European Commission

Open Access

Summary Report—Evolution of OA Policies and Availability, 1996-2013

RTD-B6-PP-2011-2: Study to develop a set of indicators to measure open access

Prepared by

Science-Metrix

Brussels | Montreal | Washington

info@science-metrix.com | www.science-metrix.com

This document has been prepared for the European Commission; however it reflects the views of the authors alone, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

This document presents condensed information from the following reports produced as part of the European Commission Contract RTD-B6-PP-2011-2—Study to develop a set of indicators to measure open access:

Archambault, E. et al. (2014). *Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels—1996–2013.* Deliverable D.1.8. (2014 Update). Version 11b.

Archambault, E. Caruso, J., & Nicol, A, (2014) *State-of-art analysis of OA strategies to peer-review publications.* Deliverable D.2.1. (2014 Update). Version 5b.

Caruso, J., Nicol, A, & Archambault, E. (2014) *State-of-art analysis of OA strategies to scientific data.* Deliverable D.2.2. (2014 Update). Version 4b.

Caruso, J., Nicol, A, & Archambault, E. (2014) *Comparative analysis of the strengths and weaknesses of existing open access strategies.* Deliverable D.2.3. (2014 Update). Version 4b.

Campbell, D., Nicol, A, & Archambault, E. (2014) *Composite indicator to measure the growth of Open Access.* Deliverable D.3.2. Version 1.

Archambault, E. et al. (2014). *Evolution of Open Access Policies and Availability, 1996–2013.* Deliverable D.4.5. Version 5b.

Open Access

Summary Report—Evolution of OA Policies and Availability, 1996-2013

RTD-B6-PP-2011-2: Study to develop a set of indicators to measure open access

Prepared by

Science-Metrix

Brussels | Montreal | Washington

info@science-metrix.com | www.science-metrix.com

Produced and written by

Éric Archambault, Didier Amyot, Julie Caruso, David Campbell, Philippe Deschamps, Aurore Nicol, Françoise Provencher, Lise Rebout & Guillaume Roberge

J^oj]`mђ-+ 2014

Definitions: Set of definitions based on the observation of the most relevant characteristics of freely available papers.

A: Access—can be open (free), restricted or paid; with unrestricted or restricted usage rights; quality controlled or not; pre-print (pre-referring), post-print (post-referring), or published version (with final copy editing and page layout); immediate or delayed; permanent or transient.

OA: Open Access—freely available online to all.

IOA: Ideal OA—free; quality controlled (peer-reviewed or editorially controlled); with unrestricted usage rights (e.g. CC BY); in final, published form; immediate; permanent.

RA: Restricted Access—access restricted to members of a group, club, or society.

PA: Paid Access—access restricted by a pay wall; includes subscription access, licensed access, and pay-to-view access.

Restricted OA—free but with download restrictions (e.g. registration required, restricted to manual download, HTML-only as opposed to self-contained format such as PDF) or re-use rights (e.g. CC NC).

Green OA—OA provided before or immediately after publication by author self-archiving.

Gold OA—immediate OA provided by a publisher, sometimes with paid for publication fee. Note that several Gold journals have right restriction: they are Gold ROA. For example, of the 38% of journals listed in the DOAJ that use a Creative Common licence, only 53% use the CC-BY licence that would allow them to qualify for the IOA definition above (Herb, 2014).

- **Gold OA Journal**—journal offering immediate cover-to-cover access.
- **Gold OA Article**—immediately accessible paper appearing in a Gold journal, or in a PA journal (the latter is also sometimes referred to as hybrid open access).

ROA: Robin Hood OA or **Rogue OA**—available for free in spite of restrictions, usage rights, or copyrights (overriding RA, PA, Restricted OA). As publishers' copyright policies and self-archiving rules are compiled by the University of Nottingham in the SHERPA/RoMEO database, Rogue OA is synonymous with Robin Hood OA.

DOA: Delayed OA—access after a delay period or embargo.

- **Delayed Green OA**—free online access provided by the author after a delay (due to author's own delay in making available for free) or embargo period (typically imposed by publisher).
- **Delayed Gold OA**—free online access provided by the publisher after a delay (e.g. change of policy that makes contents available for free) or embargo period.
  - **Delayed Gold OA Journal**—journal offering cover-to-cover access after an embargo period or after a delay.
  - **Delayed Gold OA Article**—paper appearing in a Gold journal or in a PA journal (the latter is also sometimes referred to as hybrid open access) which is available after an embargo period or after a delay.

TOA: Transient OA—free online access during a certain time.

Transient Green OA—free online access provided by the author for a certain time which then disappears. Note that a substantial part of Green OA could be Transient Green OA due to the unstable nature of the internet, websites, and institutional repositories, many of which are not updated or maintained after a period of time and are therefore susceptible to deletion in subsequent institutional website overhauls. There are also integrator repositories that can change access rules, for example after being acquired by a third party.

Transient Gold OA—free but temporary online access provided by the publisher, instead of permanent. Sometimes appears as part of promotion. Note that some Gold journals and articles sometimes become paid access after a certain time, because of revised strategies by a publisher or because they are sold to another publisher which institutes paid access.

The ineffectual situation whereby publicly-funded research results published in peer reviewed journals continue to sit behind a 'pay wall', a situation made worse by continuous increases in the price of scholarly journal subscriptions, has fuelled the popularity of open access (OA) in recent years. Borne on the back of the digital revolution, the movement towards OA to scholarly knowledge is transforming the global research communication and dissemination system. Since the pioneering years of OA in the early nineties, OA literature has come to occupy an increasing share of scholarly research across geographical regions and scientific disciplines.

In response to what many perceive to be a dysfunctional system, individual researchers, libraries, universities, research funders, and governments have become incentivised to join the campaign for OA.

Though only a few years ago research suggested that OA growth was modest, the present set of studies conducted by Science-Metrix for the European Commission provides undeniable evidence that the OA movement is disruptive and that OA is traversing the scientific, technical and medical publishing industry with the speed and force of a tsunami.

This report presents a summary of this series of studies. It examines the current state of the art of OA strategies to peer-review publications (Part I), followed by a state-of-the-art analysis of OA strategies to scientific data (Part II). A third part of the study performs an assessment of the proportion and the number of OA papers published in peer-reviewed scientific journals. The study focuses on the 28 member states of the European Union (EU28), as well as the European Research Area (ERA), Brazil, Canada, Japan, and the US.

I—Analysis of OA strategies to peer-reviewed publications

This section examines the policies and strategies of the main institutional actors in the OA movement—governments, funding bodies, and research institutions—and describes the response of scientific journal publishers.

Governmental OA strategies

Most national governments have not proposed or implemented direct legislation on OA. Instead, OA is often addressed through less formal means, such as the production of guidelines for research funding agencies. Related legislation often includes laws on copyright and licensing; in fact, the vast majority of countries covered in this study have copyright legislation that may apply to peer-reviewed publications. Legislation directly addressing OA has been implemented in the US, Spain, and Germany. Italy and Lithuania have also recently passed laws that have direct implications for OA in those countries.

- In the US, the Consolidated Appropriations Act, 2008, is the legislative basis for the OA policy of the National Institutes of Health (NIH). With this legislation, the US became the first country to adopt a national OA mandate. The Fair Access to Science and Technology Research Act (FASTR) was introduced in Congress in February 2013. At the same time, the White House's Office of Science and Technology Policy (OSTP) issued a Directive on Public Access. Both require all federal agencies with extramural research expenditures of over $100 million to develop a federal research public access policy. On January 17, 2014, an omnibus federal spending bill, similar in concept to the OSTP Directive, was signed into law by President Obama; it includes an OA mandate (Section 527 of the Consolidated Appropriations Act, 2014 [H.R.3547]).
- Spain passed the national Law on Science, Technology and Innovation (2011) (Ley de la Ciencia). Article 37 mandates that a digital version of the final copy of research accepted for publication and funded under the national public R&D funding scheme be deposited in an OA repository within 12 months of publication.
- In Germany, legislation (BGB1. I S. 3714) enacted on October 1, 2013, allows authors to make any of their articles stemming from publicly-funded research openly accessible and available for non-commercial purposes without the consent of the publishers 12 months after first publication.
- Italy's Law of 7 October 2013 (No. 112) includes a regulation for OA to scientific publications. Researchers whose research has been publicly funded (at least 50%) are required to either publish their work in an OA journal or deposit their work in OA archives within 18 to 24 months after publication.
- In Lithuania, national legislation mandating that all state-funded research be made available publically took effect on May 12, 2009.

A number of countries in the ERA have instituted national policies, programmes, and principles related to OA.

- The UK is a leader in the development of OA to peer-reviewed publications, with the Higher Education Funding Council for England (HEFCE) and the Research Councils UK (RCUK) pushing for greater public access to publicly-supported research, and the 2012 'Finch Report' raising awareness of and fostering discussion on OA in the country. Jisc, a UK registered charity that champions the use of digital technologies in UK education and research, has also had a considerable influence. The 2013 Policy on Open Access, drafted by the Working Group on Expanding Access to Published Research Findings, chaired by Dame Janet Finch, will make all government-funded research OA within five years, with a target of 45% in the first year.
- In Ireland, the National Principles for Open Access Policy Statement (2012) mandates the deposit of outputs of funded research in OA repositories.
- Since 2006, Sweden has had a national OA programme, OpenAccess.se, which has played a role in the creation of a national search portal for scholarly publications (SwePub), the Directory of Open Access Journals (DOAJ), and a number of institutional and funder policies.
- France's HAL multi-disciplinary open archive was launched by the Centre National de la Recherche Scientifique (CNRS) in 2001.

At the pan-European level, the Open Access Pilot was launched by the European Commission as part of its Seventh Framework Programme (FP7) in August 2008. Within several thematic areas of the framework programme, FP7 projects are required to deposit peer-reviewed research articles or final manuscripts resulting from projects into an online repository. Other Europe-wide initiatives include the Digital Repository Infrastructure Vision for European Research (DRIVER), established to build a cohesive network of repositories for research and education, and the Open Access Infrastructure for Research in Europe (OpenAIRE), a complementary project offering organisational and technological infrastructure for the identification, deposition, access and monitoring of FP7 and European Research Council (ERC) funded publications. Launched in February 2014, a new 30-month project titled PASTEUR4OA (Open Access Policy Alignment Strategies for European Union Research) aims to help Member States to develop and/or reinforce OA strategies and policies at the national level and facilitate coordination among Member States.

Most national governments have not proposed or implemented direct legislation on OA. Instead, OA is often addressed through less formal means, such as the production of guidelines for research funding agencies.

The UK Policy on Open Access aims to make all government-funded research OA within five years, with a target of 45% in the first year.

According to the results of the present study, the UK already had 64% of papers published in 2008-2013 available in OA as of April 2014.

OpenAIRE facilitates the identification, deposition, access and monitoring of FP7 and European Research Council funded publications.

At the international level, research-performing organisations also contribute to the global spread of OA. Since 2010, much of the World Bank's research has been made available to the public through its website and the Open Data Initiative and Access to Information Policy. Furthermore, its new OA policy, effective since July 1, 2012, adopts the Creative Commons Attribution (CC-BY) copyright licence for all of its outputs and other knowledge products, solidifying the World Bank's position as a forerunner in OA.

Funding bodies' policies and mandates

Repository development and implementation presents numerous challenges related to intellectual property rights, data curation, long-term preservation, infrastructure development and interoperability.

An analysis of funding bodies' OA policies was performed to assess the extensiveness of OA policies and to examine OA rules for grant recipients, across ERA countries and in Brazil, Canada, Japan, and the US. The country with the highest number of OA policies is the UK (with 34 mandates), followed by Canada (14), the US (9), Denmark (6), Ireland (5), and France (5). No funder mandates could be found for Bulgaria, Croatia, Cyprus, the Czech Republic, Estonia, Finland, Greece, Latvia, Lithuania, Luxembourg, Malta, Poland, Romania, Slovakia, Slovenia, Liechtenstein, Turkey, or the Former Yugoslav Republic of Macedonia. Importantly though, the number of policies alone is a weak indicator of commitment to OA in a given country.

An analysis of 48 implemented funder policies in the countries was conducted, and startling differences were found in the level of detail provided. Funding agencies that are building a new OA policy or renewing a pre-existing policy should consider a number of key points for transparency. These include the following:

- Coverage of article processing charges (APCs): 46% of funders indicated that they would cover the APCs, and 12% indicated that they would not. The remaining 42% of policies did not specify whether funders would pay the fees.
- Preference for Green or Gold OA: Except for the UK which favoured Gold OA, most favoured Green (40%), while 2% mandated both Green OA and Gold OA, and 58% expressed no preference. If Green OA is allowed, specific repositories may be identified.
- Acceptable types of documents and metadata: 80% of policies specified that articles should be deposited in their final accepted version or post-print.
- Funding share scope: Guidelines must clarify at what point the policy applies with respect to a certain percentage of funding provided or the number of authors who are grantees.
- Embargoes: If the policy supports Green OA, embargoes are frequently accepted to give publishers exclusivity for a limited time. Though this diverges from the ideal form of OA, which stipulates that results be available immediately, among the 48 funder policies reviewed, 77% accepted embargoes between 6 and 12 months.
- Though compliance with policy is currently seldom tracked or reported, clear rules should be mentioned.
- Other items that may be listed in the policy include the prescribed timing of deposits, exceptions in types of outputs, the transfer of rights, and sanctions.

Research institutions' OA strategies

The ERA countries, Brazil, Canada, Japan, and the US collectively have 293 institutional, multi-institutional, sub-institutional, and thesis OA mandates presently in place in addition to 70 proposed institutional, multi-institutional, and sub-institutional OA mandates and non-mandates. The largest number of mandates is in the US, followed by the UK, Italy, Finland, and Portugal.

A survey of head librarians at universities and higher-learning institutions was conducted for this study. The survey found that 73% of respondents agreed or strongly agreed with the statement 'Providing open access to scholarly publications is a priority in [their] organisation'. A similar proportion agreed or strongly agreed with the statement 'Providing open access to scholarly publications is a priority in [their] country'. However, only 42% of these respondents stated that their organisation has an open access policy regarding peer-reviewed scholarly publications. Among these respondents, 22% declared that their organisation's policy is not publicly available.

Within the ERA, Brazil, Canada, Japan, and the US, nearly 40 million records were spread across 1,450 institutional repositories in April 2014, as reported in OpenDOAR alone. Institutional repositories account for 28% of records in OA repositories (disciplinary repositories account for 38%, governmental repositories for 25%, and aggregating repositories for 9%) (as of April 2013). Importantly, these records are highly heterogeneous. OA institutional repositories contain digital images, music, and text. Only a portion of the text files is made up of peer-reviewed scholarly papers. Hence, the presence of records in repositories is not a robust proxy for the availability of scientific papers and won't be so until repositories are characterised with all due care.

Repository development and implementation present numerous challenges related to intellectual property rights, data curation, long-

term preservation, infrastructure development and interoperability. The greatest strength of many repositories is the large number of records they contain, not necessarily the quality and value of their data. Records are composed solely of incomplete metadata and, frequently, contain no link to articles or other digital objects.

Universities also struggle with promoting OA within the academic community. Incentives are essential for reaching researchers who are reticent about OA or who are deterred by the trade-off between the costs and benefits of making their work OA. Survey results suggest that direct advantages for researchers who make their work available in OA form remain rare at the level of institutions. Of survey respondents, 49% indicated that researchers in their organisation were encouraged to archive their scholarly work but without any formal reward, and 36% indicated that their institution had no policy in this regard. Meanwhile, only 15% of respondents indicated that self-archiving was mandatory, and 14% stated that financial support was available for researchers who published in OA journals.

Effects on and response of scientific journal publishers

In response to the changing landscape of scholarly communication, publishers have developed new products known as 'big deals'. Contracts are established between libraries and publishers whereby libraries secure access to a large set of journals distributed by the publisher, mostly in electronic format, for a set price, for a period of three to five years, and for all faculty and students at the subscribing university. Though this has had the effect of theoretically lowering the subscription cost per journal, the negotiating power of research libraries has fallen, and they do not necessarily end up with an optimal mix of journals. Another effect of this strategy is to augment the market presence of large publishers, who can impose journals of lesser interest to researchers due to their being packaged with their 'must subscribe to' journals.

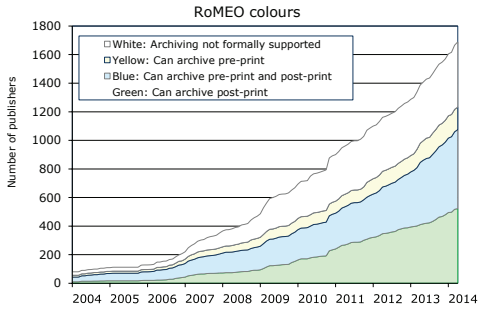
Mergers and acquisitions in the scholarly communication sector have increased the concentration of journals in the hands of a limited number of publishers. As a result, librarians have little power to opt out of big deals or negotiate the terms of subscriptions contracts. As the price of serials, particularly the highly sought after ones, has continued to rise faster than inflation, library budgets have increased moderately, stagnated or even decreased, a situation referred to as the 'serials crisis'.

There is tension between universities and publishers, which are suspected of 'double dipping', that is, charging article processing charges while subsequently obtaining revenue from journal subscriptions. In response to rising complaints, many publishers provide publication charge discounts to authors from institutions that subscribe to a relevant "Hybrid" journal (a paid access journal comprising a number of Gold OA articles). Other journals offer discounts to society members or institutions that have purchased institutional memberships.

Although it is commonly assumed that there are many vested interests in preserving the status quo of the current subscription market, many publishers have recognised that OA can lead to wider dissemination, maximised market reach, greater visibility and higher journal citation impact factors for their articles. Yet, OA journals have been confronted with the challenge of adopting a funding model that is consistent with their survival. Several models for OA publishing that differ with respect to type of content access, retention of author's rights and type of financing have emerged. These models include OA journals that are free for authors and readers; OA journals that are free for authors and readers of the online version, with subscription payment for the paper version; 'author pays' OA journals; hybrid systems; journals with free access to certain content; and journals with free access to contents after an embargo period.

In recent years, many traditional commercial publishers—including the Nature Publishing Group, Springer, and Elsevier—have established sizable OA journal operations or have extended their Hybrid OA operations. Successful 'Mega-OA' journals have also been established, most notably PLoS ONE (Public Library of Science) and Scientific Reports (Nature Publishing Group), and these have had a positive impact on the credibility of OA peer-reviewed publications.

Journal publishers are also increasingly allowing article archiving. As compiled in SHERPA/RoMEO, from January 2004 to April 2014, the number of publishers' OA policies allowing some form of archiving grew steadily, from 80 to close to 1,700. Of these, 31% allow post-print archiving, 33% allow pre-print and post-print archiving, and 9% allow pre-print archiving only. The remaining 27% of publishers do not formally allow any form of archiving but may agree to special arrangements with authors, particularly in the context of a funder mandate.



Mergers and acquisitions in the scholarly communication sector have increased the concentration of journals in the hands of a limited number of publishers. As a result, librarians have little power to opt out of big deals or negotiate the terms of subscriptions contracts.

II—Analysis of OA strategies to scientific data

This part of the study examines strategies that aim to foster open access to scientific data—such as incentives given at the researcher and institutional levels—and also examines how, and whether, these policies are monitored and enforced. The infrastructure developed to store and share OA scientific data is also examined. The analysis is supported by findings from the literature on the global progression of OA scientific data since 2000—including its growth as a segment of scholarly publishing—as well as some broader themes and debates that have emerged from the movement.

Governmental OA scientific data strategies

Most national open data policies target datasets for government data that are not necessarily generated through scientific research but may be used for research, rather than scientific data at large.

The political and administrative agendas in a rising number of countries now include the definition of open data strategies for public research through a range of projects, declarations, and policies, including legislation. To date, most national open data policies target datasets for government data that are not necessarily generated through scientific research but may be used for research, rather than scientific data at large. The countries with some of the largest numbers of open data strategies and open datasets are the United Kingdom (UK), the US, Denmark, Norway, the Netherlands, and Sweden.

The importance of comprehensive OA policies was recognised in 2004 by the Ministers of Science and Technology of the then 30 OECD countries, and of China, Israel, Russia, and South Africa. Governments may reap important economic benefits from the release of OA scientific data, such as through innovation-driven economic growth and job creation and more informed policy and research.

The US and several other national governments and major international organisations have established data portals starting in 2009. Though these data are not necessarily generated through scientific research, they may frequently have a use in research.

It has been estimated that, by unlocking the potential of big data, developed European economies could save between €150 billion and €300 billion annually in the form of operational efficiency gains and increased potential versus actual collection of tax revenue alone. A study has estimated the value of direct and indirect economic impacts of government-owned data across the EU-27 at €140 billion annually. It also estimated that, with lower barriers and improved infrastructure, this value could have been around €200 billion in 2008, representing 1.7% of the European GDP for that year.

In order to unlock the full economic potential of OA scientific data, machines must be able to access and mine this information without copyright infringement. The European Commission's Working Group on Text and Data Mining (TDM) Licences for Europe initiative has addressed this challenge through stakeholder dialogue. However, representatives from groups of researchers, science organisations, libraries, and small and medium enterprises have withdrawn from the process, arguing that the Working Group's decision to place licensing at the centre of discussions constitutes a bias against other solutions to adapt the legal framework to TDM applications.

The European Union issued a directive on the reuse of public sector information as early as 2003, while the US government was the first to establish its own open data portal, data.gov, in 2009. National public data repositories were launched in 20 countries in the following years.

Major international organisations, including the United Nations, OECD, European Commission, International Energy Agency (IEA), Nuclear Energy Agency (NEA), Programme for International Student Assessment (PISA), and International Transport Forum (ITF), have also established open data portals. The World Bank's repository, data.worldbank.org, launched in 2011, is a landmark repository in terms of its usability by machines and humans alike.

States, provinces, and municipalities own the bulk of currently available open governmental data repositories. However, a few national repositories now offer access to vast amounts of data. Notable governments and funding bodies that have initiated OA scientific databases include the NIH's GenBank, the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database, and the ClinicalTrials.gov database.

Funding bodies' strategies

Funding bodies may have an interest in archiving and disseminating data generated by their grantees to monitor the outcomes of their investments, increase the visibility of their contribution to research, and ensure long-term preservation of the data. Mining and further analysing the data collected from multiple projects could also highlight underrepresented or emerging areas of interest.

Funding bodies have far fewer OA policies for scientific data than policies for scientific publications. In an analysis conducted by Science-Metrix of 48 funders' OA policies listed in ROARMAP within ERA countries, Brazil, Canada, and the US, 23% explicitly excluded data from OA requirements, and 38% did not mention data at all. Conversely, 29% of policies mandated open data archiving and 10% encouraged it without mandating it.

Although they are still in the minority, funding agencies that expect applicants to create and

adhere to data storage, access, and management plans are increasing in number, with many agencies attaching conditions that strongly encourage or require the sharing of data with as few restrictions as possible and within a specified time frame.

Some rare funding bodies operate their own repositories as most recommend the use of institutional, disciplinary, or aggregating repositories. Institutions, government bodies, and research networks may be able to better define their needs in terms of standards, space, and functionality based on a more direct relationship with the researchers who produce the data or with the data users.

Research institutions' strategies

OA policies for scientific data are less common than OA policies for scholarly publications. In a survey of head librarians at universities and higher learning institutions conducted by Science-Metrix, 73% of respondents agreed that providing scholarly publications in open access form is a priority in their organisation, whereas only 45% agreed that providing scientific data in open access form is a priority. A similar trend is observed in the adoption of OA policies for scholarly publications and scientific data. Only 11% of respondents indicated that their institution has an OA policy for scientific data, while for scholarly publications, 42% indicated that their institution has an OA policy. Thus, establishing institutional frameworks for the diffusion of data in open access remains a secondary concern at this point.

OA scientific data has the potential to strengthen the credibility of scholarly publications and research institutions, as it opens peer-review to the entire scientific community. A growing trend has been reported in the percentage of published articles retracted for fraud or suspected fraud. Easily accessible data could extend the peer-review process beyond a small group of reviewers acting before publication, to all readers after publication.

Few institutional repositories are dedicated exclusively to research data. Institutional repositories generally support datasets on repositories devoted to books, theses, peer-reviewed publications, and multimedia objects. Moreover, the infrastructure required to host and share scientific data, while still uncommon, is more developed than the associated policy frameworks. In the survey, 36% of respondents indicated that their organisation maintains one or more repositories for OA scientific data, whereas only 11% indicated that their organisation had a policy to this effect.

The survey's results also suggest that a significant number of existing OA repositories for scientific data are not indexed in ROAR and OpenDOAR, the directories used in this study. This impression is reinforced by the absence of known open data repositories, such as the NASA's AERONET and the European Molecular Biology Laboratory's databases. The repositories listed could merely be the tip of the iceberg.

III—Analysis of the proportion and number of papers available in OA

This core part of the study assessed the free availability of scholarly publications during the 1996 to 2013 period. It is the largest scale measurement of open access availability performed to date: a sample of one-quarter of a million records was used to study the historical evolution of OA between 1996 and 2013 and a larger, one-million-record sample, was used to perform an in-depth assessment of the proportion and scientific impact of OA between 2008 and 2013 in different types of OA, for different scientific fields of knowledge, and for 44 countries comprising the EU28 and ERA countries and the world.

Compared to previous studies done on the availability of OA, the present study presents the following characteristics: (1) it used the Scopus database, which currently covers a broader range of journals from various countries and scientific disciplines than other comprehensive databases; (2) it uses a simple definition of OA—freely available online to all (no money had to be paid, no registration to a service or website had to be made); (3) it used huge samples to maximise statistical precision; (4) it made careful and extensive efforts to harvest papers wherever they could be downloaded for free without restriction, rather than using a single existing search engine in order to obtain a high 'recall' rate (that is, the capacity to retrieve a large part of the relevant records), while, in addition, carefully minimising the number of false records collected (that is, the approach maximised retrieval precision); and (5) it carefully characterised the strengths and weaknesses of the measurement instrument in order to apply a correction that would provide a truer measure based on an Adjusted OA score.

This study also provided a series of rational definitions of access, open access, and ideal open access (see above). The definitions include aspects such as restrictions, payment, delay, transiency, and legitimacy. Because of the limited means (time and budget) available for this project, it was necessary to use operational definitions of OA that do not provide all the details one may wish to obtain. Though it was easy to obtain a clear and operational definition of Gold OA by stating that it referred to papers published in Gold OA journals listed in the Directory of Open Access Journals, defining and measuring Green OA was more

Establishing frameworks for the diffusion of data in open access remains a secondary concern for research institutions at this point.

Given the growing percentage of published articles retracted for fraud or suspected fraud, extending the data peer-review process beyond a small group of reviewers acting before publication, to all readers after publication is a high value proposition.

Though it was easy to obtain a clear and operational definition of Gold OA by stating that it referred to papers published in Gold OA journals listed in the Directory of Open Access Journals, defining and measuring Green OA was more challenging. The operational definition restricted Green OA to researchers' self-archived papers in institutional and some thematic repositories listed in OpenDOAR and ROAR. This left a sizeable residual number of papers that could still be downloaded for free; these were classified as Other OA.

The percentage of peer-reviewed articles published in Gold OA journals indexed in Scopus for 1996 was only 0.9%, but grew to 12.8% for 2012. This translates into an annual growth rate of 18% for this period, which means that the proportion of articles in Gold OA journals doubles every 4.1 years.

challenging. The operational definition restricted Green OA to researchers' self-archived papers in institutional and some thematic repositories listed in OpenDOAR and ROAR. This left a sizeable residual number of papers that could still be downloaded for free; these were classified as Other OA. This comprises, for example, Gold OA papers from subscription-based journals, which are made available through article processing charges (APC). Other OA also include papers available in large repositories such as PubMed Central and on aggregator sites such as CiteSeerX. There are also Robin Hood OA or Rogue OA papers, that is, papers that infringe on copyrights by making them accessible to the public despite licenses that restrict them to being behind pay walls.

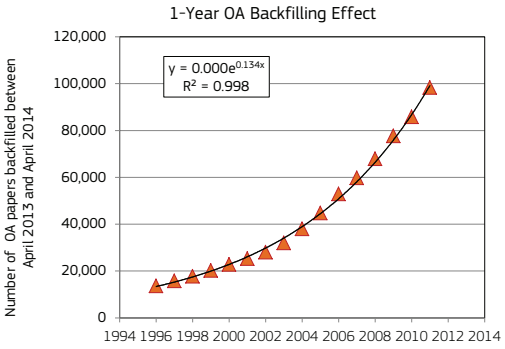
Measurement instrument calibration

A sample of 500 articles from a pilot study (December 2012) and a previous version of this study (April 2013) was used to characterise the harvester employed to measure the proportion of OA papers. Slight variations were observed in the availability of articles in this sample measured in December 2012 (47.6%), April 2013 (44.8%) and April 2014 (48.6%). It is noteworthy that 272 articles, that is, about 54.4%, were available for free at one time or another between December 2012 and April 2013. These results suggest that important transient aspects need to be taken into consideration while measuring OA availability. These results also show that the harvesting engine developed by Science-Metrix has very good retrieval precision (99.1%) and fairly good recall (86.4%), resulting in fairly robust measures of OA availability. This characterisation of the measurement instrument made it possible to calibrate to produce 'Adjusted OA' measures. The total OA measures by the harvester are multiplied by 1.146. However, because the sample size of the calibration was only 500 records, the margin of error of the calibration is a fairly large ± 4.5 percentage points.

Evolution of the proportion and the number of OA papers

As of April 2014, more than 50% of the scientific papers published in 2007, 2008, 2009, 2010, 2011, and 2012 can be downloaded for free on the Internet. This is an important finding as only one year ago, in April 2013, the proportion of papers that was freely available was just a hair below 50% (49.54%) in 2011 and did not reach that mark for any other year. On average, the citation advantage of OA papers is 40%, while the citation disadvantage is 27% for non-OA papers (based on a total sample size of 209,000 papers).

The growth of OA appears as the result of four



main forces: (1) historical growth in the interest in OA, which translates into *new papers* being increasingly available for free; (2) the growing interest in OA translates into actors increasingly making *old papers* available for free; (3) OA policies that allow for delaying OA produce a concomitant disembargoing of scientific articles, which creates additional growth in old papers being made available for free; and (4) the number of published scientific papers is growing, so even for a stable proportion of OA, the number of OA papers would keep growing.

The effect of backfilling has not been studied extensively before. Evidence suggests that during the last year alone, some 700,000 papers indexed in Scopus between 1996 and 2011 became available for free, that is, an extra 3.9 percentage points. Studying the OA availability curve for the 2004–2011 period of one year ago compared to this year (April 2013 vs. April 2014) reveals that the present curve has made an upward translation of 3.6 percentage points (measured for 2004); in addition to going up, the curve is also becoming steeper—the exponent of the curve increased from 1.9% overall growth to 2.4% growth this year. This means backfilling is accelerating over time.

The statistics just presented were on the proportion of OA; it is also relevant to assess how the number of papers is growing per se. These data show that, as of April 2014, the number of available papers increased by 9.4% per year. Of the papers published in 1996, 240,000 are now available for free, as are 950,000 of the papers published in 2012. Based on the Adjusted OA availability statistics, one can estimate that about 47% of the papers indexed in Scopus between 1996 and 2013 can be downloaded for free as of April 2014. This means that 10.1 million papers of the 21.5 million papers indexed in Scopus for that period and which can be considered to be peer-reviewed papers published in scientific journals are downloadable.

While the number of Green OA papers has grown steadily, this appears to be due to background growth, that is, the growth in published scientific papers. Green OA as a percentage of the papers indexed in Scopus appears to have levelled off from around 2004. This requires further investigation to determine whether this is measurement artefact—an effect of the imperfect operational definition

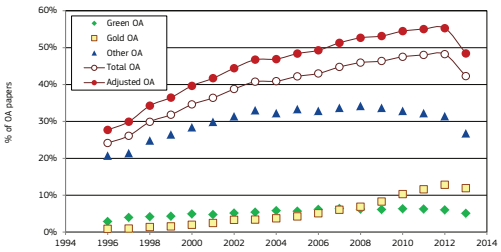
of Green OA used here—or whether Green OA is somewhat losing steam. Despite this, there were approximately 1.2 million papers available in Green OA form in repositories across the world, and the growth rate of OA papers was 8.8% between 1997 and 2011.

The percentage of peer-reviewed articles published in Gold OA journals indexed in Scopus for 1996 was only 0.9%, but grew to 12.8% for 2012, the annual growth rate being 18% for this period, which means that the proportion of articles in Gold OA journals doubles every 4.1 years. Scopus covers less than half of the quantity of journals listed in the DOAJ, so this figure likely underestimates the true extent of the role played by Gold OA journals.

Furthermore, Gold OA papers indexed in Scopus grew exponentially up until 2012. The growth rate was 24% per year between 1996 and 2012, which means that the number of papers published in Gold OA journals doubles every 3.2 years. There are currently about 1,380,000 papers from Gold journals indexed in Scopus for 1996 to 2013. This represents only 200,000 papers more than those available in Green OA form, but Gold OA journals papers are growing with a clear momentum, which is not so clearly the case with Green OA available in institutional repositories.

The evolution of Other OA is somewhat similar to that of Green OA. In terms of percentage, it increased substantially from 1996 until about 2003, at which time it levelled off until 2008, when there was an observed decline in the proportion of Other OA papers. It is not so easy to determine what creates this shape due to the heterogeneity of the underlying dataset. What seems to be obvious here is that unless we are witnessing a slowdown in OA development, embargoing and others forms of DOA (Delayed OA) are having a very tangible effect on scientific knowledge availability. This creates a situation whereby a substantial part of the material openly available is relatively old, and as some would say, outdated. As in the case of Green OA, the growth as measured in number of papers is greater than the growth of the proportion. There was a regular increase of 8.4% per year in the number of papers from 1996 to 2009, after which the increase slowed down and dropped in 2013 as the full effect of embargoes surpassed the growth in available papers (which was 6.6% per year in Scopus between 2003 and 2012).

Green OA papers, those deposited in institutional repositories, do not contribute a large share of the overall OA stock of papers. As seen previously, their number did not increase much after 2004. A word of warning here though: authors can also backfill repositories, and the current measurement does not take this into account as no baseline for Green OA was measured last year.



Other forms of OA—Gold OA Papers (that is, those with article processing charges published in subscription journal or so-called hybrid journals), Green DOA and Gold DOA (embargoed self-archiving and embargoed journals), ROA (Robin Hood or Rogue OA), and papers archived in non-institutional repositories such as ResearchGate—account for a large part of the pie. This large heterogeneous set contributes the largest proportion of OA papers, and there is therefore an urgent need to disaggregate this category. More research and more careful classification and thus finer-grained measures are required to better understand how these various categories contribute to OA growth, what their pattern of time-delay is, what their transiency is (especially of the ROA), how important backfilling is and how far back it goes.

OA papers were between 26% and 64% more cited on average for any given year than all papers combined, whereas non-OA received between 17% and 33% fewer citations (based on a sample size of at least 10,000 papers any given year). Green OA and Other OA papers have somewhat similar patterns. Papers in subscription-based journals (non-OA articles only) and Gold OA journals are also somewhat similar to one another. On average, Green OA (operational definition used here) papers have the greatest citation advantage, being cited 53% more frequently than all papers. They are followed by the Other OA category, which is 47% more frequently cited on average. Papers published in Gold OA journals have a citation disadvantage of 35% on average, compared to a disadvantage of 27% for non-OA papers.

Evolution of OA by scientific field

Considering the last three years together (2011–2013), as of April 2014 more than 50% of the papers can be freely downloaded in 12 fields out of 22. A growth index was computed by dividing the percentage of OA availability in 2011 and 2012 by that observed in 2008 and 2009 (2013 was left aside as embargoes would distort calculated growth rates). Overall, between the two periods, there was a 4% increase in OA availability (slightly less than 2 percentage points). The fields with the fastest growth during these periods were general science & technology, enabling & strategic technologies, public health and health services, visual & performing arts, clinical medicine, and built & environment design. One can suspect that

the NIH OA mandate is at play here (in public health and clinical medicine).

The fields with the greatest proportion of OA are general science & technology (Adjusted OA=90%), biomedical research (71%), mathematics & statistics (68%), and biology (66%). OA is not as commonly used in visual & performing arts (Adjusted OA=25%), communication & textual studies (31%), historical studies (34%), engineering (35%), and philosophy & theology (35%).

Green OA is particularly present in physics & astronomy (25.6%), which is certainly helped by the presence of arXiv, which probably also plays a role in mathematics & statistics (24.3%), while economics & business is the leading field in the social sciences and humanities (11.3% of papers in Green OA).

Gold OA availability is greatest in general S&T (58% of the sampled papers) and lowest in general arts, humanities & social sciences (2.6%); it is also very low in the visual & performing arts (2.8%), built environment & design (3.5%) and engineering (4.1%). Other fields with high availability in Gold journals include biology (17%), agriculture, fisheries & forestry (16%), and public health & health services (16%).

Other forms of OA are frequently encountered in biomedical research (48%), psychology and cognitive sciences (43%), biology (42%), earth & environmental sciences (38%), and clinical medicine (35%).

The absolute number of papers in OA form is rising rapidly (as there is also underlying growth in the number of papers generally). For example, the growth of the OA proportion in agriculture, fisheries & forestry was 1.02, but the number of papers grew at 1.16 (16% growth in the number of OA papers indexed in Scopus in 2011–2013 compared to the 2008–2009 period).

Overall, out of the 4.6 million scientific papers from peer-reviewed journals indexed in Scopus during the 2011–2013 period, 2.5 million were available for free in April 2014 (Adjusted OA score). A very large number of papers are freely available in clinical medicine (Adjusted OA = 680,000 papers), biomedical research, and physics and astronomy (close to 250,000 papers, as calculated with an adjusted measure). This is partly because of the policy of the NIH that mandates the use of the PubMed Central repository for supported research and because of the arXiv e-print archive, which has been largely adopted by researchers in the field of physics.

All the fields derive an OA citation advantage. Paradoxically, many of the fields where the OA proportion is low have a sizeable citation advantage, such as the visual & performing arts (80% more cited), communication & textual studies (66%), philosophy & textual studies (63%), historical studies (55%), general arts, humanities and social sciences (51%), and engineering (38%). An explanation to this is likely to be that papers from researchers in these fields are more likely to have their papers used as there are fewer OA papers available.

There is a huge citation advantage to publishing in Green OA, as has been demonstrated time and again in other serious studies conducted previously. Papers in general science & technology, in historical studies and in visual & performing arts all receive, on average, twice as many citations as

Scientific impact (ARC) of publications in Open Access by field (2009-2011)

Field	1st place		2nd place		3rd place		Least impact	
	Type	ARC	Type	ARC	Type	ARC	Type	ARC
Agriculture, Fisheries & Forestry	Green OA	1.57	Other OA	1.32	Not OA	0.88	Gold OA	0.51
Biology	Other OA	1.37	Green OA	1.30	Not OA	0.69	Gold OA	0.47
Biomedical Research	Other OA	1.23	Green OA	1.10	Gold OA	0.91	Not OA	0.65
Built Environment & Design	Green OA	1.56	Other OA	1.28	Not OA	0.86	Gold OA	0.19
Chemistry	Other OA	1.34	Green OA	1.28	Not OA	0.95	Gold OA	0.34
Clinical Medicine	Other OA	1.56	Green OA	1.08	Gold OA	0.64	Not OA	0.63
Communication & Textual Studies	Other OA	1.82	Green OA	1.51	Not OA	0.66	Gold OA	0.73
Earth & Environmental Sciences	Green OA	1.46	Other OA	1.26	Gold OA	0.98	Not OA	0.72
Economics & Business	Green OA	1.46	Other OA	1.30	Not OA	0.71	Gold OA	0.22
Enabling & Strategic Technologies	Green OA	1.68	Other OA	1.53	Not OA	0.83	Gold OA	0.52
Engineering	Green OA	1.84	Other OA	1.38	Not OA	0.83	Gold OA	0.55
General Arts, Humanities & Social Sciences	Green OA	1.74	Other OA	1.49	Not OA	0.73	Gold OA	0.13
General Science & Technology	Green OA	2.56	Other OA	2.24	Gold OA	0.69	Not OA	0.11
Historical Studies	Green OA	2.37	Other OA	1.61	Not OA	0.76	Gold OA	0.37
Information & Communication Technology	Green OA	1.62	Other OA	1.36	Gold OA	0.76	Not OA	0.69
Mathematics & Statistics	Green OA	1.35	Other OA	1.11	Not OA	0.75	Gold OA	0.67
Philosophy & Theology	Green OA	1.72	Other OA	1.63	Gold OA	0.86	Not OA	0.72
Physics & Astronomy	Green OA	1.43	Gold OA	1.18	Other OA	1.04	Not OA	0.73
Psychology & Cognitive Sciences	Other OA	1.35	Green OA	1.31	Not OA	0.66	Gold OA	0.59
Public Health & Health Services	Other OA	1.38	Green OA	1.30	Not OA	0.76	Gold OA	0.71
Social Sciences	Green OA	1.54	Other OA	1.44	Not OA	0.76	Gold OA	0.52
Visual & Performing Arts	Green OA	2.16	Other OA	1.86	Not OA	0.77	Gold OA	0.29
Total	Green OA	1.53	Other OA	1.36	Not OA	0.76	Gold OA	0.61

the overall population of papers. Two fields stand out for a fairly small Green OA citation advantage: clinical medicine (+8% vs. +56% in Other OA) and biomedical research (+10% vs. +23%). The reason may be that other sources of freely downloadable papers, classified here as 'Other OA', such as BioMed Central, are so large that the reflex of users is to first see what is available there and to shun institutional repositories. Still, 8% and 10% more citation remains a sizeable advantage, and it is worthwhile using institutional repositories and immediate Green OA to cut the delays associated with what many consider to be weak OA mandates, that is, those allowing for papers to be embargoed instead of being made available immediately.

On average, publishing in Gold OA journals is the least advantageous solution if one wants to maximise scientific impact. It is still more advantageous than strict Paid Access in seven fields, and ranks as the second best solution for physics & astronomy. Currently, there is a marked disadvantage for publishing in Gold journals in general arts, humanities & social sciences, built environment & design, economics & business and visual & performing arts. Interestingly, visual & performing arts has one of the highest advantages derived from the use of Green OA and OA generally, yet it is the field with the least prevalent use of OA.

The statistics on Gold journals require careful interpretation. First, many Gold journals are younger and smaller, and these factors have an adverse effect on the citation rate and hence on measured citation scores. Authors frequently prefer reading and citing established journals, and it is therefore a challenge to start a journal from scratch, and to have authors submit high-quality articles. It takes time to build a reputation and to attract established authors.

OA in the ERA and four selected countries

An examination of OA availability was performed for EU28 and ERA countries and four additional countries, Brazil, Canada, Japan, and the US. For the 2008–2013 period considered as a whole, the Adjusted OA suggests that all 44 countries have more than 50% of papers in OA for that period. Four EU28 countries have even reached an aggregate availability score above 70%—the Netherlands, Croatia, Estonia, and Portugal. It is interesting to note that the Netherlands, which is also scientific publishers' land of predilection, is the EU country with the largest share of papers available in OA form (74%) as a whole for papers published in the 2008–2013 period and available for free download as of April 2014.

Number of publications in Open Access per country (2008-2013)

Group	Country	Sample size	Green OA		Gold OA journals		Other OA		Total OA		Adjusted OA
			Found	%	Found	%	Found	%	Found	%	%
	Austria	8,764	821	9.4 ± 0.6	775	8.8 ± 0.6	3,450	39.4 ± 1.0	4,855	55.4 ± 1.0	63.5 ± 4.6
	Belgium	13,147	1,813	13.8 ± 0.6	968	7.4 ± 0.4	5,210	39.6 ± 0.8	7,841	59.6 ± 0.8	68.4 ± 4.6
	Bulgaria	1,707	161	9.4 ± 1.3	126	7.4 ± 1.2	558	33 ± 2	829	49 ± 2	56 ± 5
	Croatia	2,954	153	5.2 ± 0.8	687	23 ± 1.5	1,149	38.9 ± 1.7	1,876	63.5 ± 1.7	72.8 ± 4.8
	Cyprus	584	72	12 ± 3	43	7 ± 2	223	38 ± 4	329	56 ± 4	65 ± 6
	Czech Republic	7,637	521	6.8 ± 0.5	736	9.6 ± 0.6	2,598	34.0 ± 1.0	3,718	48.7 ± 1.1	55.8 ± 4.6
	Denmark	9,097	871	9.6 ± 0.6	819	9.0 ± 0.6	3,539	38.9 ± 1.0	5,127	56.4 ± 1.0	64.6 ± 4.6
	Estonia	932	81	8.7 ± 1.8	123	13 ± 2	390	42 ± 3	577	62 ± 3	71 ± 5
	Finland	7,414	659	8.9 ± 0.6	690	9.3 ± 0.6	2,838	38.3 ± 1.0	4,102	55.3 ± 1.1	63.4 ± 4.6
	France	48,991	6,881	14.0 ± 0.3	3,255	6.6 ± 0.2	16,560	33.8 ± 0.4	25,915	52.9 ± 0.4	60.6 ± 4.5
	Germany	66,268	7,575	11.4 ± 0.2	5,065	7.6 ± 0.2	21,993	33.2 ± 0.3	33,735	50.9 ± 0.4	58.4 ± 4.5
	Greece	8,043	525	6.5 ± 0.5	773	9.6 ± 0.6	3,067	38.1 ± 1.0	4,246	52.8 ± 1.0	60.5 ± 4.6
	Hungary	4,559	454	10.0 ± 0.8	356	7.8 ± 0.7	2,023	44 ± 1.4	2,782	61.0 ± 1.3	69.9 ± 4.7
	Ireland	5,150	815	15.8 ± 0.9	472	9.2 ± 0.8	1,839	36 ± 1.2	3,018	58.6 ± 1.3	67.2 ± 4.7
EU28	Italy	39,117	3,691	9.4 ± 0.3	3,112	8.0 ± 0.3	14,594	37.3 ± 0.5	21,021	53.7 ± 0.5	61.6 ± 4.5
&	Latvia	387	21	5.4 ± 2.3	57	15 ± 3	156	40 ± 5	232	60 ± 5	69 ± 7
ERA	Lithuania	1,434	65	4.5 ± 1.0	183	12.8 ± 1.7	593	41 ± 2	811	57 ± 2	65 ± 5
	Luxembourg	417	46	11.0 ± 3.0	36	9 ± 3	174	42 ± 5	253	61 ± 5	70 ± 6
	Malta	140	7	5.0 ± 3.8	30	21 ± 7	41	29 ± 7	75	54 ± 8	61 ± 9
	Netherlands	23,564	2,863	12.1 ± 0.4	1,883	8.0 ± 0.3	10,707	45.4 ± 0.6	15,177	64.4 ± 0.6	73.8 ± 4.5
	Poland	15,628	1,112	7.1 ± 0.4	2,099	13.4 ± 0.5	4,695	30.0 ± 0.7	7,416	47.5 ± 0.7	54.4 ± 4.5
	Portugal	7,190	1,169	16.3 ± 0.8	747	10.4 ± 0.7	2,636	36.7 ± 1.1	4,422	61.5 ± 1.1	70.5 ± 4.6
	Romania	5,105	271	5.3 ± 0.6	487	9.5 ± 0.8	1,994	39.1 ± 1.3	2,647	51.9 ± 1.3	59.4 ± 4.7
	Slovakia	2,372	156	6.6 ± 1.0	240	10.1 ± 1.2	798	33.6 ± 1.8	1,155	48.7 ± 1.9	55.8 ± 4.9
	Slovenia	2,586	181	7.0 ± 0.9	425	16.4 ± 1.4	871	33.7 ± 1.7	1,369	52.9 ± 1.8	60.7 ± 4.8
	Spain	35,557	3,517	9.9 ± 0.3	4,074	11.5 ± 0.3	12,119	34.1 ± 0.5	18,341	51.6 ± 0.5	59.1 ± 4.5
	Sweden	14,872	1,527	10.3 ± 0.5	1,460	9.8 ± 0.5	5,767	38.8 ± 0.7	8,587	57.7 ± 0.8	66.2 ± 4.5
	United Kingdom	73,621	8,506	11.6 ± 0.2	5,265	7.2 ± 0.2	28,173	38.3 ± 0.3	41,133	55.9 ± 0.3	64.0 ± 4.5
Total EU28		**337,231**	**31,635**	**9.4 ± 0.1**	**29,165**	**8.6 ± 0.1**	**117,793**	**34.9 ± 0.2**	**172,956**	**51.3 ± 0.2**	**58.8 ± 4.5**
	Albania	87	4	5 ± 5	16	18 ± 8	25	29 ± 10	43	49 ± 11	57 ± 11
	Bosnia and Herzegovina	362	3	0.8 ± 1.0	92	25 ± 4	139	38 ± 5	226	62 ± 5	72 ± 7
	Iceland	554	57	10.3 ± 2.5	37	7 ± 2	269	49 ± 4	354	64 ± 4	73 ± 6
	Israel	8,450	894	10.6 ± 0.6	502	5.9 ± 0.5	3,552	42.0 ± 1.0	4,882	57.8 ± 1.0	66.2 ± 4.6
ERA	Liechtenstein	40	5	13 ± 11			20	50 ± 16	25	63 ± 15	72 ± 16
Associated	FYR Macedonia	255	18	7 ± 3	75	29 ± 5	87	34 ± 6	173	68 ± 6	78 ± 7
Countries	Montenegro	104	6	6 ± 5	30	29 ± 9	38	37 ± 9	71	68 ± 9	78 ± 10
	Norway	7,280	629	8.6 ± 0.6	705	9.7 ± 0.6	2,907	39.9 ± 1.1	4,145	56.9 ± 1.1	65.3 ± 4.6
	Rep. of Moldova	160	14	9 ± 4	5	3 ± 3	52	33 ± 7	71	44 ± 8	51 ± 9
	Serbia	2,997	135	4.5 ± 0.7	906	30.2 ± 1.6	803	26.8 ± 1.5	1,786	59.6 ± 1.7	68.3 ± 4.8
	Switzerland	16,896	2,497	14.8 ± 0.5	1,547	9.2 ± 0.4	6,630	39.2 ± 0.7	10,369	61.4 ± 0.7	70.3 ± 4.5
	Turkey	17,420	475	2.7 ± 0.2	3,458	19.9 ± 0.6	4,853	27.9 ± 0.6	7,962	45.7 ± 0.7	52.4 ± 4.5
Total ERA		**375,820**	**33,766**	**9.0 0.1**	**34,932**	**9.3 0.1**	**130,244**	**34.7 0.1**	**192,202**	**51.1 0.2**	**58.6 4.5**
Others	Brazil	26,158	1,626	6.2 ± 0.3	10,482	40.1 ± 0.6	6,515	24.9 ± 0.5	17,322	66.2 ± 0.5	75.9 ± 4.5
	Canada	41,114	2,895	7.0 ± 0.2	3,098	7.5 ± 0.2	17,438	42.4 ± 0.5	23,096	56.2 ± 0.5	64.4 ± 4.5
	Japan	58,527	4,170	7.1 ± 0.2	5,382	9.2 ± 0.2	16,883	28.8 ± 0.3	25,846	44.2 ± 0.4	50.6 ± 4.5
	United States	258,815	17,865	6.9 ± 0.1	17,709	6.8 ± 0.1	119,943	46.3 ± 0.2	153,416	59.3 ± 0.2	67.9 ± 4.5
World		**1,000,000**	**60,271**	**6.0 ± 0.04**	**104,050**	**10.4 ± 0.1**	**324,637**	**32.5 ± 0.1**	**470,530**	**47.1 ± 0.1**	**53.9 ± 4.5**

All ERA countries have tipped towards having a majority of papers in OA, though in the case of the Republic of Moldova, the margin of error is quite high, and it is quite possible that the country has not tipped to OA yet. Swiss researchers contribute to making their country a leader in OA, with 70% of the papers being downloadable for free.

In countries outside the ERA, it is noteworthy that the US has passed the tipping point by a fair margin (Adjusted OA = 67.9%), as is also the case for Canada (64.4%). Even more salient is the proportion of 76% observed in Brazil. This is no doubt due to the important contribution of Scielo, which plays a key role in the Southern hemisphere in making scientific knowledge more widely available. Japan is just a hair over 50% and given the margin of error of Adjusted OA may or may not have tipped to having a majority of papers in OA form.

Within the European Union, Green OA is more widely used in Portugal (16.3%), Ireland (15.8%), France (14.0%), and Belgium (13.8%), and least used in Lithuania (4.5%), Malta (5.0%), Croatia (5.2%), and Romania (5.3%).

Publishing in gold journals is much more frequently encountered in Eastern Europe, as it is much higher in Croatia, Slovenia, Latvia, Poland, Estonia, and Lithuania (in addition to Malta). One interesting hypothesis is that researchers in these countries may use Gold journals because they more frequently allow publishing in languages other than English. Should that be the case, this may also contribute to explaining the lower citation scores received by papers in Gold journals as the readership for 'vernacular languages', as Eugene Garfield (1998) would put it, is lower and the size of the potential reference pool is consequently also smaller. This is therefore potentially fertile ground for studying the social and linguistic aspects of science by examining where and why Gold open access journals are appearing and who actually makes use of them. The countries that least use Gold OA journals are France (6.6%), the United Kingdom (7.2%), and Belgium (7.4%).

The comparatively slow growth of open scientific data may reflect the fact that individuals are more eager to decry and act upon the adverse effects of self-interest as long as it is not their own.

IV-Discussion and policy implications of OA to scientific data and peer-reviewed articles

Somewhat like open access to scientific articles, open data is evolving rapidly in an environment where citizens, institutions, governments, non-profits, and private corporations loosely cooperate to develop infrastructure, standards, prototypes, and business models. From a government perspective, open data confers a competitive advantage in an increasingly information-based economy. New products and services can be developed directly from data or through extensive transformation adding value to the information.

Many organisations use their data as a source of revenue by providing paid access to their datasets. In the case of government agencies within the EU, a review has estimated that the direct revenues generated from the sale of government-owned data represent less than 1% of each agency's revenues. However, for a few organisations, the sale of data may represent close to 20% of revenue, which is likely to hinder the development of open data policies even if the data was generated using public funds. Still, the economic benefits of open data are estimated to outweigh sales revenue by approximately two orders of magnitude within the EU28.

The heterogeneous nature of scientific data is a challenge for the development of OA in this area. A few fields of research use highly standardised formats that facilitate the aggregation and reuse of data, with genomics, proteomics, chemical crystallography, geography, astronomy, and archaeology among them. The archiving of other, less standardised types of data needs to be carefully considered in order to generate datasets that will be usable by other researchers. Overall, the growth of OA scientific data as a valid, citable form of reference is limited by the difficulties associated with the standardisation of data and metadata formats, poor indexation by internet crawlers, and the scarcity of directories or registries that could make data more visible.

Whereas researchers have a 'natural incentive' to promote OA in the case of scholarly publications, as this makes their work more widely known, in some cases there might be negative incentives to make research data public. Researchers may indeed derive a competitive advantage by filling up their own data warehouse, and widely diffusing these data may limit the growth and curtail the size of their research enterprises. The comparatively slow growth of open scientific data may reflect the fact that sometimes individuals are more eager to decry and act upon the adverse effects of self-interest as long as it is not their own.

It has been suggested that OA to scientific papers improves the speed, efficiency and efficacy of

research by allowing researchers faster access to the information they need. It increases the visibility, usage of scientific impact of research. The model may help to relieve the 'serials crisis' and save the direct costs of print publication and dissemination. For authors, it can shorten the delay between acceptance and publication in a journal. What is particularly interesting here is that the citation advantage is derived almost exclusively from the Green and Other OA portion, as Paid Access (not OA) systematically ranks behind Green OA and Other OA, and is an option yielding the lowest scientific impact (smaller number of citations received by papers on average) for seven fields. Publishing in Gold OA journals provides the least impactful choice in 15 out of 22 fields, and this low impact is probably more a reflection of the youth of these journals rather than anything else. Yet Gold OA journals beat Paid Access journals in 7 disciplines, showing that well-established Paid Access journals are losing their predominance and perhaps even their relevance in several fields of science. Despite their youth, Gold OA journals may continue to dislodge them in additional fields as their contents continue to improve. The day that articles that stay solely behind a pay wall will have lost their relevance may be edging closer faster than anyone expects.

The main bottlenecks that have prevented OA from gaining greater acceptance among stakeholders include a lack of awareness among researchers, concerns about the quality and prestige of OA journals, concerns and confusion about copyright, the dissuasive influence of article processing charges, difficulties moving beyond the current system of subscription-based journals, the lack of useful data on OA's evolution, a perceived lack of profitability surrounding OA business models, and a lack of infrastructure to support OA in developing countries. Despite these barriers, this investigation concludes that OA has become the dominant form of dissemination of peer-reviewed scholarly articles in the ERA, Brazil, Canada, Japan, and US.

Much has been said about the cost of publishing in Gold OA journals and for Gold OA articles ('hybrid publishing'). The cost of academic papers in the US is over $100,000—which is calculated by dividing the higher education expenditures on R&D (HERD) by the number of papers published by academia. In addition to or included in this amount, a $2,000 OA publication fee only accounts for a few percentage points of a typical research project budget, especially in the natural and health sciences. Green OA is free, and the majority of publishers accept that papers can be self-archived in one form or another (pre-print, post-print with final revision, or PDF) with no delay. Moreover, two-thirds of Gold OA journals do not levy author processing charges. There are free avenues to OA and cost should certainly not be seen as a barrier.

The current model of back end toll access is simply unsustainable because of its gross social inefficiency and ineffectiveness. Examining the OECD statistics on gross domestic expenditure on R&D in OECD and selected 'Non-OECD Member Economies' (Argentina, China, Romania, Russian Federation, Singapore, South Africa, Chinese Taipei), and using some conservative extrapolations, one can see that about $400 billion were spent by governments (GovERD) to support R&D. The revenue generated by science, technical and medical publishing (STM) for English-speaking countries was worth approximately $9.4 billion in the same year. Though these figures are not immediately comparable as part of the industry must derive income from non-English journals, the fact remains that a sizeable part of the research results paid for by $400 billion in public money is either delayed, restricted, or still simply behind thick pay walls to generate only $10 billion in private wealth. This is a case of gross inefficiency, one that taxpayers the world over can no longer tolerate.

Green OA advocates could respond that the lowest marginal cost to the system to make papers available for free is currently through the use of self-archiving in Green OA form: this would be a valid point. Yet, there is always a toll to be paid to create, diffuse, and use peer-reviewed papers. Few things are entirely free, be it a fee for a subscription, a fee to download an article, or a fee to publish an article; the public or philanthropic money that goes into supporting OA journals; article processing charges; or the opportunity cost associated with academics running journals instead of performing research and lecturing and training students, or providing other types of services to society.

Just as there is a need to continue to work vigilantly to remove the inefficiency created by all this public expenditure being made unavailable or available with undue restrictions, difficulties, and delays, there is also a need to closely monitor the effects of moving the scientific world from one based on Back End Paid Access (BEPA) to one based on Front End Paid Access (FEBA). BEPA created huge social inefficiency; FEBA has the potential to enlarge the rift between wealthier and more feebly financed countries, researchers, and scientific disciplines. Many mandates being promulgated at the moment run the risk of favouring a shift from BEPA to FEBA, from inaccessibility to inequality. Neither inaccessibility nor growing inequality are acceptable considering that universalism is one of the most beneficial values of scientific research.

The day that articles that stay solely behind a pay wall will have lost their relevance may be edging closer faster than anyone expects.

There are free avenues to OA and cost should certainly not be seen as a barrier. Green OA is free, and the majority of publishers accept that papers can be self-archived in one form or another (pre-print, post-print with final revision, or PDF) with no delay. Moreover, two-thirds of Gold OA journals do not levy author processing charges.

Many mandates being promulgated at the moment run the risk of favouring a shift from inaccessibility to inequality.