE – Opportunities for Data Exchange, www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/11/ODE-ReportOnIntegrationOfDataAndPublications-1_1.pdf.

Royal Society (2012), "Final report: Science as an open enterprise", Royal Society Science Policy Centre Report 02/12, https://royalsociety.org/policy/projects/science-public-enterprise/Report/, accessed 11 June 2015.

Synodinou, T.-E. (2012), "The foundations of the concept of work in European copyright law", in T.-E. Synodinou (ed.), *Codification of European Copyright Law – Challenges and Perspectives*, Kluwer Law International, The Hague,

Tamura, Y. (2009), "Rethinking Copyright Institution for the Digital Age", *WIPO Journal: Analysis and Debate of Intellectual Property Issues* 1, No. 1, pp. 63-74.

The Economist (2010), "Data, data everywhere", 27 February.

Triaille, J. (ed.) (2013), "Study on the application of Directive 2001/29/EC on copyright and related rights in the information society", De Wolf & Partners in collaboration with CRIDS. pp

Uhlir, P.F. (2012), *For Attribution – Developing Data Attribution and Citation Practices and Standards*, Summary of an International Workshop 21.

Winickoff, D.E., K. Saha and G.D. Graff (2009), "Opening stem cell research and development: A policy proposal for the management of data, intellectual property and ethics", *Yale Journal of Health Policy, Law and Ethics*.

Chapter Four

THE GOVERNANCE OF OPEN SCIENCE: ACTORS, TRENDS AND POLICIES

The key actors

Several actors of local, national and global innovation systems are involved in open science efforts. Researchers are the individuals implementing open science efforts in practice. Government ministries and research funding agencies, universities and public research organisations directly contribute to open science by defining and implementing policies and programmes, and by producing and disseminating scientific outcomes. Private scholarly publishers contribute to open science initiatives by offering new and more comprehensive services to different actors. In addition, business sector actors are affected by open science initiatives, as firms may benefit from the further dissemination of public research results as well as deliver open science-related services. Finally, supra-national entities such as the OECD, the European Union and UNESCO may play a major role in the definition of international co-ordination agreements or guidelines to address issues related to open science with an international and global perspective. The following sections provide an overview of the key actors involved in open science.

The researcher community

Researchers themselves have been at the forefront of efforts to promote open science. There are several motivations for researchers, ranging from the cultural values inherent in science (e.g. openness to scrutiny, willingness to engage society) to necessity (i.e. developing a technological infrastructure to allow for collaboration). As described in the rest of this chapter, researchers also respond to incentives from funding agencies, universities and public research institutes. They are a key actor since they are the individuals that can implement open science initiatives and make open science happen. Researchers have also brought a key contribution in advancing the knowledge and the understanding of open science, through research projects and publications. Tension may nevertheless exist between the competitive "publish or perish" paradigm and the interest in sharing data and collaborating (see section on incentive mechanisms). The role of researchers is further detailed and analysed in all sections of this chapter.

Ministries and governmental bodies

Open science efforts are often promoted by initiatives led by governments or ministries in charge of science and innovation policy, such as ministries of education and research or ministries of economic development. In several OECD countries, open science efforts are part of national innovation strategies or open government agendas. These agendas help define national-level strategic priorities that can be translated into concrete initiatives by other innovation system actors (see those mentioned below). Although several actors in the innovation system are free to develop individual open science strategies, initiatives or guidelines (such as universities or research centres, or sub-national-level authorities), actions developed at national level often contribute to steering and co-ordinating the system. Co-ordinated national-level actions typically involve large-scale investments in the infrastructure and skills necessary to promote open science efforts. Other areas where action at the central level is generally needed are the

definition of regulations and incentive frameworks. Finally, ministries may play a role in defining and setting the evaluation frameworks of open science initiatives.

Recent initiatives implemented at national level include the following. In Finland in 2014, the Ministry of Education and culture launched the Open Science and Research Initiative (ATT) with the aim of creating a national open access and open science policy and building the infrastructure necessary to reach this goal. ATT aims to make open and collaborative science more visible to innovation system actors, and to promote not only open access to research data and publications, but also transparent, collaborative research and the skills, the knowledge and the support services necessary to achieve these goals. In the framework of ATT, the Ministry plans to organise a yearly "Open Science and Research Forum" to gather all relevant stakeholders and promote fruitful discussion about ATT and its implementation.

In the United Kingdom, open access constitutes a key component of the Department for Business, Innovation and Skills (BIS) contribution to the UK Government Transparency Agenda. The guidelines developed by BIS were informed by the UK National Working Group on Expanding Access to Published Research Findings. BIS is also active in developing metrics and analysis to assess the costs and benefits of open access policies.

In Canada, the revised ST&I strategy, launched in December 2014, commits to open science policies and practices for publicly funded research by increasing public access to the results of government-funded research. An implementation plan will be developed to promote open science, including both open access and open data initiatives, within the activities of science-based departments and agencies as well as those of granting councils and the International Development Research Centre.

In other countries such as Austria, Australia, France, India and the Netherlands, ministries in charge of higher education, research and innovation are committed to investing in the infrastructure for open science activities. In Denmark, open science is one of the pillars of the newly developed national innovation strategy. As of 2014, the German Federal Ministry of Education and Research (BMBF) was developing a comprehensive open access strategy. In Spain, the Secretariat of State for Research, Development and Innovation within the Spanish Ministry for Economic Affairs and Competitiveness is active in promoting open science. In Belgium, the federal Science and Policy Office is creating an institutional repository, and is engaged in broad consultation to develop a coherent and effective open access strategy and mandate. Both are due in 2015. In 2013, the US White House Office of Science and Technology Policy (OSTP) published a memorandum to federal government science agencies, directing them to develop plans to increase and facilitate access to the results of federally funded research, in particular publications and data sets.

In addition, open science can be promoted through the disclosure of government data (Box 4.1). A number of OECD and non-member countries have adopted policies in this respect. Countries such as Australia, Belgium, Canada, Finland, France, the United Kingdom and the United States have disclosed government data on a range of different topics, from weather data to GIS data, in the framework of their open government initiatives. China has also implemented a government data-sharing programme, covering 24 sectors since the beginning of the 2000s.

Box 4.1 Promoting value creation through open government data (OGD)

In carrying out their statutory duties, government bodies produce, collect and manage a vast quantity of data (or provide funds to others to perform these responsibilities). Data are quickly becoming one of the most valuable public goods – yet, they often remain inaccessible or unaffordable to the majority of stakeholders. Enabling access to and reuse of these data has significant potential not only to improve public sector efficiency and transparency, but also to deliver **people-driven governmental actions** that increase economic and social public value.

The OECD highlights three main sets of values targeted by Open Government Data (OGD) policies and initiatives across OECD member countries, which may simultaneously benefit several actors. Potential benefits are envisaged not only in monetary and economic terms, but also from social and good governance perspectives:

- economic value (e.g. growth and competitiveness in the broad economy, fostering innovation, efficiency and effectiveness in government services)
- social value (e.g. promoting citizens' self-empowerment, social participation and public engagement in policy making and service delivery)
- public governance value (e.g. accountability, transparency, responsiveness and democratic control).

Understanding the different values is essential to guide actions aimed at clearer recognition of potential users, their demands, and the priority data to release. Different benefits require different types of data. Boosting economic growth may demand the timely provision and regular updates of specific granular data that are of interest to the business community or app developers, as they can be widely and rapidly disseminated and used. By contrast, many objectives related to accountability and good governance can be served by releasing aggregated data, or by strengthening the ties with intermediary actors playing a key role in processing "open data" to make it understandable and useful to broader society. Social value achieved through a higher level of public engagement in policy and service design and delivery may instead call for data of interest to the relevant user groups who seek active engagement. It is important to align OGD policy goals with public expectations and demands.

The OECD 2013 Open Government Data Survey shows that while political statements include citizens' engagement among the main expected achievements of OGD, public participation is not listed among the top priority objectives of national policies and strategies; These instead focus on increasing economic value for the private sector and increasing openness and transparency. The OECD methodology supports countries in conducting national impact assessment exercises and identifying metrics to support the business cases for open government data (i.e. what to measure, why and how). It also helps them design and implement OGD action plans, face challenges, and follow up on results. Interestingly, the 2013 OECD survey shows that countries consider institutional and organisational challenges the main obstacles to OGD implementation.

Source: Ubaldi, B. (2013), "Open government data: Towards empirical analysis of open government data initiatives", *OECD Working Papers on Public Governance*, No. 22, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k46bj4f03s7-en.

Research funding agencies

Research funding agencies are key actors in the promotion of open science efforts, as they are responsible for defining the mechanisms and requirements to benefit from grants and funding for research. In recent years, in many countries' funding agencies have increasingly adopted rules and mechanisms to promote and in some cases mandate open access, by including open or public access of funded research outputs as a requirement. For example, major funding agencies in Australia, Flanders and the Wallonia-Brussels Federation (Belgium), Costa Rica, Denmark, Estonia, Germany, Switzerland, the United Kingdom and the United States have mandated open or public access to the results of the research they fund. Funding agencies in Canada, Germany and Norway are also considering adopting rules for mandatory open access.

Research funding agencies actively support the development of national and in some cases international infrastructure for sharing articles, data and research material in general. The European

Commission, for example, has largely supported the creation of repositories as well as inter-repository linkages, through Framework Programmes. Funding agencies in Chile, Mexico, Portugal and Germany as well as the Nordic Countries have also supported the development of online networks, archives and platforms.

In addition to mandatory requirements and creation of the infrastructure for open access to research outputs, funding agencies may promote open science through financial support to cover open access publishing or the costs associated with the release of data and other research material. Funding agencies in Belgium, Germany, Norway, the Netherlands, Switzerland and the United Kingdom have adopted mechanisms to cover some of the costs of open access publishing. Elsewhere, governments encourage universities or research organisations to allocate funding for open access initiatives directly.

In Europe, the European Commission supports open access and open data efforts, and it requires research results financed by the Horizon 2020 programme to be made publicly available after publication (although it allows researchers to choose how they disclose research results). According to Horizon 2020 regulations, fees related to open access publishing are eligible for reimbursement under the conditions of the grant agreement. In addition, a subset of projects funded by Horizon 2020 will participate in a pilot open research data initiative that will mandate the disclosure of research data sets and the associated metadata (Box 4.14).

Finally, funding agencies may play a role in promoting an open access culture – for instance by requiring open access and open data management plans as well as by specific copyright licensing agreements, or implementing incentives and reward mechanisms to promote open science efforts, such as data collection, curation and preservation.

Universities and public research institutes

In a majority of OECD countries, universities and public research institutes have certain degrees of autonomy regarding, and are responsible for defining and implementing, STI strategies. In defining these strategies they may develop initiatives to promote open science efforts with respect to research results produced by researchers affiliated with their institutions. Open science policies may vary depending on the university or the public research institutes: for example, whether to follow a gold or green route or the duration of an embargo period. Several of these institutions have been active in developing the infrastructure to facilitate a transition towards open access practices such as repositories and platforms enabling researchers to openly disseminate the results of their work. In addition, universities and higher education institutions may play a role in training students and researchers to develop the skills necessary to enable open science practices, from basic skills related to the use of online repositories to the ones needed to implement data cleaning, curation and management.

In many countries, universities or public research organisations have been at the forefront in adopting open science mechanisms that have been subsequently translated into national strategies and efforts. Examples include the University of Southampton in the United Kingdom; the National Institutes of Health in the United States (Box 4.2); the Centre National de la Recherche Scientifique (CNRS, National Center for Scientific Research) and the Institut National de Recherche en Informatique et en Automatique (Institute for Research in Computer Science and Automation, INRIA) in France; the University of Liège in Belgium (Box 4.3), the University of Helsinki in Finland, or the University of Costa Rica in Central

America. The Indira Gandhi Institute for Development in India has launched the Open Index Initiative to develop an online bibliographic database for most of the Indian literature in the social sciences.

Universities or public research organisations may, in addition, undertake research on open science itself. In Poland for example, the Centre for Open Science is a unit devoted to the development of open science research, tools, services and promotion. Created in 2010 within the University of Warsaw, the Centre develops software tools to support open science and operates the largest Polish research open access infrastructure. It also acts as the centre of competence on open science, including its legal aspects. Other examples include the Open Data Institute in the United Kingdom (Box 4.13).

LERU, the League of European Research Universities, published a Roadmap for Research Data (LERU, 2013) providing policy recommendations to key open science actors and providing concrete examples of open research data initiatives in European Universities, including the University College London, the Swiss Federal Institute of Technology (ETH Zurich), and Dataverse in the Netherlands. Dataverse Netherlands is an example of a shared research data infrastructure for seven Dutch universities: Utrecht University, Tilburg University, Erasmus University Rotterdam, Maastricht University, University of Groningen, 3TU Datacentrum and the Netherlands Institute of Ecology.

Box 4.2 The public access and data sharing policies of the US National Institutes of Health (NIH)

The NIH has developed several policies that predate the Office of Science and Technology Policy memorandum (since as early as 2003), to promote access to publications and data. All researchers funded by the NIH are required to submit an electronic version of the final peer-reviewed paper to the National Library of Medicine's PubMed Central repository (see Box 4.5), where it will be made publicly accessible no later than 12 months after the official publication date. NIH funding may be used to cover open access processing charges and other costs of publishing, and researchers are free to choose the journal in which to publish, whether it is open access or subscription-based. Since the policy became mandatory in 2008, the NIH has funded more than half a million peer-reviewed articles, of which more than 82% have been made available through PubMed Central.

The cost of implementing the NIH public access policy varies between approximately USD 4 million and USD 4.5 million per year, depending on the number of articles submitted to PubMed Central. This budget corresponds to a small fraction of the NIH annual budget (approximately USD 30 billion).

The NIH public access policy is implemented in a way that is consistent with the US copyright laws. Publishers or authors retain copyright when submitting papers to PubMed Central, and the papers are available consistent with those rights.

The NIH has several policies to promote data sharing. The 2003 NIH Data Sharing Policy expects applicants requesting USD 500 000 or more of funding in any given year to include a data-sharing plan in the grant application or to justify why data sharing is not possible. Data-sharing plans should include a description of whether and how data will be made available, including how to account for protection of privacy, confidentiality, security, intellectual property rights; a description of the data to be shared, the timeline of sharing, data formats, procedures related to data-sharing agreements and limitations on the use of data. The policy expects data to be shared no later than the acceptance date for publication of the main findings from the final data set. In February 2015, NIH announced plans to extend its data sharing policy to all supported research, regardless of funding level.

Other NIH data-sharing policies are specifically developed to target different types of scientific data, collected during different projects researching different aspects of medicine, health and biological research. For example, the Genomic Data Sharing (GDS) Policy (issued in August 2014) is an expansion of the 2007 NIH policy for Genome-Wide Association Studies (GWAS). The GDS Policy requires researchers to register all studies using human genomic data in the database of Genotypes and Phenotypes (dbGaP), maintained and operated by the National Library of Medicine. Data resulting from the study are to be deposited in a designated public repository in a timely manner. In general, data will be available no later than six months after the initial data submission begins, or at the time of acceptance of the

first publication, whichever occurs first. The GDS Policy came into effect in January 2015.

Source: http://www.nih.gov/.

Box 4.3 The open access policy of the University of Liège (Belgium)

The University of Liège adopted its mandatory open access policy in May 2007. Researchers have to self-archive their outputs following the principle of "Immediate-Deposit & Optional-Access" (IDOA) in the institutional repository of the university, ORBi. The deposit in ORBi is mandatory as soon as the article is accepted by a scientific editor. Assessment of research performance and the evaluation of researchers within this university are exclusively based on the research outputs that are deposited in ORBi. In addition, internal grant distribution procedures are based on the statistics from the publication record of ORBi. In order to facilitate the transition period, seminars and classes to teach and explain the functioning of ORBi were organised after the policy was adopted. According to the rector of the university, the development of ORBi offered several advantages to the university and its researchers: acceleration of dissemination and the visibility of the scientific work; increased visibility for the published papers through main search engines; and the centralised and perennial conservation of publications for multiple purposes. This model has proved successful, and it is often referred to as the "Liège model" internationally.

Source: http://orbi.ulg.ac.be/bitstream/2268/102031/1/Rentier-WashDC-2011.pdf; http://roarmap.eprints.org/56/.

Libraries, repositories and data centres

Libraries and repositories are key actors and fundamental vehicles to make open science work. Thanks to ICTs, libraries have taken on a new role: they are now active in the preservation, curation, publication and dissemination of digital scientific materials, in the form of publications, data and other research-related content. Libraries and repositories constitute the physical infrastructure that allows scientists to share, use and reuse the outcome of their work, and they have played an essential role in the creation of the green open access movement. Globally acknowledged online discipline-specific repositories are for instance PubMedCentral in the life sciences, arXiv in physics, mathematics and computer sciences, and Repec in economics (Box 4.5). Issues related to the healthy functioning of libraries and repositories are related to the sustainability of investments to create and maintain the infrastructure itself and the interoperability of different systems.

In several OECD and non-member countries, there are a number of ongoing efforts to create online repositories, databases, archives and digital libraries and platforms containing information on R&D projects and researchers' CVs. For example, Estonia and Poland have created national networks of repositories and digital libraries. Finland has launched an infrastructure roadmap to promote open science. In Greece, the National Documentation Centre (EKT) is the national institution for the aggregation, documentation and preservation of research and cultural online material. China has developed online platforms for data and publication archiving. Argentina developed the SICyTAR database with information on the CVs, publications and affiliations of researchers. The European Commission has also been active in promoting the development of EU and member country repositories and platforms, as well as the inter-linking infrastructure (Box 4.6). In addition to national-level initiatives, several innovation system actors, notably universities and public research institutes, are active in the creation of digital repositories for self-archiving purposes.

As science becomes increasingly data-driven, data centres are important actors to collect, clean and curate the data (both in the short and long term), as well as to assist and provide expertise to both institutions and individual researchers. Data centres, generally supported by consortia of institutions or by public research funders, can be either national or international facilities. They generally host computer system and related ICT hardware, such as telecommunications and storage systems. Data centres can be discipline-oriented or gather and maintain data sets irrespectively of scientific fields. Internationally known data centres are found in several countries (See Box 4.4).

Box 4.4 Data centres: Storing, curating and providing access to data

The *British Natural Environment Research Council (NERC)* is the leading funding agency of research, training and innovation in environmental-related science in the United Kingdom. NERC has a network of data centres that store key information for NERC's research. The centres collect and curate data from environmental scientists active in the United Kingdom or in other countries. The data centres are responsible for preserving environmental data and making them available to all users: from researchers in academia to the business sector, governmental institutions and citizens in general. Data stored in NERC data centres are carefully curated to ensure long-term availability, by using state-of-the-art data management and preservation techniques. NERC developed a Data Catalogue Service to allow integrated search in all its data centres. Some of the centres also store different forms of research material, ranging from sample materials collected through various research activities to material supplied by third parties. Data and these materials are key resources in research dealing with environmental challenges such as climate change, conservation of endangered species or the management of water quality. NERC supports seven data centres in different environmental disciplines:

- The British Oceanographic Data Centre
- The British Atmospheric Data Centre
- The UK Solar System Data Centre (solar and space physics)
- The Environmental Information Data Centre (earth and water)
- The National Geoscience Data Centre
- The Polar Data Centre (polar and cryosphere)
- The NERC Earth Observation Data Centre

In addition, NERC has an agreement with the Archaeology Data Service to manage and share data collected through NERC-funded research in science-based archaeology.

The *UK Data Archive* is an international leader in data curation and sharing. It curates the largest collection of digital data in the social sciences and humanities in the United Kingdom. It was established in 1967 by the Social Science Research Council, with a long-term commitment of funds. The Archive collects data from surveys, questionnaires and interviews, with the aim of allowing researchers in one area to exploit already existing data sets arising from other areas of research. Since 2005, the British National Archives have designated the Data Archive to curate public records. The UK Data Archive acquires data from academia and public administrations as well as commercial sectors. It provides continuous access to the data acquired and promotes the creation of data users communities. The UK Data Archive manages the UK Data Service, a portal for research resources that hosts survey data collections, databanks, census data and qualitative data in a secure manner. The UK Data Archive is constantly involved in data management and preservation initiatives, and it provides data curation for third organisations.

The *Australian National Data Service (ANDS)* aims to create a cohesive national collection of research resources to make better use of Australia's research outputs as well as promote better use and sharing of research data and material. ANDS creates partnerships with research teams and data-producing agencies to acquire and store new data sets. It delivers services such as the interlinking of data sets from different sources and organisations, as well as data-citing tools to acknowledge the authors of different data sets. ANDS provides guidance and advice on the management, production and reuse of data, and promotes the creation of communities of practices. ADNS is also creating the *Australia Research Data Commons*, a platform for researchers aiming to provide a set of data collections ready to be shared, a description of the relevant information characterising these collections, and an infrastructure that enables data sharing and data exploitation.

The *Indian Council of Agricultural Research* (ICAR), an independent institute of the Department of Agricultural Research and Education of the Indian Ministry of Agriculture, hosts databases containing agriculture records and the related technological interventions. In addition, each of the 99 ICAR Institutes has been mandated to create an open

access institutional repository. ICAR plans to develop a central repository to collect research material and the associated metadata from all ICAR's institutes. Metadata and research material are shared freely for public research purposes, whereas commercial uses require written approval.

Source: www.nerc.ac.uk; www.data-archive.ac.uk/; www.ands.org.au/, www.icar.org.in.

Box 4.5 Examples of repositories: PubMed Central, arXiv and Repec

PubMed Central

PubMed Central (PMC) is a free archive of biomedical and life sciences journal literature developed by the US National Library of Medicine (NLM). Established in 2000, PMC has been created to collect and preserve biomedical literature; it is the digital counterpart of NLM's print journal collection. Free access to all of its journal literature is a core principle of PMC. Although providing free access, publishers and individual authors retain the copyright on the material they submit to PMC, and users must adapt to the terms defined by copyright holders. In addition, although free access is mandatory, publishers can delay the release of research material after publication (a so-called embargo period) for a maximum of 12 months. As of 2014, PMC contained more than 3.2 million research articles, and has become a key repository searched regularly by researchers in academia and industry, educators, students, the general public – and major search engines, making deposited papers more visible and accessible. It has been estimated that on a typical weekday, more than 1 million different users download more than 2 million different articles from PMC.

arXiv

arXiv is a highly automated electronic archive for research articles, created in 1991. Originally it was hosted by the server of the Los Alamos National Laboratory in New Mexico, United States. Today it is maintained and operated by the Cornell University Library under the guidance of the arXiv Scientific Advisory Board and the arXiv Sustainability Advisory Group. arXiv was originally conceived as an article pre-print archive in physics; it then expanded to cover other thematic areas such as mathematics, computer sciences, quantitative biology and statistics. Users can download papers from arXiv via the web interface, and registered authors may use the web interface to submit articles as well as update their submission records. arXiv has been a frontrunner in offering alternative methods of disseminating the results of scientific research and the open access movement.

RePEc (Research Papers in Economics)

RePEc was created in 1997. It is a collaborative effort of volunteers in 82 countries to promote the dissemination of and free access to research papers in economics and related sciences. RePEc is a decentralised bibliographic database of working papers, journal articles, books, book chapters and software components. RePEc contains about 1.4 million research items from 1 800 journals and 3 800 working paper series. Approximately 70 000 subscriptions are registered every week. Publishers themselves index their contents into RePEc, and develop the metadata according to RePEc guidelines. RePEc also delivers the following list of services: browse and search in RePEc the database; development of an academic family tree for economic disciplines; detailed download and access statistics (including co-authorship and network centrality indicators); efforts to prevent plagiarism of RePEc content; and citation analysis.

Source: www.ncbi.nlm.nih.gov/pmc/, http://arxiv.org/help/general, http://repec.org.

Box 4.6 Promoting open access at European level: The case of OpenAIRE

OpenAIRE is a project funded by the European Commission with the aim of supporting implementation of open access in Europe. It provides the means and the physical infrastructure to promote the adoption of the open access policies conceived by the European Research Council and Horizon 2020. OpenAIRE works through an extensive European Helpdesk System, based on a network of national and regional liaison offices in 27 countries. It also provides a repository facility for researchers who do not have access to institutional or discipline-specific repositories.

OpenAIRE received a budget of approximately EUR 4 million from the EC, and its total budget from 2009-11 corresponded to about EUR 5 million.

A second phase of the project, OpenAIREplus, is being currently developed to extend and facilitate open access by providing a cross-link from publications to data and funding schemes. OpenAIREplus brings together 41 European and pan-European partners, including 3 cross-disciplinary communities. OpenAIREplus will expand the current publication repository networks to include data providers, with the goal of interlinking associated scientific data. OpenAIREplus receives a total budget of approximately EUR 5 million, of which around EUR 4 million comes from the EC, covering a period of 30 months (since December 2011).

OpenAIRE also has the goal of leveraging its international connections to contribute to the definition of common standards and interoperable systems on a global level. So far, OpenAIRE covered more than 8 million publications, and 600 data sets from more than 400 repositories and open access journals.

Source : https://www.openaire.eu/

Supra-national entities

Several international organisations are directly involved in the promotion of open science. They often develop guidelines and principles related to open access and open data. They promote international co-ordination of efforts to support more effective sharing of information and research outputs. Supra-national entities have a role to play in order to ensure interoperability of systems and standards, especially across repositories. Given the large amount of public support that hard infrastructure is receiving in OECD countries and beyond, it is important to discuss those mechanisms through international platforms to make sure that different repositories are interoperable to maximise their usage – as well as co-ordinating, when possible, investments aiming to develop similar types of infrastructure or having similar goals. By promoting international co-ordination, supra-national entities may help increase the efficiency of open science efforts internationally, thus contributing to make them sustainable in the long term (Box 4.7).

In addition, supra-national entities are often committed to understanding how open science can promote capacity building and science and research advancement in developing countries. (See the section Global Open Science for more details on inter-governmental organisations and the role they play in shaping the policy agenda of member countries.)

Box 4.7 International research organisations involved in open science

International research organisations actively involved in open science (access and/or data) efforts include:

- *CERN* (the European Organization for Nuclear Research, web.cern.ch) is an international research laboratory containing the world's largest and most complex scientific instruments to study fundamental particles. CERN was founded in 1954 and is located across the Franco-Swiss border. It was one of Europe's first inter-country joint ventures and it has now 21 member states. CERN actively support open access efforts. In particular, since 1 January 2014 CERN has been hosting the Sponsoring Consortium for Open Access Publishing in Particle Physics (*SCOAP^3*, see Box 2.1). SCOAP^3 is supported by partners in 37 countries and works to make available free of cost scientific articles in the field of high-energy physics. SCOAP^3 involves the collaboration of over 1 000 libraries, library consortia and research organisations. SCOAP^3 benefits from the support of funding agencies, and has been established in co-operation with the publishing industry. As a result of SCOAP^3, articles are open access, the copyright stays with the author(s), and licensing agreements allow text and data mining.

- *The Global Research Council (GRC)* – The Global Research Council is an organisation composed of the heads of science and engineering funding agencies around the world. The GRC aims to promote the sharing of data and best practices for collaboration among funding agencies globally. The GRC has developed an Action Plan towards Open Access to Publications (GRC, 2013) to promote the diffusion and

sharing of research results. The action plan highlights the importance of raising awareness *vis-à-vis* open access in the research community; promoting and supporting open access through funding streams and by working together with publishers; and exploring new ways to assess the quality and impact of research.

- *ICSU, the International Council for Science,* is a non-governmental organisation gathering members of national scientific bodies and international scientific unions world-wide, representing 140 countries. ICSU was founded in 1931 to promote international scientific activity. ICSU's mission is to promote international science co-operation for the benefit of society. ICSU identifies and addresses major issues of importance to science and society; facilitates interaction among scientists across all disciplines and countries; and provides independent advice to stimulate dialogue among the scientific community and governments, the civil society and the private sector. The ICSU 2012-17 strategic plan has identified the following priorities: i) international research collaboration; ii) science for policy; iii) the universality of science. ICSU has recently published its statement on Open Access Principles.[46]

- *CODATA*, the Committee on Data for Science and Technology, is an interdisciplinary Scientific Committee of ICSU. CODATA works to improve the quality, reliability, management and accessibility of science and technology data. CODATA promotes awareness and cross-border co-operation of scientists. It was established in 1966 by ICSU to promote globally the compilation, evaluation and dissemination of reliable numerical scientific data. CODATA is legally independent from ICSU; it includes 23 members across different continents. Country membership often takes place through national research councils. CODATA activities include both technical discussion on standards and interoperability, and policy-level discussion on data issues. CODATA works on different aspects of data, from research data to social science data, government data, PSI, big data, etc. CODATA is concerned with all types of data resulting from experimental measurements, observations and calculations in every field of science and technology, including the physical sciences, biology, geology, astronomy, engineering, environmental science, ecology and others. Particular emphasis is given to data management problems common to different disciplines and to data used outside the field in which they were generated.

- *The Research Data Alliance* has the goal of promoting data sharing to accelerate data-driven innovation discovery, use and reuse of data, standards harmonisations, and discoverability. RDA is organised into working groups and interest groups around different themes, where experts from different countries and belonging to different communities (academia, the business sector, governmental agencies) meet and discuss. RDA was created in 2013 by a core group of organisations: the European Commission, the US National Science Foundation and National Institute of Standards and Technology, and the Australian Government's Department of Innovation. Individuals may also apply for a membership; today RDA counts around 1 600 individual members from more than 70 countries.

- *The EMBL-EBI***.** The European Bioinformatics Institute (EBI) is part of the European Molecular Biology Laboratory (EMBL), a non-profit organisation and basic research institute funded by 20 member states in Europe and Israel and one associate member, Australia. EBI is a major European laboratory for the life sciences, and it provides freely available data from life science experiments in the field of molecular biology. EBI maintains the world's most comprehensive collection of freely available up-to-date molecular databases. EBI services allow scientists to share data, perform complex queries and analyse the results. Database users can generally work locally by downloading EBI data and software and use EBI web services to access different resources. EBI serves millions of researchers world-wide active in multiple fields of life sciences, from clinical biology to agri-food research. EBI also offers training programmes to researchers in academia and the business sector to maximise the benefits of the data available in the life sciences. Some 20% of EBI users are engaged in industrial R&D, and EBI has developed an Industry Programme to collaborate specifically with firms active in bio-informatics. EBI addresses the specific needs of industry in other ways: from public-private partnerships to develop better and safer medicines for patients (the Innovative Medicines Initiative) to the provision of data infrastructure and services to SMEs, enabling bio-informatics spin-offs from EMBL, and facilitating key pre-competitive research projects with industrial partners. EBI is located on the Wellcome Trust Genome Campus in the United Kingdom.

Source: http://home.web.cern.ch/; http://www.globalresearchcouncil.org/; http://www.icsu.org/; http://www.codata.org/; https://rd-alliance.org/; http://www.ebi.ac.uk/.

[46] http://www.icsu.org/general-assembly/news/ICSU%20Report%20on%20Open%20Access.pdf.

Private non-profit

In addition to public science and innovation actors, private non-profit organisations and foundations may play a significant role in developing, raising awareness of and encouraging an open science culture. They can not only fund open access research and introduce requirements in grant agreements, but also develop and facilitate the creation of networks of stakeholders world wide. Examples of this kind of organisation are for example the Wellcome Trust or the Open Knowledge Foundation, and more recently the Bill and Melinda Gates Foundation (Box 4.8).

Box 4.8 Non-profit for open science: The examples of the Wellcome Trust, the Open Knowledge Foundation and the Bills and Melinda Gates Foundation

The *Wellcome Trust* is an independent charitable foundation supporting research related to health improvements in humans and animals. The Wellcome Trust support includes public engagement, education and research to improve health. The Wellcome Trust has been active in promoting open access and open data, by supporting unrestricted access to the published output of the research it funds, wherever possible. It requires electronic copies of any research paper supported in whole or in part by the Wellcome Trust to be made publicly available within six month after journal publication, and provides grant holders with additional funding to cover open access charges. In addition, the Wellcome Trust encourages the open release of research data by requiring data management plans to be included in grant applications; by expecting users of research data to acknowledge the sources of their data; by recognising the contributions of researchers who generate, preserve and share research data sets; and by developing best practices for data sharing in different fields, recognising that different data types raise different issues and challenges. The Wellcome Trust is based in the United Kingdom.

The *Open Knowledge Foundation (OKF)* is a global network of people promoting knowledge sharing, and access to information and data. To achieve this goal, OKF promotes awareness and a changing culture among policy makers, business organisations and civil society. OKF is committed to developing and using tools (technical, legal and educational) to make knowledge sharing easier. The OKF co-ordinates and supports international networks of individuals sharing its mission; conducts campaigns for the open release of key data and information; and monitors the level of openness world-wide. The OKF in addition offers training and consultancy services to different kinds of stakeholders, so they can better understand how to release and use open data as well as to develop the right set of skills necessary to maximise the benefits and possibilities offered by open access to data and information. The OKF is located in Cambridge, United Kingdom.

The *Bill and Melinda Gates Foundation* has recently adopted an open access policy to enable unrestricted access and reuse of all peer-reviewed published research it funds, including any underlying data sets. This open access policy became effective at the beginning of 2015. Between 2015 and 2017, a 12 month embargo period will be applied; however, from 2017 no embargo period will be allowed. The Foundation will pay open access fees if required and recommends publishing content via Creative Commons or equivalent licences that allow to copy and redistribute material for commercial purposes. The Bill and Melinda Gates Foundation headquarters are located in Seattle, United States.

Source: www.wellcome.ac.uk; https://okfn.org/about/, www.gatesfoundation.org

Private scientific publishers

The business community is involved in open science mainly in two ways. Business organisations may be the actors providing services and infrastructure for open science, as in the case of private scientific publishers offering open access publishing (for example via the gold route or publishing in hybrid journals) and related key services such as the maintenance of digital repositories and data sets or other scientific material. The role of private scientific publishers is evolving: many publishers are offering new services related to data storage, archiving and sharing to promote open science (Ware and Mabe, 2012). Many scientific publishers support open science as indicated by their signing of the 2007 STM Brussels

Declaration.[47] In recent years a number of start-ups offering open science-related innovative services, such as the storage of research data or research papers archiving, have appeared (see Box 4.9).

Box 4.9 Start-ups for open science: The case of Figshare

Figshare is an online digital repository of research data, including figures, images and videos. Figshare was launched in 2011 by a PhD student in London. Figshare users can make their research outputs available to other researchers or users in a "citable, sharable and discoverable manner". This means that they can easily share data, search for data sets and get credit for the data sets they uploaded on the website (through data set citations). Figshare allows users to upload any file format.

Figshare has recently established partnerships with other open science business actors, such as the open access publishing company *PLoS*, the Nature Publishing Group, Taylor and Francis and F1000, to allow authors to directly upload data sets linked to papers online, as well as with ORCID (see Box 3.4), a service to allow data set citations. In addition, Figshare tracks the number of downloads of research materials, and it is often used as a source of alternative metrics. All files uploaded on Figshare are released under a Creative Commons licence (see Boxes 2.3 and 2.4).

Figshare stores more than 1.5 million files, and from its original location in the United Kingdom it today has employees located in the United States and Romania. Users can sign in and upload or download content on Figshare for free. The company, however, charges for premium services (such as larger private online storage space or private collaborative spaces) to individual researchers and for services offered to publishers. In addition, it recently launched Figshare for Institutions, a service explicitly designed for research institutions around the globe.

Source: www.figshare.com.

Along with major traditional scientific publishers, such as Springer, Elsevier or the Nature Publishing Group, a number of more recent open access scientific journals have begun to emerge in the 2000s. Well-known examples include *PLoS* (Public Library of Science) and BioMedCentral. This new generation of scientific publishers offers open access publishing and requires the authors of publications to pay the costs associated with the APCs; these include, for example, the editing of the article or the costs associated with organising the peer-review process. In addition to open access publishing services, these journals often deliver additional "open science services", which contribute to making the scientific information landscape more open. They have opened blogs with discussion around scientific themes or specific articles; they compute alternative metrics for research papers; and they exploit the possibilities offered by the open source software to launch open source publishing platforms and to enable developers to create applications for smart phones or tablets. They generally develop tools and offer assistance to allow text and data mining of the articles they publish. A business model that has recently emerged in the publishing industry is the so-called Freemium model: basic services are offered for free, but the company charges for accessing a number of more sophisticated services.[48]

[47] The 2007 STM Brussels Declaration is available at: http://www.stm-assoc.org/brussels-declaration/.

[48] As defined in OECD (2015a): "The term 'freemium' is a portmanteau for the words 'free' and 'premium'. The *freemium* revenue model is one of the most dominant revenue models in the data ecosystem, which seems to be particularly attractive to start-ups. In the *freemium* revenue model, products are provided free of charge, but money is charged for additional, often proprietary, features of the product (i.e. *premium*). The freemium revenue model is often combined with the advertising-based revenue model when products are provided for free to consumers, and with the subscription-based revenue model for the premium."

The business community

The business community however is also involved in open science, as private firms may be the beneficiaries of open access publications and data to be used to develop new products and services and promote innovation more generally (OECD, 2015a). Example exists in the field of health-related research (Box 4.11). In addition to entirely privately funded initiatives, there is potential for the development of joint public-private partnerships for the delivery of open science-related services. The US OSTP memorandum explicitly encourages collaboration with the private sector, for example through public-private partnerships, to promote open science and open data. Examples of public-private partnerships involved in open science have been recently developed in Finland (Box 4.10).

Box 4.10 Public-private partnerships for open science: The Finnish SHOK, DIGILE

The Strategic Centres for Science, Technology and Innovation (SHOK in Finnish) established in Finland are new public-private partnerships aiming to speed up innovation processes, by the renewal of clusters and the development of radical innovations. SHOK centres develop and apply new methods for co-operation, co-creation and interaction.

One of Finland's Strategic Centres for Science, Technology and Innovation is directly involved in open science: DIGILE. DIGILE´s mission is to create digital business ecosystems to enable new global growth business for DIGILE's owners and partners. DIGILE aims not only to bring together R&D communities, but also to make sure that the results of scientific processes are understood, applied and adopted by companies. There are over 30 partners, including companies, research institutes and universities. DIGILE's strategy for 2015 focuses on data sharing, management and reusing as well as innovative data-intensive business models and services.

Source: www.digile.fi, Finland country note and Myllymäki, P. (2013), "Data to Intelligence (D2I) research programme on intelligent data-driven services" Presentation, www.digile.fi.

Box 4.11 Big data and health-related research

As in other research fields, the advent of ICTs, the rise of Web 2.0, the proliferation of electronic health records, smart devices and machine-to-machine communication are opening up new opportunities for advancing health-related research. Recent examples have been genetic testing models that often require the manipulation and the analytics of large amounts of data in short amounts of time or the early detection of chronic disease. In some cases data-intensive health research is managed by international consortia – as in the case of the International Cancer Genome Consortium, a multidisciplinary collaborative effort aiming to detect somatic mutations in over 24 000 tumour genomes.

Web- and mobile-based social media applications are increasingly emerging as new channels for the collection and dissemination of health and lifestyle records. Online platforms can easily reach large numbers of people and share information on therapies and disease progression. They can also collect large amounts of data on patients in different countries and with different characteristics. For example, the social network PatientsLikeMe launched global data collection processes to get information from individuals suffering from specific diseases and the usage of certain drugs. Analysis of Twitter strings of text may provide evidence on people's opinions on diseases or treatments, as well as provide information on health-related behaviours.

Source: OECD, 2015a, 2014a.

Open science and citizen involvement

The participation of amateurs in scientific processes, the interaction with professional scientists, is not a new concept; it began in the 18th century and generally related to data collection or observation, in particular in disciplines like ornithology or astronomy. However, the improvement in communication capabilities, the emergence of mobile devices, the increase of storage capacity of the information collected, the possibility of transmission not only of text but also of images and sounds, and (certainly) the existence of greater public awareness has led to the emergence of so-called *citizen science*.

Citizen science, as a specific term to identify these initiatives, was introduced in the mid-1990s by Rick Bonney in the United States and Alan Irwin in the United Kingdom**.** This is a fuzzy and complex term whose definition covers a great variety of activities, some pertaining more to education, others to scientific practise, and some that mix both – as education, learning, teaching and practising in science are always very closely connected. In order to clarify its scope, some general definitions for citizen science have been elaborated. Citizen science could be understood as "projects in which volunteers partner with scientists to answer real-world questions", as it is stated in Citizen Science Central, a website part of Cornell University's Laboratory of Ornithology in Ithaca, New York.

In an attempt to structure the variety of activities carried out, different aspects of citizen engagement have been identified. A first aspect is related to the degree of public participation, with respect to the kind of role the non-professional is playing. A second aspect is related to the role of citizens in the decision-making process of selecting the research streams that will be publicly financed. Finally, citizen science has specific organisational characteristics related to the development of networks of both professional and non-professional personnel, through dedicated events as well as the need of technical support from the scientific to non-scientific communities (Holocher-Ertl and Kieslinger, 2013). Depending on the project, the involvement of citizens varies: they may contribute to the projects by collecting samples or records, or be more actively involved in the analysis and dissemination of results (Bowser and Shanley, 2013).

It has been argued that citizen science is a means for reaching several different objectives (Riesch, Potter and Davies, 2013). For example, it allows the development of a more democratic environment in science by engaging professionals and amateurs in research and scientific efforts. The participation of civil society in these activities shows the level of commitment that has been reached. There is a clear willingness to be directly involved in the scientific process, not only as observers or data collectors, but also as practitioners, planners and evaluators. Society's participation in the process could even lead decision makers to opt for research priorities based in amateur scientist conclusions or revoke decision previously taken, as in the case of the London District of Deptford, where the UK Environment Agency revoked a scrapyard's licence after evidence resulting from citizens' collection of data on noise levels showed that the operation violated noise limits (Gura, 2013).

In addition, the involvement of citizens in scientific projects tends to have educational value, both implicit and explicit. While in the majority of projects the informal learning aspect of adult citizens is addressed, schools are increasingly considered an important target for the introduction and promotion of citizen science. Teachers play a relevant role in facilitating the deployment of experiments and transmitting the socio-scientific values of their contributions to the young audience. The involvement of citizens in scientific efforts, in addition, may have positive implications in the development of a scientific awareness culture.

Several activities promote the engagement of citizens in science. One of the first examples dates back to 1900: the Audubon Christman Bird Count gathers teams of volunteers every holiday season to survey local bird populations and contribute to monitoring the conservation of bird species. The Audubon Christman Bird Count provides more than 100 years of data and supported over 200 scientific publications in the field.

The project "Amateurs as Experts" was a three-year study of volunteer naturalists, biodiversity scientists and policy makers involved in the UK Biodiversity Action Plan process. The project began in October 2002 and lasted three years. The aim was to enrol new actors into the formal UK biodiversity policy process and have them gain experience in carrying out social experiments and analysing and assessing their progress, benefits and/or problems. Through ethnographic research methods, the study tried to clarify the social and knowledge dynamics, while also fostering patterns of interaction between social and natural scientists and policy actors. The project was a cross-disciplinary research study, involving sociologists, anthropologists (from Lancaster University) and natural scientists (from the Natural History Museum, London). It focused its objectives on effective biodiversity protection policies, in the United Kingdom and beyond.[49]

"The Open Air Laboratories (OPAL)" is a portfolio of projects that has been running in the United Kingdom since 2007, funded by a National Lottery Grant. It was set up with the aim of enhancing environmental knowledge and involving members of the public in the production of science. OPAL is an initiative that, under the direction of mostly university-based scientific teams, benefits from the participation of volunteers gathering data in areas such as biodiversity and air, water and soil pollution. Over 200 000 people have participated so far, including over 1 000 schools and 1 000 voluntary groups. Benefits include a substantial, growing database on biodiversity and habitat condition (Davies et al., 2011).

Another successful example of citizen engagement in science and research is the Zoouniverse Platform.[50] It hosts 30 separate citizen science projects dealing with astronomy, astrophysics, climate, history, ornithology, marine life and other areas. Participants contribute by performing data classification and analysis of different data in diverse formats. They not only provide the analysis but also participate in discussion forums that allow sharing of ideas and communication among participants. Today there are more than one million participants in the platform (Tinati et al., 2014).

Other examples of crowdsourcing of technical skills to solve scientific problems are, for example, the emergence of online platforms where the solution of unsolved scientific problems is requested from the public. Examples of this kind of website include Kaggle (www.kaggle.com), a web-based platform for predictive modelling and analytics, where private companies and research teams publish unsolved problems related to specific data sets (also published on the platform), and data scientists from all over the world compete to find the best solutions and the best-performing algorithms. The crowdsourcing approach relies on the fact that there are countless strategies to solve a problem, each with a different computational efficiency. The best strategy wins and receives the monetary amount advertised by the firm or the research

[49] www.lancaster.ac.uk/fss/projects/ieppp/amateurs/.

[50] www.zooniverse.org/.

team posting the problem online. It is estimated that Kaggle connects around 200 000 data scientists worldwide, and has partnered with organisations including NASA and Deloitte.

Another organisation that is based on crowdsourcing "solvers" is InnoCentive, a platform with more than 300 000 registered users who may gain a reward of between USD 5 000 and USD 1 million if their solutions work. Other examples are found in the public sector. The US Office of the National Coordinator for Health Information Technology (ONC) has launched the Investing in Innovation initiative. The initiative rewards solved challenges with prizes and aims to promote innovation in the developer community, especially by reaching IT developers active in non-health-related fields who can bring skilled expertise traditionally applied in other sectors to health research. Thanks to the initiative several applications have been developed, including some that allow patients to access and share health records (OECD, 2014a). The NASA Space Apps Challenge offers rewards for challenges solved with public data. Each challenge is written by a government or private partner that possesses the data but lacks the capabilities or the personnel to solve the challenge. So far more than 8 000 people have participated in the Space Apps Challenge, participating in almost 700 projects.

Several initiatives have recently been developed in Spain. At sub-national level, the project "atrapaeltigre.com"[51] is a pilot project of citizen science to study where and how the tiger mosquito, an invasive species from Southeast Asia that has recently settled in Catalonia, Northwest Spain, is dispersed. Other examples of Spanish initiatives include the MED-JELLYRISK Project.[52] This project brings together countries of the Mediterranean basin with the aim of assessing the socio-economic impacts of jellyfish blooms and to implement mitigation measures. The MED-JELLYRISK Consortium comprises four members: Italy, Malta, Spain and Tunisia. The project involves the participation of citizen scientists in a Jellyfish Photography Competition, and contributes in different ways to educating children with games and other teaching activities.

GripeNet.es[53] is a tool designed by researchers at the Institute of Biocomputing and Physics of Complex Systems (BIFI) of the University of Zaragoza in co-ordination with the European research consortium EPIWORK, whose primary objective is to monitor the incidence of influenza in Spain through the collaboration of anonymous volunteers via the Internet. The project has a dual purpose. First, the data collected will allow researchers to better understand the mechanisms that spread infectious diseases; second, the project will be an outreach tool to bring science to citizens and involve them directly in scientific study through the information provided.

Governance of open science: Recent policy trends

OECD and non-member countries are increasingly developing legal and policy frameworks, guidelines and initiatives to encourage greater openness in science. They are also promoting incentives and funding mechanisms. Most of the countries responding to the STI Outlook 2014 policy questionnaire

---

[51] http://atrapaeltigre.com/web/participeu/app-tigatrapp/.

[52] http://jellyrisk.eu/es/.

[53] www.gripenet.es/ .

highlighted recent changes in their policy frameworks for open science (OECD, 2014b, Country Notes). A detailed description of these initiatives is found in the Country Notes.

It has to be noted however that open access policies (i.e. policies to promote access to scientific articles) are more mature than open research data policies. This is due to the fact that data sets are not as easy to define as a finalised research article. In addition, in order to be reused, data need to be clean and linked to metadata. The release of data may also raise privacy and security concerns as well as ownership and associated intellectual property issues (OECD, 2015a; Wilbanks and Macmillan, forthcoming Lane et al., 2014).

According to a recent survey prepared for the European Commission (Science-Metrix, 2013), funding bodies have fewer open data policies than open access policies: of the 48 funders of open access policies listed on ROARMAP (a registry for open access policies) in Europe and Brazil, Canada and the United States, 23% explicitly excluded the release of data as a requirement and 38% did not mention data in the policy description. However, 29% of policies mandated the open share of research data and 10% encouraged data sharing without mandating it. Research institutions (such as universities or public research centres) behave in a similar fashion, with 42% indicating they have an open access policy but only 11% reporting to have an open data policy. This result may also be related to the fact that few institutional repositories are dedicated exclusively to research data, although the open research data infrastructure appears to be more developed than the related policy framework (36% of the respondents stated that their organisation developed repositories for open research data whereas 11% indicated that open research data policies are in place).

Open science governance and efforts may benefit from multi-stakeholder and networked consultative approaches, including all relevant actors: governments, academic and research institutions, civil society and the business sector. Policy measures to promote open science (as emerges from the sections below) may be developed and adopted by a diverse set of actors both at national and sub-national level as well as at the institution level, such as in the case of universities or public research institutes. Policy measures may include different actions and initiatives, such as *mandatory rules*, *incentive mechanisms* or *enablers*. Measures belonging to the three types of actions may be implemented together to promote open access by means of integrated and multifaceted approaches. Recent policy trends, however, have revealed that the majority of initiatives involve mandatory rules and requirements, and the development of infrastructure to enable open science. Fewer initiatives related to the definition of non-monetary incentive mechanisms have been designed so far: commonly, incentive mechanisms are in the form of funding to cover open access publishing costs, whereas the definition of new rewarding criteria for researchers involved in open access and open data activities are less common, although they are currently under discussion in a number of countries.

The sticks:[54] mandatory rules and requirements

In a number of OECD countries, major funding agencies have mandated public access to the results of the research they fund. In most countries, the requirement is limited to mandate gratis public or open

[54] The categorisation of open science instruments into sticks, carrots and enablers emerged during a workshop organised at the OECD in December 2013, after the presentation of Salvatore Mele (CERN).

access to publicly funded research results and have researchers free to choose the way they prefer to disclose research papers – that is, whether to follow a gold route (via open access journals) or to self-archive the paper by means of online repositories (the green route). Exceptions are the open access policies in the United Kingdom and the Netherlands. In the United Kingdom, following the recommendations of the Finch report, the gold route has been preferred to the green one as a way to more effectively mandate and obtain open access to research results. In the Netherlands, the national strategy for open access has a preference for gold open access publishing, although it accepts green.

In some countries and in some institutions, researchers are required to make research publications available at the date of acceptance of the manuscript, as in the case of the University of Liège (Box 4.3). In other cases, as in the case of the public access policy of the NIH (Box 4.2) or of the European Commission Horizon 2020 programme, embargo periods are allowed. The duration of embargo periods may vary, but typically it does not exceed 12 months from publication. In some cases, articles are explicitly required to be in a machine-readable format, as in the case of the EU Horizon 2020 policy (Box 4.14).

The requirement for the open disclosure of data sets has been implemented less frequently by countries, and the level of policy implementation in this respect is at a less mature stage, as witnessed by the pilot project on Open Data in EU programmes (Box 4.14). When mandatory policies with respect to data release are adopted, often the requirements include the development of data management plans (taking into account issues such as data curation, maintenance, preservation) and the development of metadata.

Requirements may be more or less specific. They may be related to the specific open access route to use, the type of legal copyright licence to use, the type of repositories to use (institutional or discipline-specific). Other requirements may be related to the release of data in specific formats to be interoperable or the development of metadata. Mandatory rules with respect to both publications and data often allow exceptions for reasons such as confidentiality, privacy and security. In addition, mandatory rules are designed in accordance with national ethical principles and copyright frameworks.

Examples of national initiatives introducing mandatory requirements have been developed in several OECD and non-member countries. In Belgium, the Wallonia-Brussels Federation regional funding agency regulation requires researchers to archive research outputs in institutional repositories. The Spanish National Plan for Scientific and Technological Research and Innovation requires the use of thematic or institutional repositories, no later than 12 months after publication. The US Office of Science and Technology Policy (OSTP) memorandum mandates the development of plans to increase public access to publications and digital data resulting from funded research, which include long-term preservation issues. The UK Research Council developed guidelines stating that publicly funded research must be available, preferably by means of gold open access, but green open access is an acceptable option. The UK Research Council guidelines also include specification of which copyright licence to use in the case of gold open access.

In Finland, the Open Science and Research Initiative (ATT) developed by the Ministry of Education and Culture is considering adoption of mandatory rules to promote open science. In addition, the Academy of Finland currently recommends open access publishing whenever possible. The Swiss National Science Foundation requires grantees to make articles or books available in discipline-specific or institutional repositories. Both Finnish and Swiss funding agencies accept either green or gold open access publishing.

The Japan Science and Technology Agency recommends open access publishing of the research it funds, via institutional repositories (green open access). The Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) mandated open access publishing of PhD theses, by ministerial decree.

The carrots: Funding and incentive mechanisms

Open science can be supported by defining the right incentive mechanisms to promote open behaviours in science and research. Incentive mechanisms may target different aspects of research processes, including for example financial support to open science efforts, acknowledgment and reward of researchers undertaking open science actions, or the use of new and broader evaluation metrics that take into account open science and its impact. So far in both OECD and non-member countries, financial support for open science efforts has been implemented more often than other types of incentives.

Acknowledgment of open science-related activities in the evaluation of researchers, for example for grant allocations or career advancement procedures, may be a powerful way to promote open science efforts. In most countries the existing framework does not promote sharing efforts, especially with respect to results, data sets or other research material at the pre-publishing stage. Science is a competitive process and researchers may fear unethical behaviour or simply may not properly take advantage of open science opportunities if not properly acknowledged and officially recognised. Surveys related to researchers' attitudes and behaviours may help in understanding how to design incentive mechanisms more effectively (See Box 4.12). In Chile for example, the national open access policy has been monitored by an evaluation committee of international experts that took into account, among other variables, the perception of higher education students, academics, researchers and editors around the policy.

The few incentives currently existing with respect to sharing data are even more pronounced. The perceived academic impact of an individual or an institution has great influence on the distribution of research funding (i.e. success or failure of research grant proposals); it also affects advancement of the individual's academic career and the status of the institution. Therefore, measures of academic impact used by research funding agencies are strongly shaping researchers' actions and decisions, encouraging or discouraging certain behaviours.

Currently, measures of academic impact focus on publications in academic journals, while sharing (publishing) data is generally not taken into account. Researchers are rewarded for publishing results obtained from analysis of data sets, rather than for publishing the data sets themselves. This may create situations in which researchers are willing to protect their data sets rather than publish them. On the other hand, some high-profile journals (such as *Nature* and *Science*) are requiring availability of all data sets that are necessary to understand, assess, replicate and build upon the published research results.[55]

Furthermore, data gathering, cleaning, validation and curation are time-consuming tasks and researchers often want to fully exploit the opportunities arising from the data they collect and generate in order to maximise the results they can derive from a single set of data. If no acknowledgment of these activities is in place, then there is little incentive for researchers to share data or to make it widely available

[55] See for example *Nature*: www.nature.com/authors/policies/availability.html, or *Science*.

to the research community by developing metadata and other necessary information for reuse. In addition to recognising these tasks in career advancement procedures or grant allocations, it has been proposed by the research community to develop data citation tools to credit the authors of data sets and metadata (Box 3.4).

Box 4.12 What do scientific authors think of open science?

The OECD NESTI survey on the behaviour of scientists

In order to help address some major evidence gaps, the OECD Working Party of National Experts on Science and Technology Indicators (NESTI) is currently in the process of experimenting with a new survey instrument targeting scientific authors worldwide publishing in scholarly journals about their experiences, to capture information that cannot be gauged from simple bibliometric analysis. Other related studies with a more narrow geographical scope include the EU-funded studies of Open Access Publishing (SOAP) and Permanent Access to the Records of Science in Europe (PARSE.Insight), a survey on Norwegian researchers carried out by DAMVAD (see below) on attitudes to sharing and archiving of publicly funded research data, and a survey on open access by the publishers Taylor & Francis (see below).

The approach adopted in the OECD pilot builds on its independence as an organisation and global perspective. The survey seeks to collect information on the inputs and outputs of the scientific process, addressing a number of specific questions:

- The role of peer-reviewed publications as a source of knowledge, and the effect of restricted access to documents and other research materials experienced.
- The funding of the research work and the types of conditions for use of its research outputs.
- The dissemination strategies adopted by authors and their teams and the access status to documents and other outputs such as data. There is an evidence gap concerning the actual modes of dissemination applied to documents within each type of OA model (e.g. who is paying under gold OA, the embargo length for green OA).
- The perceived added value of the peer review and editorial process, and preferences regarding the trade-off between access, impacts (in its multiple forms) and costs.

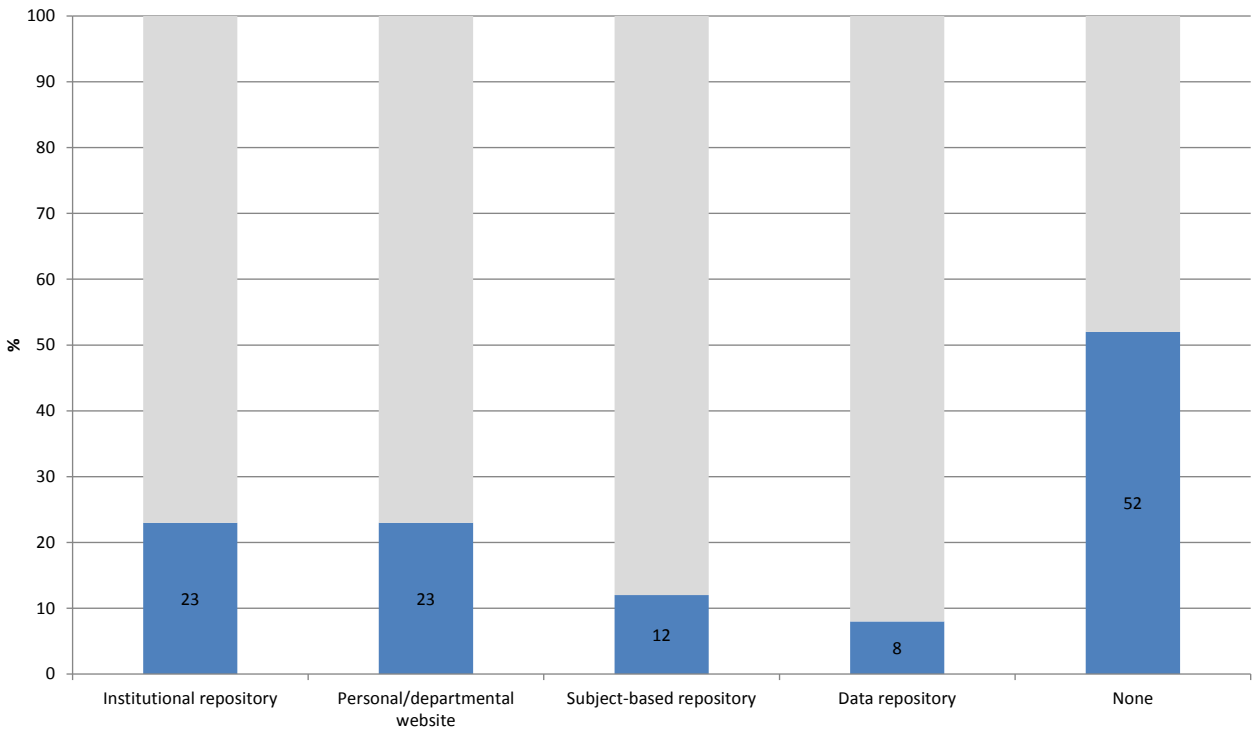
The questions in the pilot survey aim to provide reliable statistical evidence that can be used to promote a better informed debate as well as an input in the policy appraisal of different options concerning access. More information on this “proof of concept” pilot OECD survey is available on the dedicated webpage www.oecd.org/sti/survey-of-scientific-authors.htm. Further questions on this survey can be addressed to surveyonscience@oecd.org.

Evidence from the Taylor and Francis Open Access Survey

According to a recent survey conducted by Taylor and Francis (Taylor and Francis, Open Access Survey 2014), scientific authors tend to acknowledge some of the advantages offered by open access publishing, such as a wider circulation and higher visibility of articles. They also state that the articles they download from repositories are generally useful for the research they conduct. However, contrary to the evidence, the authors do not tend to believe that open access drives innovation in research or that open access publications are cited more heavily than traditional ones. Researchers also appear to be against reuse of their work for commercial gain, without prior knowledge of permission; whereas they are in favour of the re-utilisation of their results for non-commercial ends. They also have a preference for their work not to be adapted by others. More than half of the researchers who participated in the survey do not use repositories to make their papers available to others. When they do, the repositories mostly used are either institutional repositories or their personal or university webpages. The main driver to make research articles available

in repositories is personal willingness or the responsibility of making research results available to others. Requirements from funders seem to be a less important factor. The majority of researchers think that citations will still be the main metric to evaluate research output in the future, and do not attribute great importance to alternative or usage metrics (downloads).

Figure 4.1 The usage of repositories to disseminate research articles



Source: Taylor and Francis (2014), Open Access Survey 2014, June, webpage www.oecd.org/sti/survey-of-scientific-authors.htm.

Main findings of the DAMVAD Norway report to the Research Council of Norway

A 2014 survey commissioned by the Research Council of Norway on the attitude of researchers towards open research data covered a total of 1 474 researchers affiliated with many Norwegian research organisations. The survey obtained a response rate of 30.6%. It found that Norwegian researchers frequently use and share research data: 64% of the respondents had used research data from other researchers in the past three years. Most of the respondents obtained the data from other researchers in the same institution. Around 80% of the respondents agreed that open access to data improves the research process, 77% agreed that open research data facilitate education, and 74% agreed that they promote collaborative research. However, survey respondents recognised the following barriers *vis-à-vis* data sharing: i) the preparation of the data set for the release is time-consuming, ii) the technical infrastructure is inadequate, iii) open research data may reduce the options of producing further publications in the future. No significant difference across research domains or years of professional experience was found.

Source: Taylor and Francis (2014), "Open Access Survey 2014", June, webpage www.oecd.org/sti/survey-of-scientific-authors.htm; DAMVAD (2014), "Sharing and archiving of publicly funded research data", Report to the Research Council of Norway.

Examples of initiatives developed to create the incentives for open science efforts include the following. The open science national initiative in Finland (ATT) will encourage the reference to data and methods as well as the reward of teams and researchers undertaking open science efforts. In Belgium,

funding agencies in Wallonia applied the University of Liège model (Box 4.3), that is the use of institutional green open access repositories to evaluate and award researchers affiliated with that institution. In Spain, the introduction of criteria within researchers' evaluation procedures that take into account open science efforts is currently under discussion but not yet implemented. In the United States, statistics on the number of publications and downloads from the PubMed Central repository demonstrate the visibility that research papers can gain from public access. In addition, NIH regularly assigns unique identification numbers to data sets so that they may be properly cited.

The enablers: Infrastructure, skills for open science, legal frameworks

Infrastructure

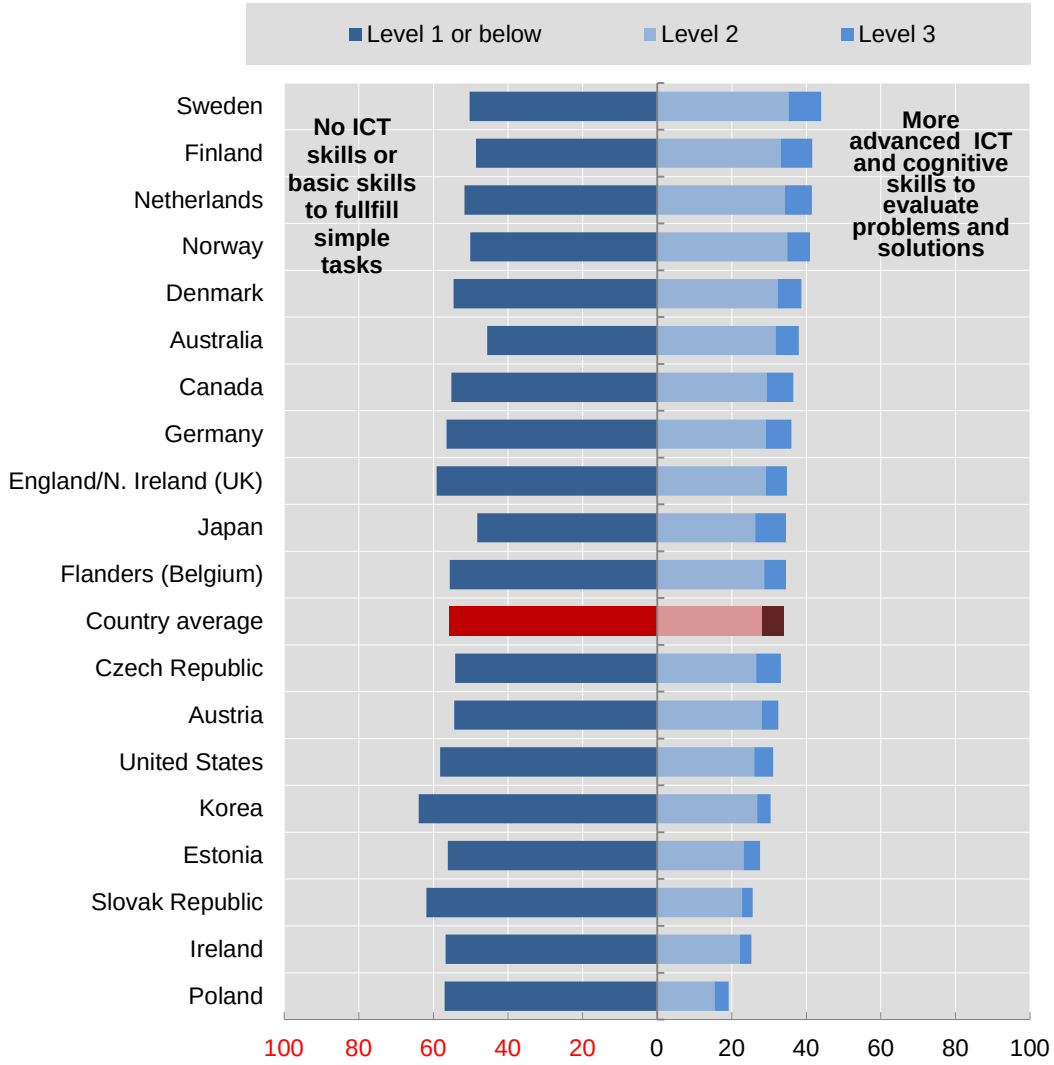
A majority of OECD and non-member countries have been investing in the infrastructure needed to promote open science. This includes online repositories, databases, archives and digital libraries and platforms containing information on R&D projects and researchers' CVs. Finland has launched an infrastructure roadmap to promote open science. China has developed online platforms for data and publication archiving. Argentina has developed the SICyTAR database with information on the CVs, publications and affiliations of researchers. Both Chile and Mexico have invested in several national repositories for sharing scientific articles and data. In Spain, RECOLECTA is the national repository and main infrastructure that allows researchers and other stakeholders to freely archive and access research publications. In Poland, several online portals and virtual libraries have been developed to facilitate open access archiving. In the United Kingdom, the E-infrastructure Leadership Council is the single co-ordinating body responsible for the UK e-infrastructure strategy, and advises BIS. The European Commission has also been active in promoting the development of EU and member country repositories and platforms (see Box 4.6). This infrastructure not only makes it possible to share scientific material, but also promotes open collaboration and the development of an open science culture. The Japan Science and Technology Agency has developed J-STAGE, an open online platform for publishers where it is possible to submit papers, manage peer-review and publish articles online.

Skills

Additional enabling factors are related to the necessary skills to allow different actors and stakeholders to use and create open science tools. Several studies have suggested that the shortage of data-related skills constitutes one of the main barriers to making use of data analytics (*Economist Intelligence Unit*, 2012; OECD, 2015a, Chapter 5). Other studies have found that the mismatch between supply and demand for specialist skills is considerable, both in the United States (McKinsey Global Institute, 2011) and in Europe (OECD 2014c). Data science-related skills are unevenly distributed across countries, as suggested by some indicators related to problem solving proficiency in technology-rich environments (Figure 4.2). As a consequence, the skills enabling open science efforts need to be identified and strengthened in education curricula. In addition, training and skills development may target scientists and researchers along all education and working cycles. Not only students but also adult researchers and scientists need to have or acquire the necessary skills to use open access repositories and archives, learn to develop metadata, and clean and maintain the data they produce.

Figure 4.2 Adult population by level of proficiency in problem solving in technology-rich environments, 2012

As a percentage of 16-65 year-olds



1. Problem solving in technology-rich environments requires "computer literacy" skills (i.e. the capacity to use ICT tools and applications) and the cognitive skills required to solve problems.

2. The OECD Survey of Adult Skills as part of the OECD Programme for the International Assessment of Adult Competencies (PIAAC) assesses the proficiency of adults aged 16-65 in literacy, numeracy and problem solving in technology-rich environments. It collects in particular a range of information on the use of information and communication technologies at work and in everyday life, and on a range of generic skills, such as collaborating with others and organising one's time.

Source: OECD (2013a), *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264204256-en.

Recent initiatives targeting the development of skills for open science include the Data Management Guide developed by the Finnish Ministry of Education and Culture; the Guide covers issues related to data management and describes existing services for data management available in the country. In Finland, training sessions will be launched in higher education institutions to train researchers and students in data management and data ownership. In Poland, a new research centre on big data, OCEAN, is developed by

the Interdisciplinary Center for Mathematical and Computational Modelling of the University of Warsaw and it is co-funded by the National Centre for Research and Development, with the aim to not only provide the e-infrastructure to store data but also the expertise and training for big data analytics. In addition, Poland is one of the partners in a recently started project, FOSTER, which aims to support young researchers in adopting the open access approach and in complying with the open access policies. The project intends to establish a European-wide training programme on open access and open data, consolidating training activities across Europe. The project also involves partners from Denmark, France, Germany, the Netherlands, Portugal, Spain and the United Kingdom.

France recently established several initiatives to develop open science-related skills. The national Digital Scientific Library created a working group focusing on the new professions around open access and open data. The Centre for the Direct Scientific Communication (CCSD) offers training in the national open access platform (HAL). Finally, seven regional training units offer classes to researchers and PhD students on open access and open data research.

In the United States, the OSTP memo highlights the importance of supporting training, education and workforce development related to scientific data management, analysis, storage, preservation, and stewardship. Within the NIH alone, the development of skills for data analytics is one of the four main pillars of the big data to Knowledge initiative, having the aim to develop further skills in the science of big data in addition to further developing skills in data usage and analysis in the biomedical field. In addition, a number of NIH Institutes and Centres offer funding for training in biomedical informatics.

Under the Action Plan on Open Government 2.0, the government of Canada will maximise access to federally funded scientific research, to encourage greater collaboration and engagement with the scientific community, the private sector, and the public. Deliverables to be completed by 2016 include: building a profile of Canadian's digital skills competencies and increasing understanding of the relationship between digital skills and labour market and social outcomes; developing online tools and training materials to improve the digital skills of individuals; funding private sector and civil society initiatives aimed at improving digital skills.

In the UK, the Data Capability Strategy focuses on (among other issues) human capital and skill development for data analytics as well as data accessibility and data-sharing skills in consumers, business and academia. In addition, the creation of centres for doctoral training on big data has been announced in several universities and higher education institutions in the country. With similar goals, the Open Data Institute has been recently established in London, United Kingdom (Box 4.13). The Indian National Informatics Centre has organised several workshops to promote open science awareness among academics. In Germany the Helmholtz Association, a partnership of 18 scientific-technical and biological-medical research centres, regularly provides training and classes on the maintenance and management of research data to promote an open science culture among researchers.

Box 4.13 The Open Data Institute, London, United Kingdom

The Open Data Institute (ODI) is a non-profit organisation largely funded by the UK Government, located in London. The goal of ODI is to analyse and promote an open data culture to create economic, environmental and social value.

ODI works together with data-intensive start-ups, organises events to promote open data, and provides training to develop data-related skills. ODI develops ties with different stakeholders, from the business sector to academia and the public sector. ODI delivers a variety of data-related services to early-stage start-ups, micro-businesses and SMEs to exploit the potential arising from open data. ODI will train a number of open data technologists and entrepreneurs and ODI staff works directly within firms to make sure that they exploit all possible benefits and opportunities from open data.

In addition to the delivery of data-related services, ODI commissions research from top universities working on open data and aims to provide guidance to the government on the opportunity arising from open data for the public sector. ODI is also active in developing international collaboration around the theme of open data.

ODI has developed the Data as Culture programme to engage artists and individuals that use data as an art material; they commission and exhibit data artwork. The programme aims to disseminate the power of data exploitation outside the usual networks of stakeholders. and reach a broader community of individuals.

Source: www.theodi.org.

Legal frameworks

Open science-friendly legal frameworks are additional means to promote open science efforts. Some countries are currently discussing modification of intellectual property rules for research or exemptions. For example, Australia and Finland are currently discussing modifications of the existing legal framework around the publication of publicly funded research results, to make the copyright legislation increasingly open science-friendly. In Germany, the national copyright act was modified in 2013 to allow publicly funded scientists and researchers to retain the legal right to upload their publications on line, even if they have transferred their exploitation rights to the publishers, after an embargo period of up to 12 months. The United Kingdom has recently passed a series of amendments to its legal framework for copyright (that came into force in 2014), which include greater freedom of reuse of copied or recorded material for education and non-commercial research purposes. For more information on recent policy trends around open science and legal frameworks, see OECD, 2015b.

Global open science

The governance of open science from an international and, in some cases, global perspective is facilitated by international governmental organisations (IGOs). IGOs play a critical role in promoting inter-governmental co-ordination at international level and in shaping the political agenda through the development of guidelines and principles around specific themes, to be subsequently adopted and implemented by member countries and beyond.

IGOs such as the OECD, UNESCO, the EU (Box 4.14) and the World Bank have been active in recent years in promoting open science efforts of member and, in some cases, non-member countries. In June 2013, the G8 Science Ministers issued a statement that, among other subjects, acknowledge the importance of open research data and open access to publicly funded research articles (G8 Science Ministers Statement, June 2013[56]).

The OECD has been active in developing guidelines and principles on open science-related themes, including access to research data (see Boxes 3.2 and 3.3) or public sector information (OECD, 2008, 2013b). The OECD, together with WIPO and other international organisations, had a fundamental role in

[56] www.gov.uk/government/news/g8-science-ministers-statement.

developing the new Creative Commons license for IGOs (see Box 2.4). Many OECD committees and working parties have worked on issues related to open data and open science. For example, the potential of open science and open data efforts to advance research to address Alzheimer's disease and dementia has been recently highlighted by the OECD expert consultation "Unlocking Global Collaboration to Accelerate Innovation for Alzheimer's Disease and Dementia" (OECD, 2014a) as well as in other medical areas. At European level, the European Union has adopted and promoted open science efforts in the most recent Framework Programme for Research and Innovation; in particular, Horizon 2020 requests open access to research publications sponsored through this programme (Box 4.14).

The United Nation Educational, Scientific and Cultural Organisation (UNESCO) is committed to promoting open access to scientific material of publicly funded research organisations. UNESCO conducts actions to raise awareness and promote open access practices among policy makers, researchers and knowledge managers. Actions are promoted in co-operation with the global network of regional and local offices of the organisation. UNESCO devotes special attention to the benefits arising from open access to African countries and other developing countries, where open access efforts are less developed. UNESCO developed a set of policy guidelines on the adoption and implementation of open access efforts (UNESCO, 2012), in order to provide information to UNESCO member countries; to help member countries choose the open access policy that best suits their specific context; and to promote adoption of open access policies.

Finally, the World Bank has adopted a fully open access internal policy for the publications it produces, and is considerably advanced in providing data openly to all possible users and stakeholders.

Box 4.14 Open science in Horizon2020

The 2014-20 European programme for science, research and innovation – Horizon 2020 – is committed to supporting open science in several ways. European open access policy brings together elements from several policy efforts: the Digital Agenda for Europe, the Innovation Union Policy, the communication "A Reinforced European Research Area Partnership for Excellence and Growth". Openness is a key principle in Horizon 2020: researchers receiving grants from Horizon 2020 must deposit a machine-readable electronic copy of the published version or peer-reviewed manuscript accepted for publication in an open repository, although there is no specification of timing or embargo periods. OpenAIRE infrastructure is recommended (see Box 4.6). Authors of scientific publications are free to choose how they wish to share research results: both green and gold open access mechanisms are accepted. Costs incurred in open access publishing are eligible for reimbursement from Horizon 2020 grants.

In addition, Horizon 2020 includes a pilot project on open research data. Researchers involved in projects participating in the pilot will be asked to make publicly available the data forming the basis of the project research results, to be used by other researchers and projects, innovative industries and citizens, as well as to develop data management plans. Over 2014-15 the Open Research Data Pilot will receive around EUR 3 billion. The pilot project on open data targets all key thematic areas of Horizon 2020 (Future and Emerging Technologies, Research Infrastructure, Leadership in Enabling and Industrial Technologies, Societal Challenges, Science with and for Society).

Researchers participating in the pilot have the possibility to opt out of the pilot to protect intellectual property or personal data; for security concerns; or if the main objective of their research can be compromised by making data openly accessible. The pilot has the aim of providing a better understanding of what supporting infrastructure is needed, as well as the role of limiting factors such as security, privacy or data protection, or any other reason that will induce researchers to opt out of the pilot. The pilot also aims to contribute to a better understanding of the best mechanisms to define the right incentives for researchers to curate and share the research data they produce. The pilot will be closely monitored during the implementation phase of Horizon 2020.

Source: http://europa.eu/rapid/press-release_IP-13-1257_en.htm; http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

REFERENCES

Bowser, A. and L. Shanley (2013), *New Visions in Citizen Science*, Commons Lab, Science and Technology Innovation Program, Woodrow Wilson International Center for Scholars, November.

DAMVAD (2014), "Sharing and archiving of publicly funded research data", Report to the Research Council of Norway.

Davies, L. et al. (2011), "Open Air Laboratories (OPAL): A community-driven research programme".

Economist Intelligence Unit (2012), "The deciding factor: Big data & decision making", commissioned by Capgemini, 4 June, www.capgemini.com/insights-and-resources/by-publication/the-deciding-factor-big-data-decision-making/.

Foreign & Commonwealth Office, United Kingdom (2013), G8 Science Ministers Statement, www.gov.uk/government/news/g8-science-ministers-statement, accessed 18 June 2015.

Gura, T. (2013), "Citizen science: Amateur experts", *Nature*, Vol. 496, pp. 259-61, http://dx.doi.org/10.1038/nj7444-259a.

Holocher-Ertl, T. and B. Kieslinger (2013), "Green paper on citizen science citizen science for Europe: Towards a better society of empowered citizens and enhanced research", SOCIENTIZE Project, http://ec.europa.eu/digital-agenda/en/news/green-paper-citizen-science-europe-towards-society-empowered-citizens-and-enhanced-research-0, accessed 18 June 2015.

Lane, J. et al. (eds.) (2014), *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, Cambridge University Press.

McKinsey Global Institute (2011), "Big data: The next frontier for innovation, competition and productivity",McKinsey & Company,www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation, accessed 18 June 2015.

OECD (2015a), *Data-Driven Innovation for Growth and Well-Being*, OECD Publishing, Paris.

OECD (2015b), *Inquiries into Intellectual Property's Economic Impact*, OECD Publishing, Paris.

OECD (2014a), "Unleashing the power of big data for Alzheimer's disease and dementia research: Main Points from the OECD Expert Consultation on Unlocking Global Collaboration to Accelerate Innovation for Alzheimer's Disease and Dementia", OECD Publishing, Paris, http://dx.doi.org/10.1787/20716826.

OECD (2014b), *Science, Technology and Industry Outlook 2014*, OECD Publishing, Paris, http://dx.doi.org/10.1787/sti_outlook-2014-en.

OECD (2014c), *Measuring the Digital Economy: A New Perspective*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264221796-en.

OECD (2013a), *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264204256-en.

OECD (2013b), "Public sector information: A review of the Recommendation", Working Party on the Information Economy, DSTI/ICCP/IE(2012)2/REV2.

OECD (2008), *OECD Recommendations of the Council for Enhanced Access and More Effective Use of Public Sector Information*, OECD Publishing, Paris.

Riesch, H., C. Potter and L. Davies (2013), "Combining citizen science and public engagement: The Open Air Laboratories Programme", *Journal of Science Communication*, Vol. 12, No. 3.

Science-Metrix (2013), *Open Data Access Policies and Strategies in the European Research Area and Beyond*, European Commission, 26 August, www.science-metrix.com/pdf/SM_EC_OA_Data.pdf, accessed 18 June 2015.

Taylor and Francis (2014), *Taylor and Francis Open Access Survey, June 2014*, Taylor and Francis/Routledge, www.tandf.co.uk/journals/explore/open-access-survey-june2014.pdf, accessed 18 June 2015.

Tinati, R. et al. (2014), "Collective intelligence in citizen science – A study of performers and talkers", arXiv:1406.7551, accessed 18 June 2015.

Ubaldi, B. (2013), "Open government data: Towards empirical analysis of open government data initiatives", *OECD Working Papers on Public Governance*, No. 22, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k46bj4f03s7-en.

Ware, M. and M. Mabe (2012), *The STM Report: An Overview of Scientific and Scholarly Journal Publishing*, Third Edition, November, STM: International Association of Scientific, Technical and Medical Publishers, The Hague.

Wilbanks, J. and S. Macmillan (forthcoming), "Making Energy Data Re-Useful".

BIBLIOGRAPHY

Anderson, C. (2008), "The end of theory: The data deluge makes the scientific method obsolete", *Wired*, 23 June, www.wired.com/science/discoveries/magazine/16-07/pb_theory/.

Anderson, W.L. (2004), "Some challenges and issues in managing, and preserving access to, long-lived collections of digital scientific and technical data", *Data Science Journal*, Vol. 3, 30 December, pp. 191-202, retrieved from www.jstage.jst.go.jp/article/dsj/3/0/191/_pdf.

Archambault, E. et al. (2014), "Proportion of open access papers published in peer-review journals at the European and world levels – 1996-2013", European Commission study prepared by Science Metrix, http://science-metrix.com/files/science-metrix/publications/d_1.8_sm_ec_dg-rtd_proportion_oa_1996-2013_v11p.pdf.

Arzberger, P. et al. (2004), "Promoting access to public research data for scientific, economic and social development", *Data Science Journal*, Vol. 3, November, pp. 1777-78.

Aymar, R. (2009), "Scholarly communication in high-energy physics: Past, present and future innovations", *European Review*, No. 1, pp. 33-51.

Bell, G., T. Hey and A. Szalay (2009), "Beyond the data deluge", *Science*, Vol. 323, No. 5919, 6 March, pp. 1297-1298, retrieved from www.cloudinnovation.com.au/Bell_Hey_Szalay_Science_March_2009.pdf.

BIAC (2011), "BIAC thought starter: A strategic vision for OECD work on science, technology and industry", Business and Industry Advisory Committee to the OECD, 12 October.

Björk, B.-C. et al. (2013), "Anatomy of green open access", *Journal of the American Society for Information Science and Technology*, Vol. 65, No. 2, pp. 237-50.

Björk, B.-C. and D. Solomon (2014), "Article processing charges in OA journals – Relationship between price and quality", *Scientometrics*, Vol. 103, No. 2, pp. 373-85.

Björk, B.-C. et al. (2010), "Open access to the scientific journal literature: Situation 2009", *PLOS ONE*, Vol. 5, No. 6, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0011273, 23 June, accessed 10 June 2015.

Blue Ribbon Task Force on Sustainable Digital Preservation and Access (2010), "Sustainable economics for a digital planet: Ensuring long term access to digital information", February, http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf, accessed 17 June 2015.

Bollier, D. (2010), *The Promise and Peril of Big Data*, The Aspen Institute, Washington, DC.

Bornmann, L. (2014), "Measuring the broader impact of research: The potential of altmetrics", ResearchGate, www.researchgate.net/publication/263506984_Measuring_the_broader_impact_of_research_The_potential_of_altmetrics, accessed 10 June 2015.

Bowser, A. and L. Shanley (2013), *New Visions in Citizen Science*, Commons Lab, Science and Technology Innovation Program, Woodrow Wilson International Center for Scholars, November.

Brase, J. et al. (2009), "Approach for a joint global registration agency for research data", *Information Services and Use*, Vol. 29, No. 1, pp. 13-27.

CEBR (2012), "Data equity: Unlocking the value of big data", Centre for Economic and Business Research, London, www.sas.com/offices/europe/uk/downloads/data-equity-cebr.pdf, accessed 10 June 2015.

CED (2012), *The Future of Taxpayer-Funded Research: Who Will Control Access to the Results?*, Committee for Economic Development, Washington, DC.

Chan, L., B. Kirsop and S. Arunachalam (2005), "Open access archiving: The fast track to building research capacity in developing countries", *SciDevNet*, November.

Chavan, V. and L. Penev (2011), "The data paper: A mechanism to incentivize data publishing in biodiversity science", *BMC Bioinformatics*, Vol. 12, Supplement 15, p. S2.

Clark, J. (2013), *Text Mining and Scholarly Publishing*, Publishing Research Consortium.

CODATA-ICSTI (Committee on Data for Science and Technology - International Council for Scientific and Technical Information) Task Group on Data Citation Standards and Practices (2013), "Out of cite, out of mind: The current state of practice, policy and technology for the citation of data", *Data Science Journal*, Volume 12, pp. 1-75.

Corrado, C., C. Hulten and D. Sichel (2009), "Intangible Capital and U.S. Economic Growth", *Review of Income and Wealth*, Series 55, No. 3, September, www.conference-board.org/pdf_free/IntangibleCapital_USEconomy.pdf.

Costas, R., Z. Zahedi and P. Wouters (2014), "Do 'altmetrics' correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective", http://arxiv.org/abs/1401.4321, accessed 10 June 2015.

Costas, R. et al. (2013), "The Value of Research Data – Metrics for datasets from a cultural and technical point of view", A Knowledge Exchange Report, www.knowledge-exchange.info/datametrics, accessed 11 June 2015.

Cragin, M.H. et al. (2010), "Data sharing, small science and institutional repositories", *Philosophical Transactions A of the Royal Society*, Vol. 368, No. 1926, pp. 4023-38, http://rsta.royalsocietypublishing.org/content/368/1926/4023, accessed 10 June 2015.

Criscuolo, C. and C. Menon (forthcoming), "Do patents matter for venture capital investments? Evidence from the green sector".

da Rosa, I.B. and D.R. Lamas (2013), "Easing access to digital libraries with m-DSpace", *eLearning Papers*, No. 34, October, www.openeducationeuropa.eu/sites/default/files/asset/From-field_34_1.pdf, accessed 18 June 2015.

Dallmeier-Tiessen, S. et al. (2011), "Highlights from the SOAP project survey: What scientists think about open access publishing", http://arxiv.org/abs/1101.5260, accessed 10 June 2015.

DAMVAD (2014), "Sharing and archiving of publicly funded research data", Report to the Research Council of Norway.

David, P.A. (2003), "The economic logic of "open science" and the balance between private property rights and the public domain in scientific data and information: A primer", in P. Uhlir and J. Esanu (eds.), *National Research Council on the Role of the Public Domain in Science*, National Academy Press, Washington, DC.

Davies, L. et al. (2011), "Open Air Laboratories (OPAL): A community-driven research programme".

Davis, P.M. et al. (2008), "Open access publishing, article downloads, and citations: Randomised controlled trial", *BMJ*, 2008;337:a568**,** www.bmj.com/content/337/bmj.a568, accessed 10 June 2015.

Dwork, C. and A. Roth (2014), "The algorithmic foundations of differential privacy", *Foundations and Trends in Theoretical Computer Science*, Vol. 9, Nos. 2-4, pp. 211-407, http://dx.doi.org/10.1561/0400000042.

EC (2010), "Riding the wave: How Europe can gain from the rising tide of scientific data", Final report by the High-level Expert Group on Scientific Data, European Commission, October, http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf, accessed 10 June 2015.

Economist Intelligence Unit (2012), "The deciding factor: Big data & decision making", commissioned by Capgemini, 4 June, www.capgemini.com/insights-and-resources/by-publication/the-deciding-factor-big-data-decision-making/.

Fienberg, S.E., M.E. Martin and M.L. Straf (1985), *Sharing Research Data*, National Academy Press, Washington, DC.

Filippov, S. (2014), "Mapping tech and data mining in academic and research communities in Europe", *The Lisbon Council*, Issue 16/2014.

Force11 (2012), "Improving the future of research communications and e-scholarship", Force11 white paper, www.force11.org/white_paper.

Foreign & Commonwealth Office, United Kingdom (2013), G8 Science Ministers Statement, www.gov.uk/government/news/g8-science-ministers-statement, accessed 18 June 2015.

Fradsen, T.F. (2009), "The integration of open access journals in the scholarly communication system: Three science fields", *Information Processing and Management*, Vol. 45, No. 1, pp. 131-41.

Frischmann, B.M. (2012), *Infrastructure: The Social Value of Shared Resources*, Oxford University Press.

Gardner, D. et al. (2003), "Towards effective and rewarding data sharing", *Neuroinformatics*, Vol. 1, No. 3, pp. 289-95.

Gargouri, Y. et al. (2012), "Green and gold open access percentages and growth, by discipline", Paper presented at the 17th International Conference on Science and Technology Indicators (STI), Montreal, Canada.

Gargouri, Y. et al. (2010), "Self-selected or mandated, open access increases citation impact for higher quality research", PLOS ONE, Vol. 5, No. 10, Art. No. e13636, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0013636, accessed 11 June 2015.

Geist, M. (2013), "Fairness found: How Canada quietly shifted from fair dealing to fair use", in M. Geist (ed.), *The Copyright Pentalogy*, University of Ottawa Press, Ottawa, pp. 157-186.

Gentil-Beccot, A., S. Mele and T.C. Brooks (2010), "Citing and reading behaviours in high-energy physics", *Scientometrics*, Vol. 84, Issue 2, August, http://link.springer.com/article/10.1007%2Fs11192-009-0111-1, accessed 11 June 2015.

Gervais, D. (2005), "Towards a new core international copyright norm: The Reverse Three-Step Test", *Marquette Intellectual Property Law Review*, Vol. 9, p. 1.

Goldschmidt-Clermont, L. (2002), "Communication patterns in high-energy physics", *High-Energy Physics Libraries Webzine*, Issue 6, March.

Grieneisen, M.L. and M. Zhang (2012), "A comprehensive survey of retracted articles from the scholarly literature", PLOS ONE, Vol. 7, No. 10, Art. No. e44118.

Groth, P. and T. Gurney (2010), "Studying scientific discourse on the web using bibliometrics: A chemistry blogging case study", in *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*, April 2010.

Groves, T. (2010), "The wider concept of data sharing: View from the BMJ", *Biostatistics*, Vol. 11, No. 3, pp. 391-92.

Groves, T. (2009), "Managing research data for future use", *BMJ: British Medical Journal* (Overseas & Retired Doctors Edition), Vol. 338, Issue 7697, 28 March, p. 729.

Guibault, L. (2011), "Owning the right to open up access to scientific publications", in L. Guibault and C. Angelopoulos (eds.), *Open Content Licences: From Theory to Practice*, Amsterdam University Press, Amsterdam, pp. 137-167.

Guibault, L. and T. Margoni (2014), "Legal aspects of open science and open data", Background paper for the OECD CSTP/TIP project on open science, Instituut voor Informatierecht, University of Amsterdam.

Gura, T. (2013), "Citizen science: Amateur experts", *Nature*, Vol. 496, pp. 259-61, http://dx.doi.org/10.1038/nj7444-259a.

Hardisty, D.J. and D.A.F. Haaga (2008), "Diffusion of treatment research: Does open access matter?", *Journal of Clinical Psychology*, Vol. 67, No. 7.

Heuer, R.D., A. Holtkamp and S. Mele (2008), "Innovation in scholarly communication: Vision and projects from high-energy physics", *Information Services and Use*, Vol. 28, No. 2, pp. 83-96.

Hey, T. and A. Trefethen (2003), "The data deluge: An e-science perspective", in F. Berman, G.C. Fox and A.J.G. Hey (eds.), *Grid Computing: Making the Global Infrastructure a Reality*, John

Wiley & Sons, Ltd., Chichester, England, pp. 809-24, retrieved from http://eprints.ecs.soton.ac.uk/7648/1/The_Data_Deluge.pdf.

Hilty, R. et al. (2013), "Zum Referentenentwurf eines Gesetzes zur Einführung einer Regelung zur Nutzung verwaister Werke und weiterer Änderungen des Urheberrechtsgesetzes sowie des Urheberrechts-wahrnehmungsgesetz", Stellungnahme des Max-Planck-Instituts für Immaterialgüter- und Wettbewerbsrecht zur Anfrage des Bundesministeriums der Justiz vom 20 February 2013, www.ip.mpg.de/files/pdf2/Stellungnahme-BMJ-UrhG_2013-3-15-def1.pdf.

Holocher-Ertl, T. and B. Kieslinger (2013), "Green paper on citizen science citizen science for Europe: Towards a better society of empowered citizens and enhanced research", SOCIENTIZE Project, http://ec.europa.eu/digital-agenda/en/news/green-paper-citizen-science-europe-towards-society-empowered-citizens-and-enhanced-research-0, accessed 18 June 2015.

Houghton, J. and P. Sheehan (2009), "Estimating the potential impacts of open access to research findings", *Economic Analysis and Policy*, Vol. 29, No. 1, pp. 127-42.

Houghton, J. and A. Swan (2013), "Planting the green seeds for a golden harvest: Comments and clarification on going for gold", *D-Lib Magazine*, Vol. 19, No. 1/2.

Houghton, J., B. Rasmussen and P. Sheehan (2010), "Economic and social returns on investment in open archiving publicly funded research outputs", Report to the Scholarly Publishing and Academic Resources Coalition (SPARC), Center for Strategic Economic Studies, Victoria University.

Houghton, J., A. Swan and S. Brown (2011), "Access to research and technical information in Denmark", Technical Report, School of Electronics and Computer Science, University of Southampton.

Iida, K. et al. (2011), "Question Q216B exceptions to copyright protection and the permitted uses of copyright works in the hi-tech and digital sectors", the Japanese Group of AIPPI (Association Internationale pour la Protection de la Propriété Industrielle), p. 9.

Jirotka, M. et al. (2006), "Collaboration in e-research", *Computer Supported Cooperative Work (CSCW)*, Special Issue, Vol. 15, No. 4, pp. 251-55, http://dx.doi.org/10.1007/s10606-006-9028-x.

JISC (2014), "The value and impact of data sharing and curation: A synthesis of three recent studies of UK research data centres", JISC, March, http://www.cni.org/news/jisc-report-value-impact-of-data-curation-and-sharing/, accessed 11 June 2015.

JISC (2012), "The value and benefits of text mining", JISC, www.jisc.ac.uk/sites/default/files/value-text-mining.pdf, accessed 11 June 2015.

Johnson, R. (2015), "Making open access work for authors, institutions and publishers", Report on an Open Access Roundtable hosted by Copyright Clearance Center, Inc., www.copyright.com/content/dam/cc3/marketing/documents/pdfs/Report-Making-Open-Access-Work.pdf, accessed 16 June 2015.

Kotarski, R. et al. (2012), "Report on best practices for citability of data and on evolving roles in scholarly communication", *Opportunities for Data Exchange*, www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/08/ODE-ReportBestPracticesCitabilityDataEvolvingRolesScholarlyCommunication.pdf.

Kowalczyk, S. and K. Shankar (2010), "Data sharing in the sciences", *Annual Review of Information Science and Technology*, No. 45, pp. 247-94.

Laakso, M. (2014), "Green open access policies of scholarly journal publishers: A study of what, when, and where self-archiving is allowed", *Scientometrics*, Vol. 99, No. 2, pp. 475-94, http://dx.doi.org/10.1007/s11192-013-1205-3, accessed 10 June 2015.

Laakso, M. and B.-C. Björk (2012), "Anatomy of open access publishing: A study of longitudinal development and internal structure", *BMC Medicine*, Vol. 10, pp. 124, http://www.biomedcentral.com/1741-7015/10/124, accessed 11 June 2015.

Lakhani, K.R. et al. (2007), "The value of openness in scientific problem solving", HBS Working Paper No. 07-050, Harvard Business School, http://hbswk.hbs.edu/item/5612.html, accessed 11 June 2015.

Lane, J. et al. (eds.) (2014), *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, Cambridge University Press.

Lansingh, V.C. and M.J. Carter (2009), "Does open access in ophthalmology affect how articles are subsequently cited in research", *Ophthalmology*, Vol. 116, No. 8, pp. 1425-31.

LERU (2013), "LERU Roadmap for Research Data", LERU Research Data Working Group, League of European Research Universities, Advice Paper No. 14, December, www.uzh.ch/research/LERU_Roadmap_for_Research_data.pdf, accessed 19 June 2015.

Lewis, D.W. (2012), "The inevitability of open access", *College and Research Libraries*, Vol. 73, No. 5, pp. 493-506.

Loshin, D. (2002), "Knowledge integrity: Data ownership", 8 June, www.datawarehouse.com/article/?articleid=3052.

McCabe, M. and C.M. Snyder (2014), "Identifying the effect of open access on citations using a panel of science journals", *Economic Inquiry*, Vol. 42, No. 4, pp. 1284-1300.

McKinsey Global Institute (2011), "Big data: The next frontier for innovation, competition and productivity", McKinsey & Company, www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation, accessed 18 June 2015.

Merton, R.K. (1973), *The Sociology of Science: Theoretical and Empirical Investigations*, University of Chicago Press.

Miguel, S., Z. Chichilla-Rodrígues and F. de Moya-Anegón (2011), "Open access and scopus: A new approach to scientific visibility from the standpoint of access", *Journal of the American Society for Information Science and Technology*, Vol. 62, No. 6, pp. 1130-45.

Mivule, K. (2013), "Utilizing noise addition for data privacy: An overview", *Proceedings of the International Conference on Information and Knowledge Engineering (IKE 2012)*, pp. 65-71, http://arxiv.org/pdf/1309.3958.pdf, accessed 17 June 2015.

Mooney, H. and M.P. Newton (2012), "The anatomy of a data citation: Discovery, reuse, and credit", *Journal of Librarianship and Scholarly Communication*, Vol. 1, No. 1, Art. No. eP1035, http://dx.doi.org/10.7710/2162-3309.1035, accessed 10 June 2015.

Moscon, V. (2013), "Open access to scientific articles: Comparing Italian with German law", Kluwer Copyright Blog, http://kluwercopyrightblog.com/2013/12/03/open-access-to-scientific-articles-comparing-italian-with-german-law, accessed 16 June 2015.

Murray, F.P. et al. (2009), "Of mice and academics: Examining the effect of openness on innovation", NBER Working Paper No. 14819, National Bureau of Economic Research, www.nber.org/papers/w14819, accessed 11 June 2015.

Narayanan, A. and V. Shmatikov (2008), "Robust de-anonymization of large sparse datasets", *SP 08 Proceedings of the 2008 IEEE Symposium on Security and Privacy*, IEEE Computer Society, Washington, DC.

Narayanan, A. and V. Shmatikov (2007), "Robust De-anonymization of Large Datasets (How To Break Anonymity of the Netflix Prize Dataset)", University of Texas at Austin, 22 November, http://arxiv.org/abs/cs/0610105v2.

Neylon, C. and S. Wu (2009), "Article-level metrics and the evolution of scientific impact", PLOS Biology, 2009, Vol. 7, No. 11, Art. No. e1000242, http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1000242, accessed 11 June 2015.

OECD (2015a), *Data-Driven Innovation for Growth and Well-Being*, OECD Publishing, Paris.

OECD (2015b), *Inquiries into Intellectual Property's Economic Impact*, OECD Publishing, Paris.

OECD (2014a), "Unleashing the power of big data for Alzheimer's disease and dementia research: Main Points from the OECD Expert Consultation on Unlocking Global Collaboration to Accelerate Innovation for Alzheimer's Disease and Dementia", OECD Publishing, Paris, http://dx.doi.org/10.1787/20716826.

OECD (2014b), *Science, Technology and Industry Outlook 2014*, OECD Publishing, Paris.

OECD (2014c), *Measuring the Digital Economy: A New Perspective*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264221796-en.

OECD (2013a), "Strengthening health information infrastructure for health care quality governance", in *OECD Health Policy Studies Series 2013*, OECD Publishing, Paris.

OECD (2013b), "New data for understanding the human condition: International Perspectives", OECD Global Science Forum Report on Data and Research Infrastructure for the Social Sciences, OECD Publishing, Paris, www.oecd.org/sti/sci-tech/new-data-for-understanding-the-human-condition.pdf, accessed 17 June 2015.

OECD (2013c), "Public sector information: A review of the Recommendation", Working Party on the Information Economy, DSTI/ICCP/IE(2012)2/REV2.

OECD (2013d), Background paper for the TIP workshop on Open Science and Open Data, unpublished, DSTI/STP/TIP(2013)13.

OECD (2013e), "Knowledge Networks and Markets", *OECD Science, Technology and Industry Policy Papers*, No. 7, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k44wzw9q5zv-en, accessed 11 June 2015.

OECD (2013f), *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264204256-en.

OECD (2011), "Quality Framework and Guidelines for OECD Statistical Activities", 17 January, www.oecd.org/std/qualityframeworkforoecdstatisticalactivities.htm, accessed 19 June 2015.

OECD (2008), *OECD Recommendations of the Council for Enhanced Access and More Effective Use of Public Sector Information*, OECD Publishing, Paris.

OECD (2007), *OECD Principles and Guidelines for Access to Research Data from Public Funding*, OECD Publishing, Paris, www.oecd.org/sti/sci-tech/38500813.pdf, accessed 19 June 2015.

OECD (1997), *Recommendation of the Council concerning Guidelines for Cryptography Policy*, 27 March, C(97)62/FINAL, OECD, Paris, www.oecd.org/internet/ieconomy/guidelinesforcryptographypolicy. htm

Ohm, P. (2009), "The rise and fall of invasive ISP surveillance", *University of Illinois Law Review*, 2009 Volume, No. 5, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1261344, accessed 17 June 2015.

OpCit Project (2012), Open Citation (OPCIT) Project.

Open Society Institute (2005), *Open Access Publishing and Scholarly Societies – A Guide*, OSI, New York, p. 6.

Paltoo, D.N. et al. (2014), "Data use under the NIH GWAS Data Sharing Policy and future directions", *Nature Genetics*, Vol. 46, No. 9, September.

Pfitzmann, A. and M. Hansen (2010), "A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management", v0.34, 10 August, http://dud.inf.tu-dresden.de/Anon_Terminology.shtml.

Piwowar, H.A. and W.W. Chapman (2008), "Identifying data sharing in biomedical literature", *AMIA Annual Symposium Proceeding Archive*, pp. 596-600.

Piwowar, H. and T.J. Vision (2013), "Data reuse and the open data citation advantage", *PeerJ*, 1:e175, http://dx.doi.org/10.7717/peerj.175.

Piwowar, H., R.S. Day and D.B. Frisma (2007), "Sharing detailed research data is associated with increased citation rate", PLOS ONE, Vol. 2, No. 3.

Priem, J. and B.M. Hemminger (2010), "Scientometrics 2.0: Toward new metrics of scholarly impact on the social web", *First Monday*, Vol. 15, No. 7, 5 July.

Priem, J. et al., (2010), "Altmetrics: A manifesto", www.altmetrics.org/manifesto, accessed 11 June 2015.

Reichman, J.H. and R.L. Okediji (2012), "When copyright law and science collide: Empowering digitally integrated research methods on a global scale", *Minnesota Law Review*, Vol. 96, p. 1362-1480.

Reilly, S. et al. (2011), "Report on integration of data and publications", ODE – Opportunities for Data Exchange, www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/11/ODE-ReportOnIntegrationOfDataAndPublications-1_1.pdf.

Ricketson, S. (2003), "WIPO study on limitations and exceptions of copyright and related rights in the digital environment", World Intellectual Property Organization, www.wipo.int/edocs/mdocs/copyright/en/sccr_9/sccr_9_7.pdf, accessed 16 June 2015.

Riesch, H., C. Potter and L. Davies (2013), "Combining citizen science and public engagement: The Open Air Laboratories Programme", *Journal of Science Communication*, Vol. 12, No. 3.

Romeu, C. et al. (2014), "The SCOAP3 initiative and the open access article-processing-charge market: Global partnership and competition improve value in the dissemination of science", http://cds.cern.ch/record/1735210/files/?ln=it, accessed 11 June 2015.

Rowlands, I. and D. Nicholas (2005), "Scholarly communication in the digital environment: The 2005 survey of journal author behaviour and attitudes", *Aslib Proceedings*, Vol. 57, Issue 6, pp. 481-97.

Royal Society (2012), "Final report: Science as an open enterprise", Royal Society Science Policy Centre Report 02/12, https://royalsociety.org/policy/projects/science-public-enterprise/Report/, accessed 11 June 2015.

Science-Metrix (2013), *Open Data Access Policies and Strategies in the European Research Area and Beyond*, European Commission, 26 August, www.science-metrix.com/pdf/SM_EC_OA_Data.pdf, accessed 18 June 2015.

Senftleben, M. (2004), *Copyright, Limitations and the Three-Step Test*, Kluwer Law International, London.

Sparks, S. (2005), *JISC Disciplinary Differences Report*, Rightscom, London.

Spiegler, D.B. (2007), "The private sector in meteorology: An update", *AMS Journals Online*, American Meteorological Society, http://journals.ametsoc.org/doi/pdf/10.1175/BAMS-88-8-1272, accessed 11 June 2015.

Suber, P. (2012), *Open Access*, MIT Press.

Swan, A. (2010), "The open access citation advantage: Studies and results to date", Research on Institutional Repositories (IRs), SelectedWorks, http://works.bepress.com/ir_research/31/, accessed 11 June 2015.

Synodinou, T.-E. (2012), "The foundations of the concept of work in European copyright law", in T.-E. Synodinou (ed.), *Codification of European Copyright Law – Challenges and Perspectives*, Kluwer Law International, The Hague, pp. 93-113, p. 101.

Tamura, Y. (2009), "Rethinking Copyright Institution for the Digital Age", *WIPO Journal: Analysis and Debate of Intellectual Property Issues* 1, No. 1, pp. 63-74.

Taylor and Francis (2014), *Taylor and Francis Open Access Survey, June 2014*, Taylor and Francis/Routledge, www.tandf.co.uk/journals/explore/open-access-survey-june2014.pdf, accessed 18 June 2015.

Taraborelli, D. (2008), "Soft peer review: Social software and distributed scientific evaluation", *Proceedings of the 8th International Conference on the Design of Cooperative Systems* (COOP'08), Carry-Le-Rouet, 20-23 May.

The Economist (2010), "Data, data everywhere", 27 February.

Thelwall, M. et al. (2013), "Do altmetrics work? Twitter and ten other social web services", PLOS ONE, Vol. 8, No. 5, Art. No. e64841, http://dx.doi.org/10.1371/journal.pone.0064841.

Tinati, R. et al. (2014), "Collective intelligence in citizen science – A study of performers and talkers", arXiv:1406.7551, accessed 18 June 2015.

Triaille, J. (ed.) (2013), "Study on the application of Directive 2001/29/EC on copyright and related rights in the information society", De Wolf & Partners in collaboration with CRIDS, pp. 39-40.

Ubaldi, B. (2013), "Open government data: Towards empirical analysis of open government data initiatives", *OECD Working Papers on Public Governance*, No. 22, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k46bj4f03s7-en.

Uhlir, P.F. (2012), *For Attribution – Developing Data Attribution and Citation Practices and Standards*, Summary of an International Workshop 21.

UNESCO (2012), *Policy Guidelines for the Development and Promotion of Open Access*, UNESCO Publishing.

Wagner, B. (2010), "Open Access Citation Advantage: An Annotated Bibliography", *Issues in Science and Technology Librarianship*, No. 60.

Ware, M. (2009), *Access by UK Small and Medium-Sized Enterprises to Professional and Academic Literature*, Publishing Research Consortium, Bristol.

Ware, M. and M. Mabe (2012), *The STM Report: An Overview of Scientific and Scholarly Journal Publishing*, Third Edition, November, STM: International Association of Scientific, Technical and Medical Publishers, The Hague.

Ware, M. and M. Monkman (2008), *Peer Review in Scholarly Journals: Perspective of the Scholarly Community – An International Study*, Publishing Research Consortium, Bristol.

Wilbanks, J. and S. Macmillan (forthcoming), "Making Energy Data Re-Useful".

Williams, H. (2010), "Intellectual property rights and innovation: Evidence from the human genome", NBER Working Paper 16123, National Bureau of Economic Research, July, http://www.nber.org/papers/w16213, accessed 11 June 2015.

Winickoff, D.E., K. Saha and G.D. Graff (2009), "Opening stem cell research and development: A policy proposal for the management of data, intellectual property and ethics", *Yale Journal of Health Policy, Law and Ethics*.

Wouters, P. and R. Costas (2012), *Users, Narcissism and Control: Tracking the Impact of Scholarly Publications in the 21st Century*, SURF Foundation, Utrecht.

Zahedi, Z., R. Costas and P. Wouters (2013), "How well developed are altmetrics? Cross disciplinary analysis of the presence of alternative metrics in scientific publications", *Scientometrics*, Vol. 101, Issue 2, http://dx.doi.org/10.1007/s11192-014-1264-0.