OECD publishing

Please cite this paper as:

Boselli, B. and F. Galindo-Rueda (2016-09-22), "Drivers and Implications of Scientific Open Access Publishing: Findings from a Pilot OECD International Survey of Scientific Authors", *OECD Science, Technology and Industry Policy Papers*, No. 33, OECD Publishing, Paris.
http://dx.doi.org/10.1787/5jlr2z70k0bx-en

OECD Science, Technology and Industry Policy Papers No. 33

Drivers and Implications of Scientific Open Access Publishing

FINDINGS FROM A PILOT OECD INTERNATIONAL SURVEY OF SCIENTIFIC AUTHORS

Brunella Boselli,

Fernando Galindo-Rueda

OECD

FOREWORD

This document was authored by Brunella Boselli and Fernando Galindo-Rueda from the Economic Analysis and Statistics Division (EAS) at the OECD Directorate for Science, Technology and Innovation (STI). This draft has benefited from helpful suggestions and feedback from Giulia Ajmone-Marsan, Laudeline Auriol, Mario Cervantes, Alessandra Colecchia, Tomohiro Ijichi, Daniel Ker, Greg Peterson, and Carthage Smith. Brigitte van Beuzekom provided assistance with the bibliometric data. The authors wish to thank survey participants and researchers who contributed to testing the questionnaire. Any errors or omissions are the authors' sole responsibility.

A previous version of this document was presented and discussed by the OECD Committee for Scientific and Technological Policy (CSTP) and its Working Party of National Experts on Science and Technology Indicators (NESTI). This document has also been presented to and discussed by members of the OECD Global Science Forum (GSF) and the Working Party on Innovation and Technology Policy (TIP). This report was approved and declassified by the CSTP on 22 October 2015 and was prepared for publication by the OECD Secretariat.

Note to Delegations:

This document is also available on OLIS with the cote:
DSTI/EAS/STP/NESTI(2015)4/FINAL

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 document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area. The statistical data for Israel are supplied by and under the responsibility of the relevant third party data providers. 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. Latvia was not an OECD member country at the time of preparation of this publication. Accordingly, Latvia does not appear in the list of OECD member countries.

© OECD 2016
You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for commercial use and translation rights should be submitted to rights@oecd.org.

TABLE OF CONTENTS

FOREWORD....2

ABSTRACT....4

EXECUTIVE SUMMARY....5

The study....5
Main findings....5
Implications from the pilot study for the data infrastructure on scientific research....7

DRIVERS AND IMPLICATIONS OF OPEN ACCESS PUBLISHING: FINDINGS FROM A PILOT INTERNATIONAL SURVEY OF SCIENTIFIC AUTHORS....9

1. Introduction....9

2. Survey design and implementation....11

2.1. Related studies and literature....11
2.2. Survey design....13
2.3. Sampling approach....14
2.4. Survey testing and launch....15
2.5. Response patterns and potential impacts....16

3. Main results on access to research outcomes....20

3.1. Access status to authors' reference publication....20
3.2. The research context for open access....28
3.3. Open access and citation impact....31
3.4. The marketplace for dissemination and quality assurance - publishing charges....34
3.5. Scientist demand for open access –contingent valuation estimates....37
3.6. Disclosure, access to data and research reproducibility....42

4. Perspectives on the profiles of scientific authors....46

4.1. Affiliation of scientific authors....46
4.2. Age and gender profile of scientific authors....46
4.3. Research status....49
4.4. Educational background of scientific authors....50
4.5. Sources of funding for research....52
4.6. Mobility patterns of scientific authors....53
4.7. Scientist earnings....55

5. Concluding remarks and next steps....57

ANNEX 1. SURVEY METHODOLOGY....59

REFERENCES....65

NOTES....67

DRIVERS AND IMPLICATIONS OF SCIENTIFIC OPEN ACCESS PUBLISHING: FINDINGS FROM A PILOT INTERNATIONAL SURVEY OF SCIENTIFIC AUTHORS

ABSTRACT

This paper presents the results of a new and experimental initiative aimed at investigating the research and publishing activities of scientific authors and the extent to which new approaches can be used to collect data on scientist behaviour and impact. This email-based pilot survey targeted a stratified random sample of corresponding authors with publications listed in 2011 in a major global scientific publication index. The results provide evidence on the extent of journal and repository-based open access (OA) (55% of documents are available through either channel), data sharing practices, the nature of the citation advantage conferred by different forms of OA, and its relationship with the quality accreditation role played by journals and their reputation. Under the prevailing scientific peer review and publication system, revealed and contingent author willingness to pay for OA is proven to be sensitive to funding availability and to publication charges, up to modest amounts. The results point to a potential decoupling of access and quality assurance objectives and the need for policy makers to consider the role of economic incentives and social norms when considering different policy options. The results also provide evidence on pathways for societal impacts of science and on the extent of scientist mobility and career progression, including on the extent of gender pay bias notwithstanding similar citation impacts.

Keywords: science, authors, research, open access, open science, survey.

EXECUTIVE SUMMARY

The study

This paper presents the preliminary results of a new initiative aimed at investigating the research and publishing activities of scientific authors and the extent to which new approaches can be used to collect statistical data on scientist behaviour and impact. This pilot survey was based on a stratified random sample of corresponding authors with publications in one or more of seven hand-picked but diverse science domains that had been published in 2011 and listed in Scopus, a major global scientific publication index.

This study pursued two main research objectives:

- Obtain statistical evidence on the activity of scientific publishing and access to outputs of scientific research (open access –OA– and open science –OS–) in fulfilment of commitments to carry out empirical work in this area to complement policy-based work being undertaken within the OECD Directorate for Science, Technology and Innovation under the aegis of the OECD Committee for Scientific and Technological Policy (CSTP).
- Contribute to defining the OECD research agenda on the activities of scientific researchers and their impact, by assessing the case for developing a complementary statistical infrastructure that can be flexibly used to meet a wide range of current and future user interests.

The survey took place early in 2015 following earlier pre-testing of a batch of one thousand survey requests in December 2014. The preliminary round of analysis took place in February-March 2015 - the outcomes of which are reported in this document. As a first of its kind, some degree of caution is required when interpreting the survey results, which should be treated as experimental. Assessed against conventional quality measures, the technical outcome of the survey data collection can be described as in line with expectations for this type of survey, population and resources available. Nearly 61 thousand scientific authors (corresponding authors) were targeted, with approximately 10% complete responses collected from them. The overall response rates were broadly comparable with those of other multi-country initiatives of this kind among similar populations. Weight adjustments based on limited information for this population for non-response bias provide a partial correction mechanism.

Main findings

The main findings from the pilot survey and its analysis can be tentatively summed up as follows. Based on this study's estimates of the incidence of open access:

- Approximately 50-55% of documents are openly available 3-4 years after publication. Documents in fields receiving higher levels of commercial funding for research are less likely to be available on an open access basis.
- Authors from emerging and developing countries tend to rely more on open access journals than their OECD counterparts for open access publishing. Accounting for open repository-based access reduces observed OA differences across countries based solely on journal status.

- Publisher and repository-based access to documents are intertwined. Their relative importance varies by field, and it is associated with the prevalence within a field of major and reputable repositories for working papers.

Concerning the impacts of open access, the results lend support to the hypothesis that OA policy choices have to be considered within the broad incentives set out by the peer review and accreditation system for scientific research.

- The evidence found is consistent with the hypothesis of citation advantage for OA. Openness is robustly associated with higher citation impact, but the choice of OA vehicle is related to different dissemination/citation impact channels.
- Thus, citations in peer reviewed articles covered by major indices (both as self-reported by the author and as calculated from Scopus) see no boost from journal/publisher-based OA, but are significantly higher in the case of repository-based OA. Journal prestige, as implied by its past citation record, is a key but imperfect determinant of article citation rates.
- Citations in other literature, working papers, or user documents (policy, technical, media, etc…) appear to be significantly boosted by open access, while journal prestige plays a much less important role.

The analysis of author processing charges indicates that authors publishing in OA journals are more likely to pay for review and publication, and that payment of fees does not seem to reflect measures of review quality nor journal prestige. “Prestige” appears to provide journals with the ability to recoup costs through subscriptions while benefiting from unpaid contributions by reviewers.

The survey has also facilitated an investigation of authors’ latent demand for OA for their articles by carrying out a contingent valuation exercise in which their willingness to pay for ensuring publisher-based OA to documents was tested using a methodology frequently applied in the valuation of private and semi-public goods. Contingent valuation provides a lower bound estimate on the overall (private to author and external) value of OA:

- Approximately only 15% of authors who did not pay author processing/publication charges (APCs) would have been willing to pay USD 299 for OA – a figure that drops to 6% when the randomly proposed APC is USD 999, or increases to 30% when the amount proposed is USD99. This is consistent with standard, price sensitive demand curves in economic textbooks, and suggests that access is underprovided given its marginal cost.
- Scientific authors in the life/health sciences domains have the largest willingness to pay for OA, while those in the humanities and business are willing to pay the least.
- Our analysis also shows significant differences in willingness to pay for small differences of 1 USD per proposed amounts reflecting the existence of “sweet” pricing spots that are as a matter of fact used by publishers who tend to avoid rounded price points.
- Authors who already paid an APC (but not giving right to journal/publisher OA) have a somewhat higher willingness to pay for open access.
- Willingness to pay is very sensitive to the prestige of the journal, and a high percentage of individuals who refuse to pay the proposed amount would have been willing to pay in the hypothetical case of acceptance in a higher ranked journal within their domain.
- Refusal to pay is often explained on the basis of limited available funds, and willingness to pay is larger among those with higher earnings (as measured at the time of the survey) and

report having access to dedicated resources for dissemination. This is relevant towards the policy analysis of how to fund OA systems.

Our analysis also provides some relevant evidence on a number of characteristics of scientific authors listed as corresponding contacts: Women are heavily under-represented in this population, more so than in other statistics on researchers and doctorates. The average corresponding author is affiliated to a higher education institution, and is typically in his late 40s/early 50s, but there are significant variations in age and sectoral affiliation across countries. Full time employment with partial dedication to research is the most common form of labour market engagement. Most authors were grant-funded by governments for their work but institutional funds also played a significant role. Contract research and funding from business is largely confined to the fields of chemical engineering and materials science among those fields covered in this study.

There is a very large heterogeneity across countries in the share of resident authors born elsewhere. Those with the largest share of foreign born authors also tend to have the largest net attractors of talent, if one compares their shares on a residence and birth country basis.

We conclude our analysis of author's characteristics and open access publishing by estimating the determinants of reported scientist earnings. Our results point to a significant earnings penalty on women which does not disappear when accounting for several other attributes, including measures of scientific impact or productivity on which male and female corresponding authors do not appear to differ. Earnings are also lower for those affiliated to higher education institutions, those in fixed term contracts and those fully dedicated to research. We observe a general premium on age and international mobility. We also observe a high premium on the prestige of the journal in which the reference document was published, and a small additional effect from actual citations which disappears one we account for whether the individual is a reviewer and/or an editor. Among the self-reported measures of citation vehicles, peer-reviewed journals and policy documents are the only ones showing a significantly positive correlation with earnings.

Among this wide range of results, one key theme emerges from the analysis of the data based on prima facie evidence on the role of career incentives, credentials, prestige and peer review. These stand out as major contextual factors for examining demand for open access and the potential impact of policies. Notwithstanding several comments about the usefulness of knowledge dissemination, authors' preferences and behaviour are fully aligned to the goals that the market in which they operate values the most - namely publications and in the highest possibly ranked journals.

Implications from the pilot study for the data infrastructure on scientific research

The choice of OA as the subject of this pilot has helped deliver numerous insights on the barriers and incentives that policy makers need to take into account when designing open access and open science policies. These findings can further support the parametrisation of cost-benefit analysis that so far had to rely on untested assumptions.

The implementation of a survey linked to bibliometric information has also allowed for testing the consistency of different data sources and expanding our knowledge about dissemination mechanisms for which OA can be relevant. This instrument has the potential to become promising approach for extending measurement practice beyond currently used indicators, towards indicators that reflect in a more comprehensive way the full breadth of impacts from scientific research and the factors that shape its production, dissemination and use.

The ability to consider several countries through a single study with identically formulated questions is another major advantage in a world in which access to micro-data on researchers is still limited and cannot be pooled effectively across countries or linked to other sources for analytical purposes.

Within the resource envelope available for this project and with competing pressures, a number of quality improvements could not be reliably implemented within the available time. Given the wealth of evidence and results gathered, the initiative may be deemed to represent good value for money as a second or third best-solution with respect to the ideal of a distributed data collection approach involving national statistical agencies willing to share and pool their micro-data.

The intention of the approach tested in this pilot is not to replace standard approaches but complement them, acting as a form of continued data collection experiments from which standard surveys can draw inspiration and bring new concepts and ideas into the mainstream. In line with the major themes developed in the Blue Sky Forum of 2006 (OECD, 2007) and the 2010 OECD Innovation Strategy (OECD, 2010), the approach tested in this pilot exercise could be in principle applied to other major in-demand policy topics for which traditional instruments currently cannot offer appropriate solutions. The process leading to the identification and prioritisation of those topics can also provide a unique opportunity to strengthen the communication between the policy and statistical communities within the OECD.

This performance can be deemed to provide a lower bound for potential future efforts, bearing in mind the limited financial and time resources available to conduct this pilot. Subject to resource availability, a number of modifications and additional control steps can be introduced in possible future editions in order to yield significant improvements in standard survey quality measures.

i) implementing a more thorough cleaning of the underlying database and seeking to identify more recent email addresses for the authors through additional search work;

ii) implementing a shorter questionnaire;

iii) allowing respondents to respond to the survey in the language of their choice;

iv) identifying an appropriate mechanism for non-response follow-up contacts;

v) referring respondents to the results of previous related studies such as this pilot;

vi) developing an OECD internal survey tool with internal server hosting to increase user trust in the survey;

vii) securing national institutional endorsement that is visible in the survey tool;

viii) evaluating how to provide incentives to further motivate individual response.

Taking into account the significant scope for improvement, the pilot has resulted in the collection of a unique global set of information on the characteristics, track record, behaviour and preferences of scientific authors that is also linked to standard bibliometric indicators. Given the wealth of information collected and the range of analysis enabled by this survey, the approach can be considered as a promising and potentially valuable addition to the OECD toolbox for the statistical analysis of major science and research policy questions. The OECD plans to conduct a second pilot on issues identified as of high relevance in the recent declaration of science ministers who met in Dajeon (Korea) in October 2015.

DRIVERS AND IMPLICATIONS OF OPEN ACCESS PUBLISHING: FINDINGS FROM A PILOT INTERNATIONAL SURVEY OF SCIENTIFIC AUTHORS

1. Introduction

This report presents the preliminary findings of a new and experimental project aimed at investigating the research and dissemination activities of scientific authors and the extent to which new data collection approaches can be used to provide new insights on scientist behaviour and impact. The OECD Committee for Scientific and Technological Policy (CSTP) and its working parties have recently been carrying out policy work on subjects related to the production and dissemination of scientific knowledge and the role played by scientists (OECD, 2015a,b). Some of the statistical evidence produced by OECD and its Working Party of National Experts on Science and Technology Indicators (NESTI) on human resources on S&T has made considerable contributions. However, several gaps remain when it comes to addressing questions often posed in the context of science policy decision making at a country or global level in this rapidly changing environment.[1]

A number of limitations often apply to collecting empirical, representative and comparable evidence on a global scale. In the case of evidence from surveys conducted within countries under a common model, the range of phenomena and policy questions that can be analysed is limited as surveys carried out on a country by country basis cannot, for confidentiality reasons, be combined to generate pooled micro-data on which to explore individual, institutional and country characteristics in a consistent fashion. Furthermore, science is a highly globalised activity and some of its key features, spanning across national boundaries, cannot be easily captured by instruments with a purely national scope. Links and knowledge flows are particularly hard to measure through instruments that target subjects (OECD, 2013). This is even harder when the subjects of analysis move from one jurisdiction to another and where their outputs can be easily transferred across boundaries. In contrast, standard bibliometric data-based analysis, while sometimes global in scope and able to trace a number of relevant linkages, is constrained by the scholarly and administrative processes that give rise to the available data, reflecting only one specific dimension of the activities of scientists. Citations are for example too often identified as evidence of impact and knowledge flows, when the reality can be significantly more complex and diverse. Reliance on evidence that captures a single and partial dimension of knowledge flows may lead to badly informed policies.

In 2014, NESTI discussed and agreed on the piloting of a new survey-based initiative with the overarching goal of testing the feasibility of implementing a survey instrument to collect evidence from publishing scientists worldwide on matters of relevance to the CSTP policy making community. This decision pursued two main research objectives:

- Obtain statistical evidence on the activity of scientific publishing and access to outputs of scientific research (open access –OA– and open science –OS–) in fulfilment of commitments to carry out empirical work in this area to complement policy-based work being undertaken within the OECD Directorate for Science, Technology and Innovation under the aegis of the OECD Committee for Scientific and Technological Policy (CSTP).

- Contribute to defining the OECD research agenda on the activities of scientific researchers and their impact, by assessing the case for developing a complementary statistical infrastructure that can be flexibly used to meet a wide range of current and future user interests.

A key advantage of implementing a centralised data collection, as opposed to the more standard approach of distributed data collection and analysis, is that, by having full control over its design and execution, it is possible to minimise biases arising due to variation in country-level implementation. It is also a potentially cost-effective way of collecting statistical data for micro analysis. Without a pooled set of micro-data for multiple countries, identifying the impact of country-level features is significantly more complicated.[2]

Developing and putting into practice a new survey instrument is always a challenging task, particularly if resources are very limited. Targeting for the first time an expert population that typically receives a considerable influx of messages can be difficult if the organisation behind the survey is not yet widely recognised by target respondents as an authoritative and impartial actor in this area. Scientists commonly receive several email surveys and messages inviting them to join conferences, submit papers and join editorial boards for a wide range, often newly created journals which they know little or nothing about. Few scientists have assistants who deal with their email correspondence, and as result, many messages naturally go dismissed or ignored unless they manage to capture the interest of the targeted respondent through the subject heading. It is fundamental to be able to appeal credibly to some form of individual responsibility in sharing the information requested. Without any form of reporting enforcement mechanism, the challenge is to make respondents realise that that by sharing their experience better policy decisions can be made at a national or global level. Other potential motivations such as the ability to benchmark themselves against peers were not considered so far, as they require a more considerable investment on the survey and data platform.

This pilot's working hypothesis for the concept of an international survey of scientific authors (ISSA) was that the OECD could eventually be perceived, relative to other institutions or individual researchers, as a trusted producer of policy-relevant statistics, and that this could position it to carry out a high quality voluntary global data collection. One challenge is that the broad visibility of OECD is more often linked to other policy areas (such the PISA study for education or country-based economic projections and structural reform recommendations). In contrast, statistics on R&D, science and innovation, have never before been collected directly by the OECD but by national agencies following OECD standards, without necessarily refer to OECD as such in their surveys. The better known STI-measurement standards are primarily aimed at collecting data from institutions, not individuals, with relatively recent exceptions such as surveys conducted under the model provided in the OECD/Eurostat/UN guidelines for collecting data on the Careers of Doctorate Holders. In summary, it was foreseen that respondents would not necessarily immediately recognise the OECD as a relevant direct collector of information on this subject and that this might dampen response rates in the pilot.

An important consideration was the avoidance of avoidable overlap with the activities of national statistical organisations (NSOs). In this respect, the scope for overlap was deemed to be very small. Career of Doctorate Holders (CDH)[3] surveys (or variations thereof for master's level students or graduate cohort surveys) for example target the population of individuals with doctorate degrees, not all of which publish in scientific journals. It is also the case that not all individuals contributing to scientific publications have doctorate degrees. These two activities are therefore highly complementary. Some countries carry out surveys of academics with the purpose of improving their estimates of statistics on higher education, dealing for example with time use in academia between research, teaching and other functions, but these are often of an *ad hoc* nature and the targeted nature of our pilot minimised the risks of impact on national official surveys. Larger overlaps could be found with respect to ad hoc studies such as the Mobility of Researchers (MORE) surveys[4] sponsored by the European Commission, or other academic or publisher-driven studies focusing on the same type of population.

The experience of this first pilot survey launched at the beginning of 2015 has revealed both the strengths and weakness of our experimental approach. Response rates have been low yet at a realistic level

expected for a survey of this type, with very marked differences across countries that in our opinion largely reflect the degree of awareness and perception of the OECD and the constraint of not being able to provide a multilingual survey questionnaire. The wealth of information about both respondents and non-respondents facilitates the implementation of statistical adjustments to the data to account for varying propensities to respond. Those adjustments prove to have limited impact on some of the main behavioural estimates and correlations, supporting the analytical value of the resulting database. The pilot nature of the collection and sample sizes restricts the detail that can be provided given the precision of estimates required.

With all its limitations, the survey has resulted in a unique and detailed repository of information about the research and dissemination activities of individual authors from all over the world. This information has been used to enquire about the extent to which scientific documents are openly available and in which forms, about the factors that contribute to or hinder openness, and the potential impacts. New exploratory analysis has also been undertaken to infer the scientific community's demand for open access to their publications. This analysis sheds new light on the role of career incentives and peer review mechanisms in driving behaviour, and the importance of user channels that are not captured by traditional bibliometric indicators.

This report is structured as follows: Section 2 describes the design of the survey and its implementation. Section 3 presents the study's main results in relation to access to research outputs, including descriptive statistics and econometric analysis examining the relationship between individual experiences, behaviours and journal and article attributes. This section also presents the findings of an experimental exercise aimed at identifying authors' willingness to pay for open access. Section 4 presents some new evidence on the characteristics of scientific authors and their careers. Section 5 concludes.

2. Survey design and implementation

2.1. Related studies and literature

Surveys of scientific authors

A number of surveys have previously targeted the population of scientific authors as the subject of statistical survey analysis, with variations in country coverage and scope. Selected examples of studies with similar design features to our project comprise those by Franzoni, Scellato and Stephan (2012), who targeted a panel of corresponding authors for articles published in 2009 in four fields of science within one of 16 core countries, focusing on the mobility patterns of scientific authors, and the Hitotsubashi-NISTEP-Georgia Tech Scientists' Survey in Japan and the United States (Nagaoka et al, 2011).

Other surveys have dealt specifically with the subject of open access in its various possible modalities. These surveys are discussed in OECD (2015a) and tend to focus on eliciting opinions from respondents about the implications of open access, for example, the EU-funded studies on Open Access Publishing (SOAP: http://project-soap.eu/) and Permanent Access to the Records of Science in Europe (http://www.parse-insight.eu/). Within the SOAP project, a large-scale survey of the attitudes of researchers on open access publishing found a positive attitude among researchers towards open access publishing, higher in the humanities (90% of positive responses) than in the natural sciences (around 80%) (Dallmeier-Tiessen et al., 2011). Most of the respondents identified funding as the main barrier to open access, followed by the quality of open access journals.

A 2014 survey on Norwegian researchers carried out by (DAMVAD, 2014) on attitudes to sharing and archiving of publicly funded research data covered nearly 1,500 researchers affiliated to Norwegian research organisations. The study found that Norwegian researchers frequently use and share research data:

64% of the respondents had used research data from other researchers in the last three years. Most of the respondents obtained the data from other researchers in the same institution. Around 80% of respondents agreed that open access to data improves the research process, with similar proportions for facilitating education and promoting collaborative research. However, survey respondents identified the efforts to prepare data-sets for release, technical infrastructures and potential impacts on producing further publications in the future as main challenges. A global survey on open access by academic publishers Taylor & Francis (Taylor and Francis, 2014) reports that wider circulation and higher visibility of articles are cited by authors as major advantages of OA publishing. This study presents an opposing view on whether authors believe that OA drives innovation in research and helps increase citations and less favourable attitudes towards data sharing. Slightly more than half (52%) of the researchers who participated in the survey do not use repositories to make their papers available to others. When they do, the most-often used repositories are either institutional repositories or their personal or university webpages and main reported reasons have to do with personal willingness to make research results available, rather than requirements from funders. This study also reports on attitudes towards citations versus other usage metrics as means to evaluate research output. More recently, Wiley Publishers contacted 90,000 researchers across a wide array of disciplines and received more than 2,250 responses from individuals engaged in active research programs in relation to questions on data sharing,[5] in which 52% of respondents said they had made their data publicly available, the largest proportion via supplementary material in journals.

A recent study by Zinner et al (2016) surveyed separate samples of life scientists at US universities receiving the most National Institutes of Health funding before and after a series of major reforms incentivising the use of open access repositories. Respondents were asked to estimate the number of requests for information, data, and materials they made to and received from other academic researchers in the past three years. They were also asked about potential consequences of sharing and withholding. The authors found that the proportions of authors who received, denied, made, or were denied at least one request were not statistically different. However, the total volume of requests received from or made to other scientists dropped substantially over the two waves, suggesting a possible move away from peer-to-peer sharing.

Other sources on open access

Open access has also been studied in the literature by means of indirect approaches. Publisher based OA (also defined in the literature as “gold” OA) is in principle an attribute of journals and can be assessed fairly easily by comparison with inventories of OA journals. However, variation can be found at the article level within a journal as a charge may be levied as standard practice by the journal or as a premium in order to render the document publicly available online (Laakso and Björk, 2012). Open repository archiving (green OA) is also difficult to establish because it depends on author’s and third party decisions and may apply to slightly different versions of the same document. Some studies have investigated data from the journals databases such as Sherpa/Romeo covering publisher OA policies (e.g. Miguel et al, 2011) and found that an additional percentage of journals explicitly allow self-archiving manuscripts. Publishing in such journals is neither a necessary nor sufficient condition for the document to be publicly available. The detection of a publication’s open access status has been more recently studied by means of computer-implemented online queries on a large sample of documents, which seek to identify whether a (near) identical version is freely available online (for example, see Gargouri et al., 2010 and 2012). Information on the access status from publishers is obtained directly from the publishers’ information systems.

Bibliometric and other type of scientometric analyses, such as web-presence and download statistics have been used to investigate the potential drivers and impacts of different access models. The challenge is twofold: a) whether the measures of “impact” fully capture the knowledge flow mechanisms that result in

societal value; b) whether it is possible to identify the causal effect of openness on "impact" measures over and above other types of confounding effects. Our study provides some new insights on this issue by enquiring on other impact mechanisms but provides only correlational type evidence as no quasi-experimental settings are readily available.

2.2. Survey design

The concept of an OECD international survey of scientific authors on the dissemination of scientific research began to be developed around November 2013 in preparation for a workshop of the OECD Working Party on Innovation and Technology Policy (TIP) on Open Science and internal discussions on the main evidence gaps concerning our understanding of the drivers and impacts of open access to a range of outcomes of scientific research. A previous OECD project on mapping knowledge flows (OECD, 2013) had already considered issues of measurability of multiple types of knowledge flows and the role of access policies and infrastructures. This project also pointed to the relevance of aiming for increased integration of administrative and survey-based sources and highlighted a number of relevant initiatives in this area. The availability of contact information for corresponding authors in the bibliometric database used by the OECD secretariat provided an opportunity to exploit synergies between different data sources. In preparation, different online survey software solutions were investigated, considering the pros and cons of different options given the lack of an adequate OECD corporate survey tool for the type of survey considered. In early spring 2014 the OECD secretariat began to prepare an outline questionnaire aiming to combine a good coverage of the questions of interest to the TIP-led project on Open Science and the experience accumulated in developing recommendations for surveys on the Careers of Doctorate Holders.

In particular, the design efforts were channelled towards developing an instrument aimed at capturing the experience of scientific publishing and the salient features of the publication context that are relevant to the analysis of incentives, costs and benefits of alternative access models. Key questions related to:

- Collecting a rich set of demographic characteristics, and complement this with information available from existing registers.
- The role of peer-reviewed publications as a source of knowledge, and access restrictions to documents and other research materials experienced.
- The funding of research work and the types of conditions on use of research outputs.
- The dissemination strategies adopted by authors and the access status to documents and other outputs such as data.
- The perceived added value of the peer review and editorial process, and preferences towards the trade-off between access, impacts (in its multiple forms) and costs.

As a general approach, it was considered very important to ensure that interest in open access did not bias participation and drive changes in perception of OECD statistics and empirical analysis. As part of this, potentially "loaded" opinion questions were avoided.

From the perspective of first-in-kind pilot data collection at OECD, the survey aimed at testing the feasibility of OECD directly collecting data from the population of scientists, with the aim of taking a first step towards building credibility and "recognition" through the outcomes of the pilot. Indirectly, it was also intended to use the pilot to help identify promising survey research topics, develop internal competencies on survey implementation and analysis to allow the secretariat to address potential future committee evidence needs, with this or related populations, and learn relevant lessons from the pilot's experience, participant feedback and any shortcomings and design and implementation errors.

A draft questionnaire was put together by the project team with input from policy colleagues in the secretariat and presented to NESTI in June 2014. Delegates agreed on the relevance and merit of this initiative for OECD, in particular bridging the gap between survey and other data sources, while noting the demanding timetable set out for the pilot survey and the potential implementation challenges to be faced. For resource reasons, the actual survey implementation had to be postponed until after December 2014 when the questionnaire and the sampling approach were finalised. The questionnaire is available in the project's webpage (oe.cd/issa).

2.3. Sampling approach

Aiming to capture a diverse range of fields but at the same time contain the scope of the pilot to a limited, manageable size, the survey targeted Scopus-listed scientific authors publishing in 2011 in journals allocated into either of the following seven domains according to the All Science Journal Classification (ASJC) system:

Table 1. Selected fields for sampling

Selected field	ASJC code
Arts and Humanities(all)	12
Business, Management and Accounting(all)	14
Chemical Engineering(all)	15
Immunology and Microbiology(all)	24
Materials Science(all)	25
Neuroscience(all)	28
Physics and Astronomy(all)	31

Source: ASJC classification. http://www.elsevier.com/online-tools/scopus/content-overview

The domains in **Table 1** were handpicked on the basis that they capture a diverse group of fields across all broad domains of knowledge, ensuring coverage of the social sciences and the humanities, as well as attempting a balance between basic research, applied research and experimental development-oriented fields. These selected fields account for approximately 20% of total scientific production recorded by Scopus for 2011. The sampling procedure started with an extraction of the full set of Scopus documents in at least one of these fields which were published in 2011, and whose corresponding authors had a valid Scopus author ID. This extraction is based on Elsevier's 2014 Scopus custom database used under licence by the OECD.

Records corresponding to patently wrong email addresses were discarded, but no hands-on-cleaning of email addresses took place for resource reasons. We chose not to exclude any specific affiliation country from this pilot. We undertook some minimal editing of the affiliation country field, imputing country names on the basis of country domains in the email address (e.g. ".au" for Australia; ".be" for Belgium, etc. and we further identified a number of US-based authors by those having the ".edu" email domain).

The data selection process first addressed the issue of multiple attribution of journals to more than one subfields within the 7 fields selected, by choosing a single one of the possible attributions at random, resulting in 554 256 unique documents. Next, a document was chosen, also randomly, for each author, resulting in 379 218 author-documents pairs. With an indicative target sample size of approximately 60 000 authors, a proportionate quota sampling approach was implemented based on a stratification process by country of institutional affiliation and field. An absolute quota was introduced per stratum in order to ensure that, whenever possible, at least 30 authors were targeted within a given stratum – there were

31 strata with fewer than 30 observations in them, accounting for 483 author-publication pairs. This selection resulted in a total number of 61 361 authors being targeted.

The online questionnaire was delivered through the SurveyGizmo® platform. For time and resource reasons, the survey was launched in English only. Although the platform allowed for multi-language implementation, it was not possible to translate it with the resources available. This shortcoming was expected to reduce participation from countries where knowledge of English is less common. Knowledge of English among the global scientific community is widespread but far from complete and Scopus indexing language requirements relate only the availability of an abstract in English. Implementing a multi-language survey can also be challenging. The underlying database only identifies the affiliations of an author at the time of the publication, not his/her country of residence, nationality or language of preference. This makes it difficult to tailor the survey invitation emails in a precise way.

2.4. Survey testing and launch

The survey questionnaire was initially tested in November 2014 with 10 authors from different countries being asked to complete the survey and then debriefed on the experience. Comments were also collected from NESTI delegates. Some of the questions were also tested by Statistics Canada in the context of its testing of an academic time use questionnaire, but it was not possible to secure the results for its own testing before the pilot had to be implemented.

Before the complete survey launch, 1 000 individuals within the sample frame were initially approached to complete the survey in early December 2014. In this testing sample, we already identified the extent of bounce rates and the challenges in ensuring that email recipients clicked on the link that opened up the survey for completion. Conversely, those who clicked appeared to have little problem in completing the survey almost in its entirety. We therefore focused our efforts in making the invitation cover email as attractive and engaging as possible, rather than reducing the survey length. Communicating the relevance of survey completion to potential respondents was challenging as knowledge about OECD activities in this area (in terms of data collection from scientists and reporting about them) appeared to be rather limited and it was not possible to refer to related OECD studies. To help mitigate this problem, a webpage (www.oecd.org/sti/survey-of-scientific-authors.htm) on the project was developed explaining the research objectives and providing links to other related and relevant work and email (surveyonscience@oecd.org) and telephone contact information were also provided.

The survey was launched to the entire target sample in January 2015, with substantially revised introductory messages and a more personalised approach. Two reminders were sent over the course of the following two weeks. Based on evidence of higher response rates for online surveys on Monday mornings, it was decided to go for a late Sunday launch. A third reminder was considered but decided against as it could encourage additional un-subscription requests and increase the scope for bias in the achieved sample. No further resource was available to conduct a non-response follow up by phone. An email-based approach would have been unlikely to provide any satisfactory results from previous non-respondents.

A number of respondents contacted the OECD, through the dedicated contact address provided within the survey and the cover email, but also through generic OECD email and switchboard, to verify that the survey was genuinely coming from OECD. Although the latter was not initially foreseen, the messages were eventually forwarded to the survey team. This experience confirmed the potential role for this pilot in providing a first step towards addressing lack of scientist awareness of OECD and its outputs in the area of science, technology and innovation. In future surveys, it should be possible to point directly to results from this survey to increase perceived authenticity and relevance among target respondents. In that respect, the pilot has met one if its objectives, namely to provide a first batch of survey evidence to build upon and refer to in future studies and surveys.

A possible additional approach to address this problem would be for future surveys to secure and indicate endorsement from a relevant national institution with responsibility for science policy and implementation. It is however difficult to find a mechanism to assign a specific official endorsement (e.g. country based) to an individual respondent. For example, should the relevant ministry be linked to the author on the basis of her or his last recorded affiliation? Leaving aside the complexity of securing endorsement from organisations in multiple countries, it is also not entirely clear whether a national institutional endorsement would uniformly encourage participation among its scientist community. For example, scientists in disagreement with national policies may opt not to reply if such an endorsement is included in the survey cover.[6]

In addition to verifying the authenticity of our emails, targeted authors also had other types of questions and comments that required considerable survey manager time to respond to. Common issues raised concerned the fact that the survey asked about documents that did not represent themselves peer-reviewed articles, but reviews and introductory documents which by mistake were not fully eliminated from the scope of the survey. This made some parts of the questionnaire less relevant to some respondents and might have been avoided with additional but relatively straightforward data cleaning at the early stage of the project. Respondents that noted this problem dropped from the survey as they found the questions not to apply to them. This is reflected in the non-response rates and suggests considerable scope for methodological improvement.

Some respondents also candidly shared with us their personal and "philosophical" objections to the survey; for example, one respondent disapproved of what (s)he perceived as a "positivist" approach towards science underpinning the survey and even provided a number of suggestions. A couple of authors happened to be renowned business school experts in the analysis of science and technology. A few authors did not fully recognise themselves as corresponding authors when, on the basis of the records available and the actual publications, they clearly are presented as such.

2.5. Response patterns and potential impacts

Attaining adequate response rates is a major concern for survey-based research, particularly one in which survey completion is not mandatory. While response representativeness is a far more important concern than attaining a high response rate, (unit) response rate is important specially if it bears on representativeness for the types of inferences the data will be used for (high response rates also matter for precision). A number of diagnostics tests are provided on the survey response patterns in this subsection and in **Annex 2**, which contains detailed information on response patterns by country or economy of institutional affiliation.

The first barrier for an email-based survey to overcome is the possibility that the email bounces back. A bounce occurs when an automated electronic mail message is received from a mail system informing the survey sender (Surveygizmo in this case) about a delivery problem.[7] Respondents with affiliations in developed countries tend to exhibit relatively higher bounce rates and can be as large as 20% or even more in some countries such as Switzerland. The bounce rate may be driven by the author having moved and the email address in our records being cancelled, or actual blocking by the servers hosting the email. Author mobility is a major consideration for any study of this kind to take into account, for any mobile author is likely to have very distinctive features from the rest of the population with respect to the phenomenon under study. In future surveys, a possible strategy may be to trace the individual author's more recent affiliations and contact information for traces of mobility. However, there will always be a lag between the submission of information by the author to the publisher and the survey data that will impinge more seriously on the coverage of more mobile individuals. A possible but more resource intensive strategy would entail using social networks to try and retrieve the individual's most up-to-date affiliation.

For those emails that did not bounce back, the majority of targeted respondents that received the invitation email do not click through to the survey (82% out of this subgroup). This is the main source of survey non-response, and it is broadly in line with most types of online surveys, if not lower. The share of non-clicks greatly varies across countries. Among those who click through to the survey, the incidence of non-response is very limited but still significant; this may be the main area for potential improvement in future (e.g. through improved design). Shortening the questionnaire should also help reduce the extent of partial responses (3% of all target respondents).

Survey completion rates are lowest for authors with affiliations in the People's Republic of China (hereafter China), Hong Kong (China), Singapore, Korea and Chinese Taipei, possibly reflecting a combination of a lack of awareness of OECD as well as potential language problems, as no translation was provided for time and resource reasons within this pilot. A lack of national institutional endorsement may also have played a role in these particular countries. Language may have been a challenge for several authors with limited command of English, but our experience suggests this is not the primary driver of non-response across countries. Among English speaking countries, the United States which accounts for a very large share of corresponding authors has a lower response rate than Australia and the United Kingdom (**Table 2**). Several non-English speaking countries have much higher response rates, for example Italy at 19% or Spain at 17%. Japan's response rate is as high as the United States'. European countries, especially Nordic countries, also have high response rates and countries recently engaged in the accession process to the OECD have relatively high response rates (Chile and Estonia at 20%, Slovenia at 22%, Israel at 23%, Colombia at 16%), possibly reflecting media coverage driving awareness of the OECD (see **Annex 2** for details).

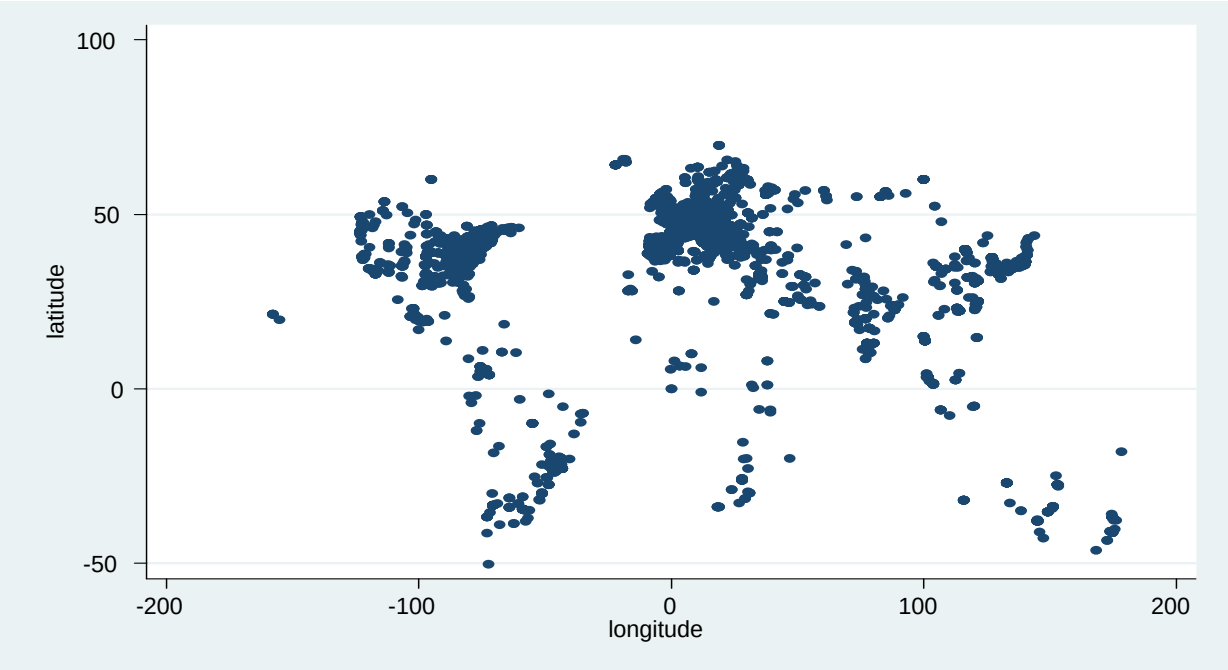
Table 2. Response rate by country of author's affiliation, countries with largest number of authors

Country	Targeted authors (Number)	Bounced (%)	No click (%)	Click, non-response (%)	Partial response (%)	Complete, response submitted (%)
USA	10962	11	76	2	3	8
CHN	9643	15	81	1	2	*2*
DEU	3243	18	66	2	3	**10**
JPN	3122	13	72	2	4	8
GBR	3032	17	69	2	2	10
Other countries	2413	10	67	4	6	**13**
FRA	2247	15	67	3	4	**11**
IND	2118	7	76	3	5	9
ITA	1756	6	67	3	5	**19**
ESP	1587	7	66	4	5	**17**
KOR	1567	6	85	1	4	*5*
CAN	1462	9	75	3	3	**10**
RUS	1394	12	68	5	5	**10**
BRA	1122	6	75	2	4	**13**
AUS	1094	11	69	3	3	**14**
Combined	60697	12	73	2	3	**10**
Combined, excluding bounces	53466	--	82	3	3	**11**

Source: OECD Pilot Survey of Scientific Authors 2014/15. For full table, see Annex section.

Overall, the survey provides very substantial global coverage. A scatterplot of respondent geo-coordinates (**Figure 1**) reveals the high concentration of responding authors in Europe and the United States, but also in localised regions all over the world.

Figure 1. Worldwide distribution of survey respondents



Source: OECD Pilot Survey of Scientific Authors 2014/15.

Response by fields

Different research fields exhibit very distinctive response patterns (**Table 3**). We observe significantly higher response (and completion) rates in the social science and humanities fields captured by the survey, while response is lowest is materials science, a field in which a significant share of authors work for the business sector.

Table 3. Response patterns by field of science

ASJC	Field	Targeted authors (Number)	Bounced (%)	No click (%)	Click, but non response (%)	Partial response (%)	Complete, response submitted (%)
12	Arts and humanities	5516	10	67	3	5	**15**
14	Business, management and accounting	5521	11	70	2	4	**12**
15	Chemical engineering	5980	11	73	2	4	**9**
24	Immunology and microbiology	7234	10	74	3	3	**9**
25	Materials science	12829	13	75	2	3	**7**
28	Neuroscience	6290	10	75	3	3	**8**
31	Physics and astronomy	17327	13	72	2	4	**9**
	Combined	60697	12	73	2	3	**10**
	Combined, excluding bounces	53466	--	82	3	3	**11**

Source: OECD Pilot Survey of Scientific Authors 2014/15.

Additional checks and controls for non-response

In light of the observed response rates, the representativeness of the resulting sample should be carefully examined. Starting with the type of response outcome (bounce, no click, clicked non-response, partial or full response), multinomial logit estimates indicate a slightly higher propensity for non-response among authors of articles in OA journals and/or with affiliation in OECD country or partner economy. Emails sent to authors of articles in higher impact journals are less likely to bounce, and more likely to lead to survey response. By disciplines, bounce is more likely for Materials science and Physics, while the likelihood of providing a response is highest for Arts/Humanities and Social sciences.

We also studied the changes in the sample composition induced by the use of reminders, as a form of wave analysis. The addition of further waves assisted substantially in increasing the coverage of authors affiliated to institutions in Brazil, China and India. Compared to survey "dropouts" (i.e. those who clicked through but did not submit a response), respondents tended to have more publications in 2011 and also publish in journals with a higher citation impact rating. For most countries, the use of reminders tended to make the composition of the achieved sample shift closer to that of dropouts and non-respondents, but while they encourage extra clicking, within survey drop-out rates also increased, from 4% in the first wave, 8% in the second, jumping to more than 50% in the third. It does seem that two reminders is the appropriate number to use.

As an attempt to correct for potential non-response bias, we evaluated the expected probability of survey completion using journal impact, OA status and authors' numbers of publications in 2011 to derive "predicted" estimates of response probability within each stratum. These estimates were used to construct adjusted sampling weights and provide a means for checking the extent to which any estimates are sensitive to the response patterns identified, based on those observable characteristics.[8] By and large, most estimates and especially regression results are not particularly sensitive to the choice between a pre-sampling and an adjusted weighting scheme. The adjustments appear to have a significant impact on the estimated distribution of country of current residence and birth, which may be indicative of some over-correction for survey under-coverage of authors based in China. Most descriptive statistics presented in this report are based on the adjusted-weighting scheme (very often both are jointly reported), while econometric regression and related analyses are un-weighted but control for strata fixed effects and all the characteristics used for modelling response rates.

For countries and economies with very low response rates, any estimates should be taken with considerable caution at this stage. It was decided not to supress such authors from the global estimates and as a general rule we only present statistics for individual countries when there is a large enough number of observations. In preparing these preliminary results, no imputations or specific weight adjustments were implemented for item non-response other than those implied by the sampling weight adjustment.

Overall assessment

It is not easy to benchmark the response results from this survey to other studies. Quality indicators are rarely reported in full. Furthermore, it is not appropriate to compare an initiative like this with established and well-resourced OECD initiatives with high global visibility such as the Programme for International Student Assessment (PISA) or the Programme for the International Assessment of Adult Competencies (PIAAC). Such surveys are also typically conducted by NSOs or other official organisations as part of a formal agreement with the OECD, and are based on a well-defined population with reference demographic characteristics (e.g. based on the census of population). They tend to attain between 40 and 60% response rates in their relevant generic populations. For thematically closer studies implemented on a distributed basis by NSOs, such as CDH, OECD guidelines recommend the use of non-response surveys if the response rate falls below 30%. However, no information has been made available to OECD on

response rates and potential biases in individual countries collecting such data. Furthermore, official surveys tend to first confirm with respondents their willingness to participate in the survey and subsequently send out a questionnaire. In cases where that initial agreement is not given, a replacement with similar demographic characteristics is found in the population. This approach helps boost response rates if measured conditional on a questionnaire being submitted.

Closer to our study in terms of population and methodology, a very recent Statistics Canada email-based survey targeting a population of academics, based on a contact list web-scraped by a private provider appears to have attained a 30% response rate[9]. Given the reputation of a leading national statistical organisation like StatsCan, this rate provides an upper bound as the most optimistic scenario for any multi-country, non-official and monolingual survey. An even closer possible comparison is with the EU sponsored MORE study of researchers in higher education institutions (MORE, 2011), which attained similar response rates for a majority of countries under the common scope of these surveys (see **Annex** 1).[10] Franzoni et al. (2012) attained significantly higher response rates of 35% for a selected group of countries without China and Korea where response rate was also much lower. Nagaoka et al. (2013) attained an overall response rate of 27% in their survey of US and Japanese authors. Unlike our study, these studies managed to effectively clean and remove all review and commentary articles.

Despite the many challenges encountered, the survey tool was fairly easy for respondents to use and the branching routines worked without problems. The survey was not unduly long to complete. The median respondent took 21 minutes from start to end, including possible interruptions, but it is acknowledged that a shorter and more professionally designed survey sent from an OECD server would have contributed to raising completion rates *vis a vis* partial responses. We found some evidence that individuals with higher earnings dedicated less time to completing the survey, a possible reflection of the opportunity cost of time for these individuals.

Additional, specific lessons from the survey implementation include the following recommendations:

- Focus on improving individual response "incentives" (i.e. the respondent's value proposition) and enhance the visual design of the contact emails to increase substantially the "click" rate.
- The content of the webpage dedicated to the survey is very important for credibility reasons, hence by being able to share results from a previous survey, responses can be further encouraged.
- Avoid grid/matrix-type questions, as these were found to encourage dropout behaviour.

The online survey tool needs to continue to be reviewed against alternatives. Unfortunately, there is no corporate OECD tool available for conducting surveys of this scale. As revealed by follow-up contacts, several targeted respondents did not appear completely persuaded that the Surveygizmo sender was doing so genuinely on behalf of OECD and some servers may systematically treat messages from its servers as spam.

3. Main results on access to research outcomes

3.1. Access status to authors' reference publication

Our questions did not pursue to establish directly the immediate gold (i.e. publisher-provided) or green (e.g. repository-based) access status of documents upon publication and preferred to focus on the current status of the document. Evidence on the access status of scientific publications from the authors themselves was retrieved through a series of questions:

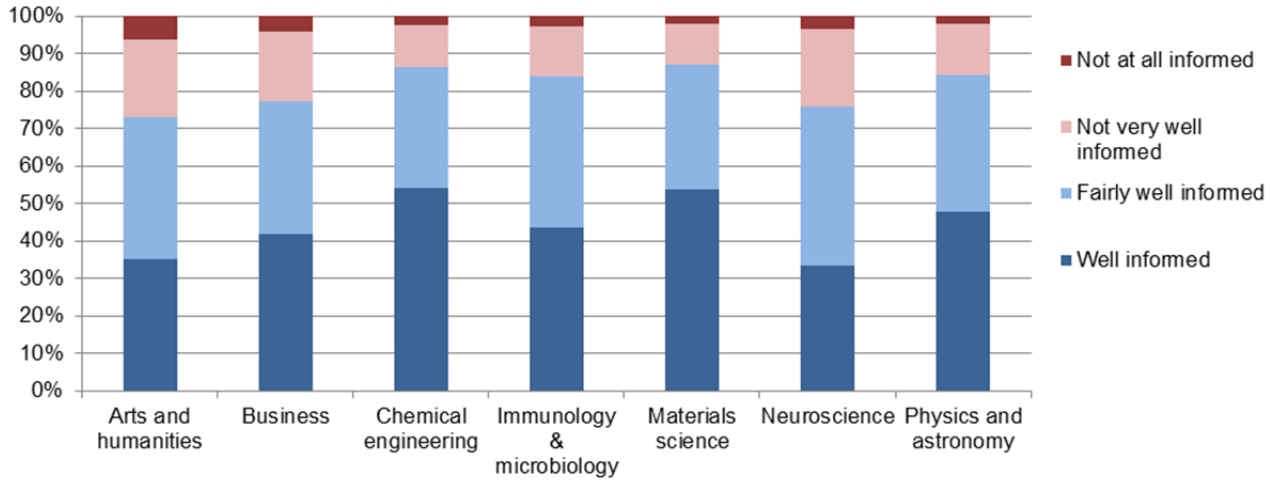
- Whether the article is available *from the publisher* on an open and free of charge basis (Q48).

- Whether a *version* of the document is available on a *freely accessible and searchable repository* (Q49).
- Whether the document was or is *still subject to temporary embargo* and the length of this period (Q50 and Q51)
- Whether an *early version of the paper* is available in an open repository (Q27).

Information on the OA status of the journal in which the document was published was obtained from the source title Scopus files.

Additional information was collected on how well respondents feel informed about author licensing terms. As **Figure 2** shows, there is some considerable degree of uncertainty among authors, with nearly 20% not feeling well informed. Actual responses to specific questions reveal an even larger degree of uncertainty.

Figure 2. Authors' perception about information on the conditions in the author licence agreement

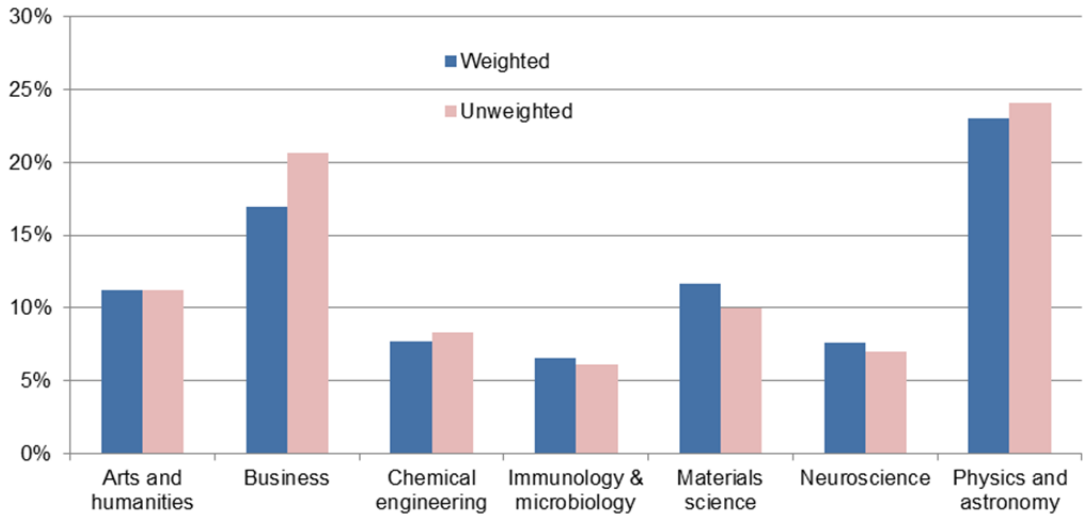


Source: OECD Pilot Survey of Scientific Authors 2014/15.

Logistic regression analysis shows that, after controlling for field and country of affiliation, female authors report feeling significantly less well informed – 21% less informed (s.e.=8%) – and also younger authors and those with fewer citations of their work. There is no systematic difference for authors affiliated to in higher education institutions in comparison to authors with other affiliations.

The release of documents before going through peer review is an important instrument for enhancing access to research outputs. This practice seems to be particularly common in the domains of Physics and astronomy and Business, with more than 20% of authors reporting having placed early versions in public open repositories (**Figure 3**). The prominent place occupied by physics and business can be associated to the widespread role of repositories such as ArXiv and RePEC, respectively.

Figure 3. Documents with an early version of paper available in a public repository



Source: OECD Pilot Survey of Scientific Authors 2014/15. Weighted means results adjusted for non-response.

The combined representation in **Table 4** of responses collected for Q48, Q49 and OA journal status indicates that between 50 to 55% of documents in the selected fields can be considered to be available on an OA basis in early 2015. Combining the information on journal status and affirmative responses about access from publishers, it appears that a significant proportion of documents have open access status from publishers beyond what can be expected from the information available about the status of the journal.

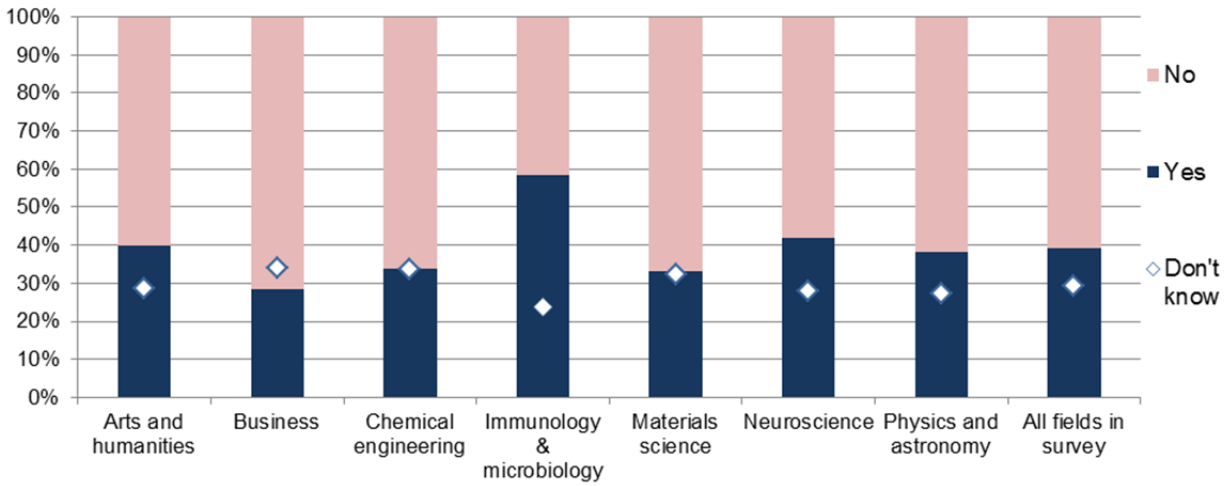
Table 4. Access patterns among survey respondents
Un-weighted responses

Journal status (2014)	Q48. Access from publisher (self-reported)	Q49. Access from repository (self -reported)	Percentage	Observations on likely open access (OA) status
Not in DOAJ	Not known	In open repository	6%	Green type access
Not in DOAJ	No	In open repository	15%	Green type access
Not in DOAJ	Yes	In open repository	7%	Hybrid/green access
Not in DOAJ	Not known	Not in open repository	18%	Most likely not OA
Not in DOAJ	No	Not in open repository	29%	Not open access
Not in DOAJ	Yes	Not in open repository	11%	Hybrid or post embargo
In DOAJ	Not known	In open repository	1%	Gold/green access
In DOAJ	No	In open repository	1%	[Possible response error]
In DOAJ	Yes	In open repository	3%	Gold/green access
In DOAJ	Not known	Not in open repository	2%	Possible gold access
In DOAJ	No	Not in open repository	2%	[Possible response error]
In DOAJ	Yes	Not in open repository	4%	Gold access
Summary				
Document in DOAJ journal			13%	
OA from publisher			26%	
OA from repository			34%	
OA from publisher or repository			49%	
OA: journal, publisher or perspective			53%	

Source: OECD Pilot Survey of Scientific Authors 2014/15.

Reflecting the uncertainty on licensing terms, there is also some considerable uncertainty among authors concerning access status from publishers, around 30% do not know the answer. This pattern is fairly uniform across disciplines (**Figure 4**). Access through publisher is likely to vary over time and it may not be easy to assess from the perspective of authors with institutional online access who do not directly encounter access problems themselves. Weighted estimates by field indicate that among those who report to know the access status, rates of publisher OA are largest among papers in Immunology and microbiology, at over 50%, while they range between 30 and 40% for the other fields covered in the survey, with business and administration being the least open of all. It is important to note that this does not necessarily represent OA status at the time of publishing, which is the formal definition of gold OA.

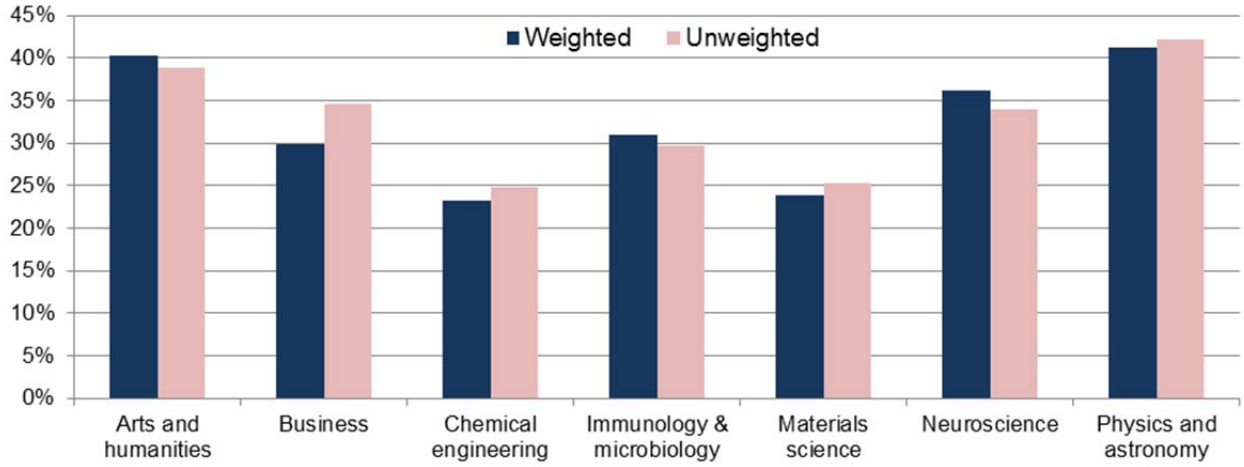
Figure 4. Document available from publisher on open access basis, as self-reported by author (Q48)



Source: OECD Pilot Survey of Scientific Authors 2014/15.

Regarding access from repositories, we find evidence of significant access rates that also vary across fields, reaching up to nearly 40% in Physics and astronomy and also in Arts and humanities (**Figure 5**). The lowest repository-based access rates are found in Chemical Engineering and in Materials Science.

Figure 5. Document available in an open and free of charge repository (Q49)

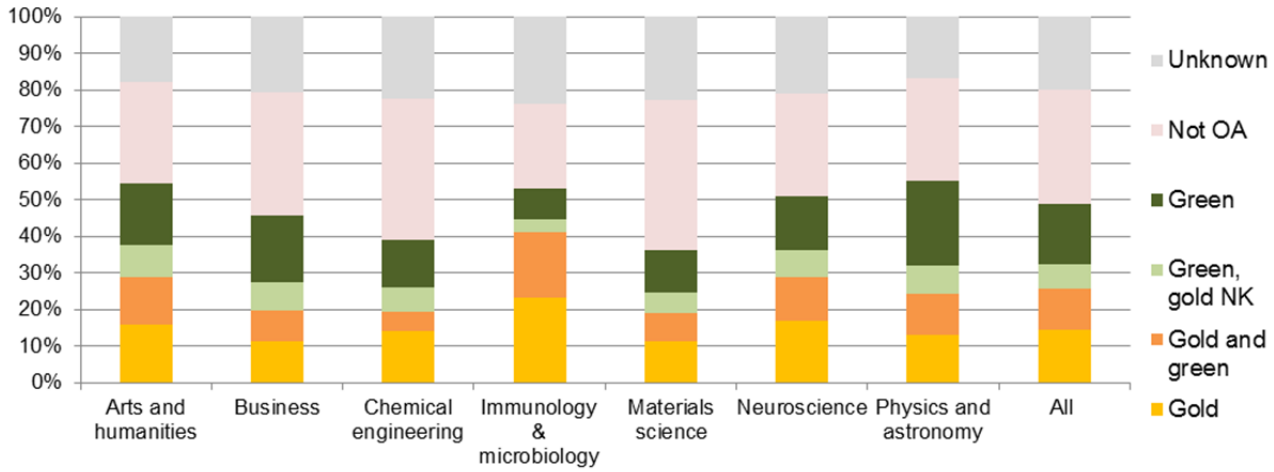


Source: OECD Pilot Survey of Scientific Authors 2014/15.

Because of the overlap between the two access categories reported in Q48 (publisher) and Q49 (repository), we combine them and attempt to map onto tentative gold and green access categories (**Figure 6**). These results point to the fields of Materials Science and Chemical Engineering as the least open, while the fields of Arts and humanities, Business and Physics and astronomy gain a major accessibility boost from the addition of green access documents to those with publisher-based OA.

Figure 6. Open access status, by field, based on author's self-reported information

Weighted results, by field

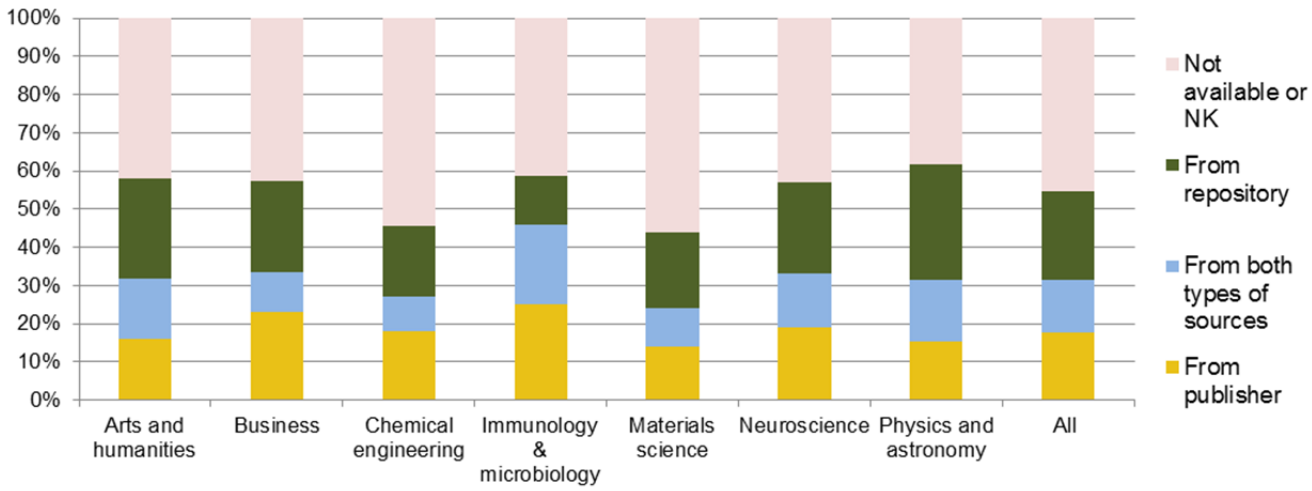


Source: OECD Pilot Survey of Scientific Authors 2014/15. NK means not known.

Finally, we use additional information on the OA status of the journal as well as information on the online availability of early versions to reduce the extent of unknown access status. The results are available in **Figure 7**, which present a very similar qualitative picture. These results again indicate an average OA rate of nearly 55%, just a small majority of documents, with repositories playing a very important role for at least 20% of all documents.

Figure 7. Document access status, combined information

Weighted results, by field

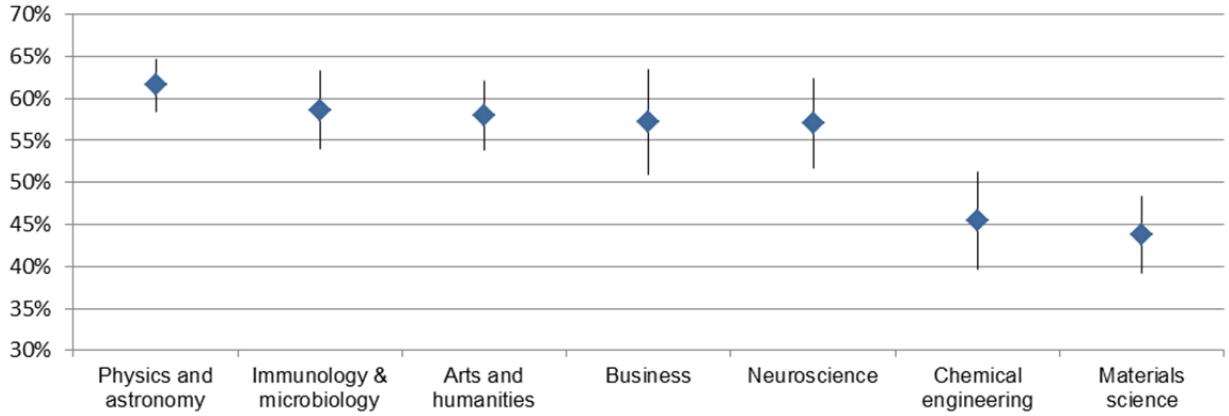


Source: OECD Pilot Survey of Scientific Authors 2014/15.

In order to help assess the precision of these estimates, open access rates are presented in **Figure 8** with their estimated 95% confidence intervals. The main message from our results is that open access is significantly lower in the Chemical engineering and Materials science domains, relative to the other fields considered. The Arts and the humanities reach comparable access rates to their natural science counterparts of Physics, Immunology and Neuroscience thanks to the use of repositories.

Figure 8. Document open access status, by field

Weighted results, average and 95% confidence intervals



Source: OECD Pilot Survey of Scientific Authors 2014/15.

To put these results in a broader context, OECD (2015a) provides an overview of recent literature documenting access status to scholarly journal articles for potential comparison. Laakso and Björk (2012) point that about 17% of scientific articles published in 2011 and indexed in Scopus were openly available through journal publishers (i.e. gold OA). Most articles were immediately available (12%) whereas the remaining 5% were available 12 months after publication. Hybrid open access articles[11] accounted for 0.7% of the total published articles in 2011. Our publisher OA estimate of 30% is somewhat higher, which may be explained by a different discipline composition and/or the time lag after which the document is made available by publishers. We turn to this point below.

It is also possible to benchmark green access estimates but this is somewhat more complicated, as researchers archive articles not only on official repositories but also on personal webpages or on other digital infrastructures. As reported in OECD (2015a), a number of computerised internet search and comparison-based estimates have been developed. More conservative estimates such as Björk et al. 2013 suggested 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. Gargouri et al. (2012) suggest that the share of articles available through green open access was slightly above 20% as of 2011. Archambault et al. (2014) have found that as of spring 2014 more than 50% of the scientific papers published between 2007 and 2012 can be accessed and downloaded for free online. This particular estimate is very close to ours. Both our estimates and Archambault et al.'s are upper bound estimates of the actual access status at the time of publication because of the effect of embargo policies.

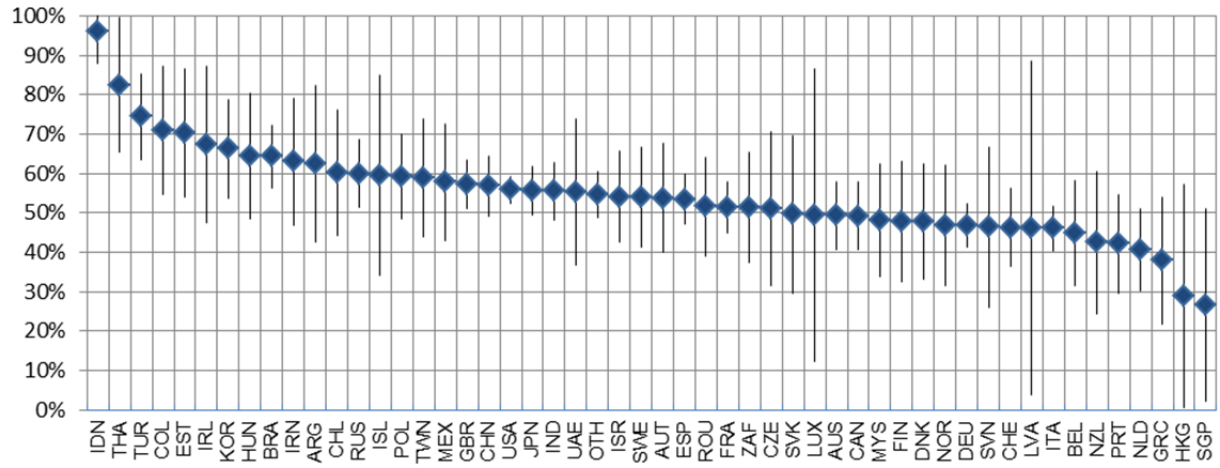
Our results are also broadly consistent with findings in the literature concerning access variation by discipline and the relative importance of green versus gold access. According to a number of studies (Björk et al., 2010, UNESCO 2012), the uptake of green open access is high in Physics and astronomy and Social Sciences relative to gold access, whereas it is low in chemistry and chemical engineering as well as in biology and related fields.

Open access by country

The adoption of open access publishing within country or economy of institutional affiliation (**Figure 9**) can be more precisely estimated for larger economies while estimates for smaller economies are of lower quality. In general, conditional on being captured in this population, authors based in developing economies tend to have higher OA rates. This is largely driven by access status from the publisher, revealing systematic differences in the journals where authors have their papers published.

Figure 9. Open access (all types) status, by country/economy of affiliation

Weighted results, average and 95% confidence intervals

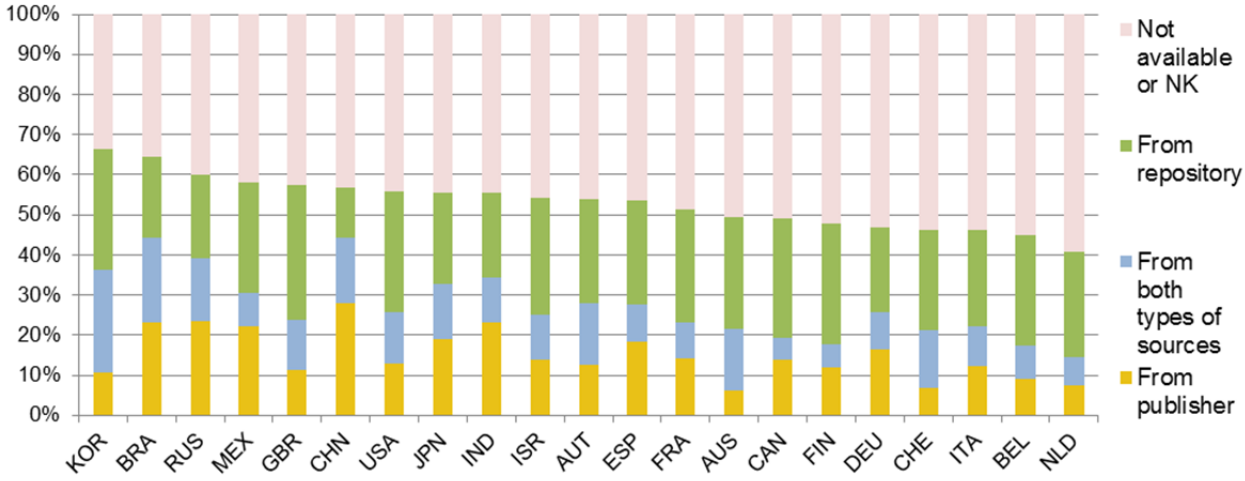


Source: OECD Pilot Survey of Scientific Authors 2014/15.

The relative importance of publisher or repository-based access varies significantly across countries. **Figure 10** reports the distribution of access modes for selected, larger economies where estimates are more precise. It is worth noting in particular the strong relative importance of green access for authors based in the Netherlands, Switzerland, United States and the United Kingdom. Green OA appears to be a more popular choice of OA mode within the more developed countries.

Figure 10. Open access status, by type and country of affiliation

Selected, larger economies

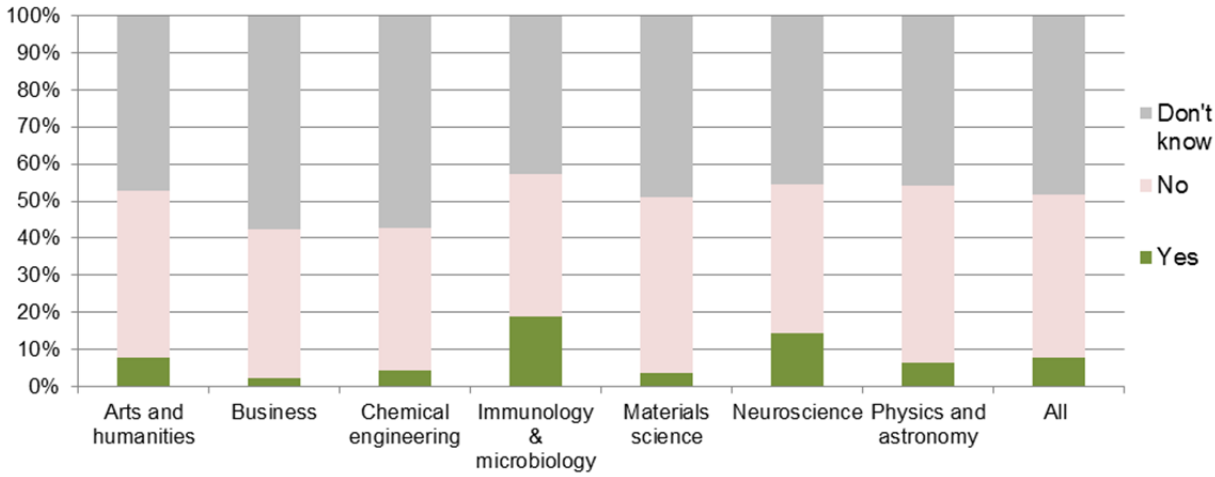


Source: OECD Pilot Survey of Scientific Authors 2014/15. NK= not known.

The use of time-defined embargo periods for documents is reported in **Figure 11**. It is possible to note a very substantial degree of uncertainty among authors concerning such practices, over and above the reported uncertainty concerning their licensing agreement. The practice of open release subject to embargo is not that common and can be approximately 20% assuming that unknowns are not significantly different from the rest of the population. This practice is far more diffused in the life and health science domains within this survey, which may also reflect the NIH funding conditions previously alluded to.

Figure 11. Document subject to embargo period

Weighted estimates, by field

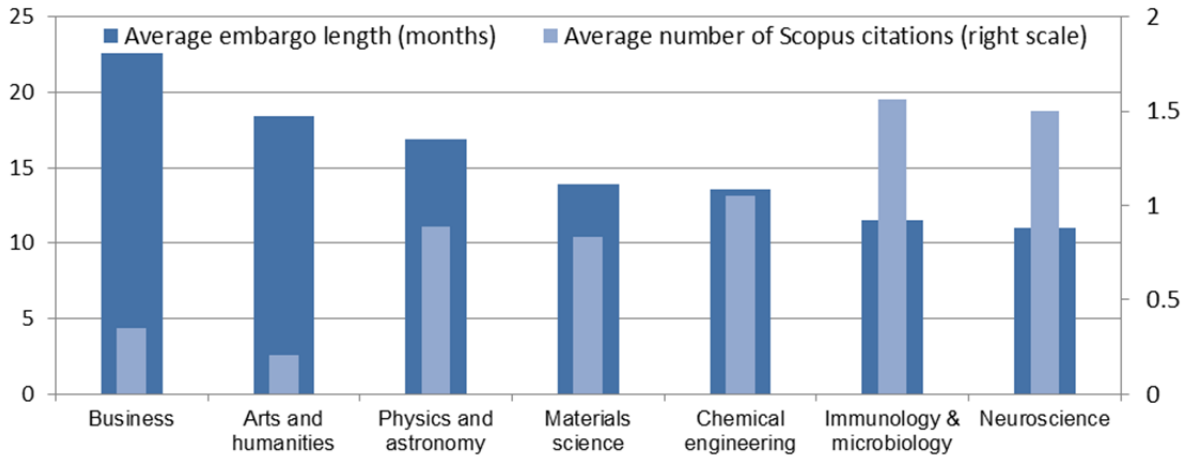


Source: OECD Pilot Survey of Scientific Authors 2014/15.

There is considerable variation in the length of embargo periods in different disciplines, as also indicated in the literature (Laakso, 2014). The embargo length appears to be related to the citation patterns by discipline: longer embargo spells are related to lower (and possibly deferred) citations (**Figure 12**). It is however not possible to tell whether there is a causal effect that connects both and the direction of such an effect, a relevant policy consideration when setting guidelines on embargo periods for publicly-funded research. Publishers may fix their embargo dates to reflect use patterns, while the latter may in turn depend on accessibility.

Figure 12. Average embargo length and citation patterns by field

Average embargo length defined for documents subject to embargo period



Source: OECD Pilot Survey of Scientific Authors 2014/15.

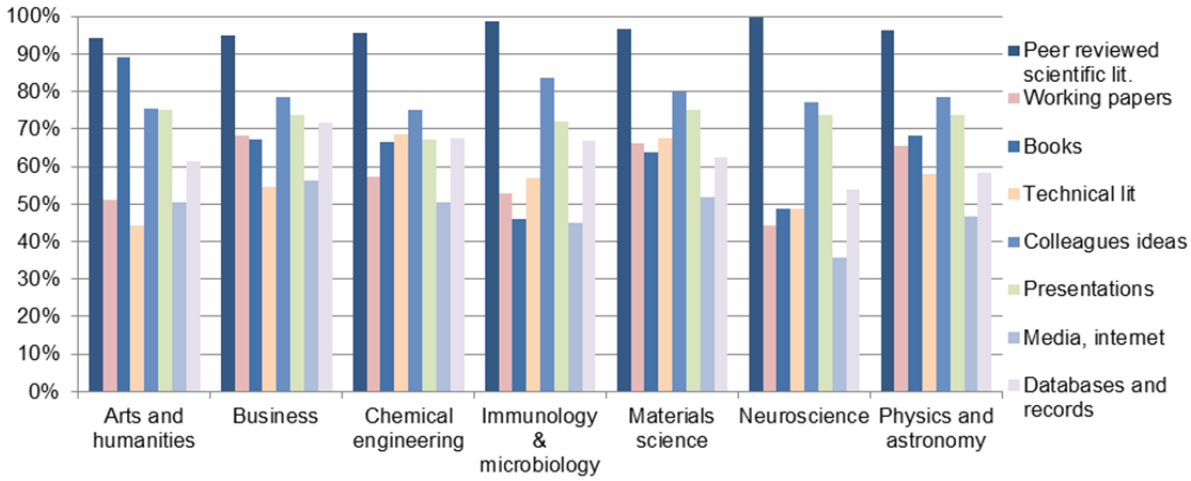
While there is a significant negative correlation at the article level between the embargo length and the number of registered Scopus citations by 2014, our analysis shows that this correlation greatly diminishes after controlling for unobserved field characteristics, and vanishes completely once the journal impact indicator (SNIP), i.e. a measure of journal prestige, is taken into account. On the basis of these results, it appears that further work is needed in this area if the use of embargos as part of open access policies is to become more widespread. There is a risk that single embargo reference policies across all fields may be too blunt of a policy instrument.

3.2. The research context for open access

Problems experienced by the authors accessing sources

There is some evidence that both research and innovation system actors may experience difficulties in accessing non-openly available scientific material. OECD (2015a) reports on access difficulties within the academic community in the United States and Europe. This survey shows that access to peer-reviewed scientific results matters. For more than 90% of respondents, peer reviewed scientific literature is an important source for their research work (**Figure 13**). The importance assigned to other areas varies greatly by field. For example, in the arts, books were the second most cited important sources, while in other fields that position was allocated to ideas from colleagues and supervisors. Conference presentations are the third most cited important source of knowledge. Technical literature including patents and standards had a prominent role in chemical engineering and materials science. Databases and records were cited as important in more than 50% of cases, reaching more than 70% in the case of Business and finance. This field ranks media and the internet more frequently than any other, whereas in Neuroscience media an internet are of the lowest importance. Working papers play an important role in Business, but also in Materials science and in Physics and astronomy.

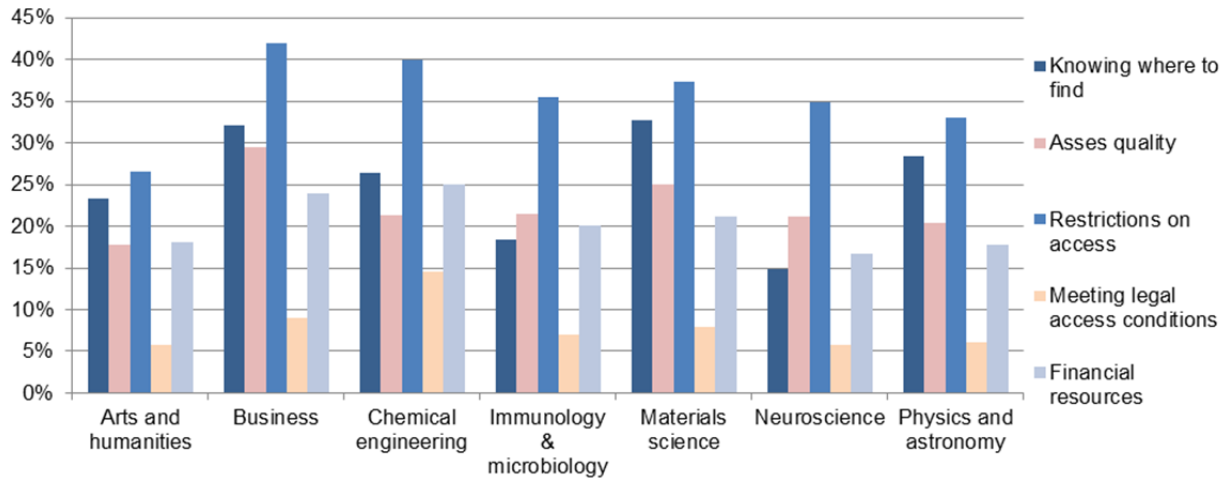
Figure 13. Important sources of knowledge, by field



Source: OECD Pilot Survey of Scientific Authors 2014/15.

Authors participating in the survey were also asked to report the major obstacles faced in using external sources of information and data in their research work (**Figure 14**). Access restrictions were the most recurrent challenge, followed by lack of information on where to find the sources, confirming the importance of ensuring retrievability of information as part of OA policies.

Figure 14. Obstacles encountered in using external sources in research



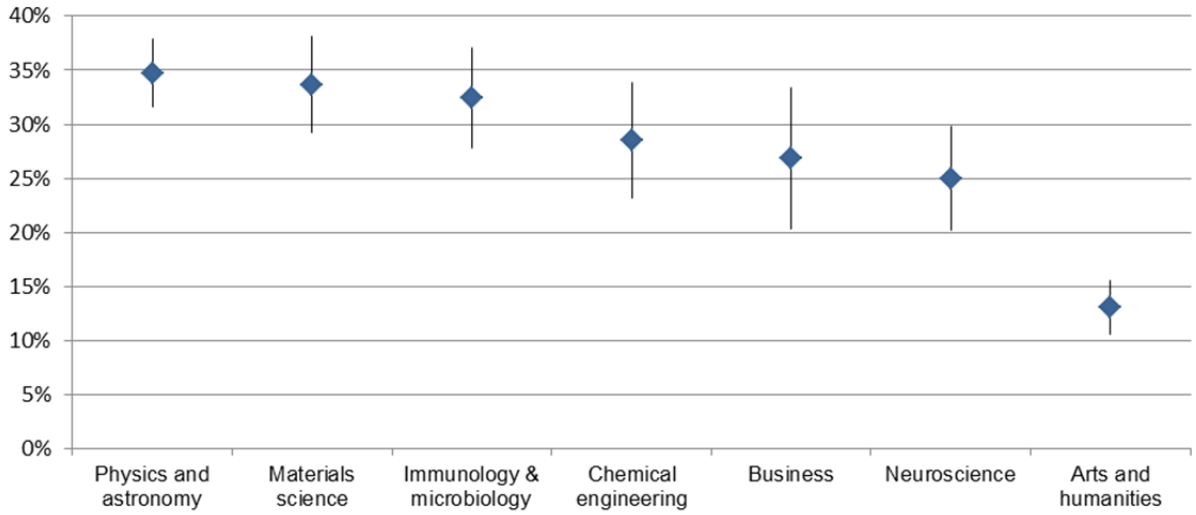
Source: OECD Pilot Survey of Scientific Authors 2014/15.

Resources and requirements for publication and dissemination

Funding streams available for research in different domains appear to place different emphasis on supporting dissemination and publication (**Figure 15**). In the arts and humanities, less than 15% of authors indicate receiving funding for that specific purpose. Support for dissemination does not appear to be automatically factored into financial support for research projects.

Figure 15. Resources provided by research funders for dissemination, by subject field

Percentage of authors receiving funding for dissemination, and 95% confidence internals

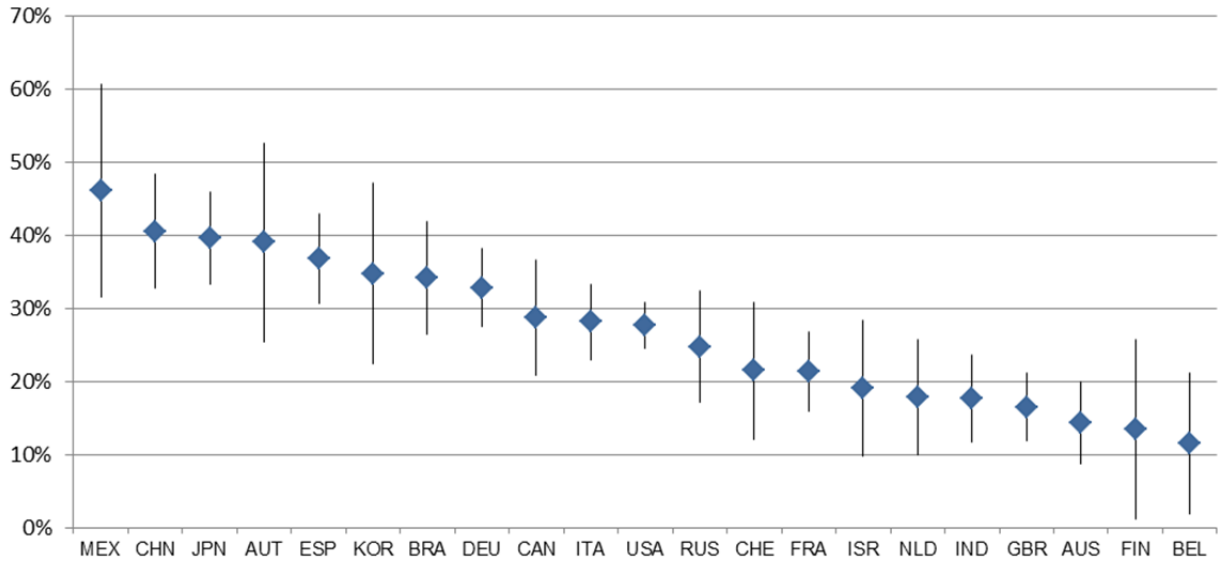


Source: OECD Pilot Survey of Scientific Authors 2014/15.

There appears to be significant variation across countries of affiliation in terms of whether funding organisations provide dedicated resources for publication and dissemination purposes. Such support does not necessarily entail resources for open access funding, let alone specific national programmes with that objective. The results for the largest countries are available in **Figure 16.** In comparing these results, it is also important to note that within each country, the group of corresponding authors in Scopus relative to all scientists will greatly differ, with different biases towards seniority, connections, etc...

Figure 16. Resources provided by research funders for dissemination, by country

Percentage of authors receiving funding for dissemination, and 95% confidence intervals

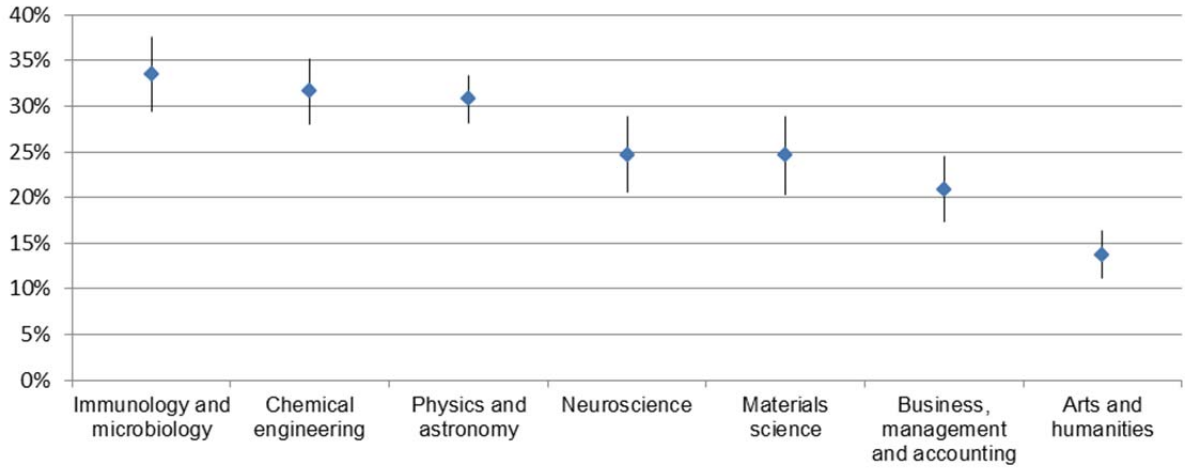


Source: OECD Pilot Survey of Scientific Authors 2014/15.

Authors reported that only a minority of research funders/sponsors had established in advance rules or conditions for the dissemination of research results (**Figure 17**). The highest rates were found in Immunology & microbiology, Neuroscience and Materials science. In addition to other idiosyncratic features, one possible driver is the requirements set out by a major funder in the life sciences domain, the US National Institutes of Health (NIH), which recently implemented a mandatory public access policy for publications resulting from its funded research (OECD, 2015a).

Figure 17. Did funders establish any rules or conditions for dissemination of research results?

Percentage and 95% confidence intervals



Source: OECD Pilot Survey of Scientific Authors 2014/15.

3.3. Open access and citation impact

The academic literature has long debated whether open access publications receive more citations than non-open access publications by attempting to estimate the so-called *open access citation advantage*. OECD (2015a) concludes that a majority of studies show that open access is associated with a larger citation impact, while there is no general consensus on the intensity in the increase in citations. The causal nature of these effects is also highly controversial (McCabe and Snyder, 2014). The empirical challenge in demonstrating the true impact of open access is the double selection process that makes the occurrence of OA a non-random event driven by choices and attributes directly related to impact and impact expectations. Bias can arise as a result of a journal (OA or not) decision on whether an article meets its own standards, while the author previously decides whether a particular journal is the optimal submission choice (on basis of its prestige, openness, etc…) given the uncertainty it may or may not be accepted. Given the absence of readily available and connected access "experiments", this study can only report correlations that control for a number of characteristics.

Table 5 presents the estimated elasticity of the number of citations captured in the Scopus database with respect to a number of author and document characteristics, controlling for country and field strata fixed effects. Journal impact is a key factor in accounting for observed citations. Publisher-based OA has a negative but not significant effect, suggesting a limited citation impact through this channel, whereas OA through repositories has a statistically significant positive association with citations.[12] Measures of author quality and connections, as implied by whether they are reviewers or editors are also positively associated with citations. There are no visible links with gender, age or Higher Education (HE) affiliation.

Table 5. Explaining document citations captured in Scopus, 2014

Least square estimates of log (1+citations) with country-field strata fixed effects

	Coefficient	Std. Err.	P-value
Journal impact (SNIP)	**0.167**	0.010	0.000
Missing journal impact	-0.054	0.030	0.078
OA publisher	-0.027	0.019	0.160
OA other repository	**0.045**	0.017	0.009
Author's reviewer status			
Remunerated and non-remunerated	**0.087**	0.043	0.043
Non remunerated	**0.101**	0.029	0.001
Remunerated	**0.117**	0.043	0.006
Author's editor status			
Remunerated and non-remunerated	**0.125**	0.058	0.031
Non remunerated	**0.065**	0.019	0.000
Remunerated	0.050	0.041	0.228
Female	0.027	0.020	0.180
Age	-0.001	0.001	0.396
HE affiliation	0.019	0.019	0.311
Constant	0.074	0.048	0.120
Number of observations		5090	
F(13, 4744)		26.06	
Prob > F		0.000	
R-squared		0.21	
Adjusted R-squared		0.15	

Source: OECD Pilot Survey of Scientific Authors 2014/15.

It is often noted that scientists and academics are not the only groups that can benefit in principle from greater open science efforts and access to results. We therefore compared the factors that are associated with citations in peer reviewed journals contained in major indices with those revealed by looking at other forms of possible citations, using the information provided by the authors themselves (**Table 6**).

Table 6. The role of open access in accounting for self-reported citations, by source of citation

Weighted estimates of citation probability and T-test for difference

	Not Open Access	Open access	Difference OA vs non OA (T-stat and sig-levels)	
Peer-reviewed literature - major index	0.81	0.81	0.30	
Peer reviewed literature - other	0.28	0.31	1.67	*
Working papers	0.12	0.18	4.40	***
Technical documents	0.07	0.08	1.63	*
Policy/government documents	0.05	0.07	2.05	**
Practitioner's protocols	0.02	0.02	0.00	
Media	0.06	0.07	1.88	*
Legal proceedings	0.02	0.03	2.21	**
Other	0.06	0.05	-0.79	

Source: OECD Pilot Survey of Scientific Authors 2014/15.

Note: A positive T value denotes a higher estimated level of citations for OA. Significance levels: *=10%, **=5%, ***=1%.

The results of the analysis of citation advantage across different referencing vehicles, which controls for the type of OA and other factors, are reported in **Table 7**. They show a number of highly relevant findings:

- There is substantial consistency between estimates derived from actual citations captured in Scopus and those self-reported by authors.
- Journal "prestige" citation impact matters the most. The analysis also confirms the absence of citation advantage from publisher-based open access, and the existence of such an advantage from repositories.
- When it comes to other peer-reviewed literature, not captured in major indices, journal prestige and reviewer status become notably less important. Here we find some evidence of small positive citation advantage from publisher open access. Within this class of citations, older authors tend to be less cited.
- In the case of citation within working papers, open access through repositories is the most distinctive access feature associated to citation, which is probably related to the speed with which results become available and the citing document can be released. Older authors also tend to be less cited in working papers.
- Regarding technical documents such as patents or standards, the impact of repository access on citation is positive although the result is not estimated precisely enough to be significant. Female, older authors in higher education institutions are less likely to be cited in this type of literature.
- Open access from the publisher appears to matter for citation in practitioners' protocols, for example in medical guidelines. Higher education (HE)-based authors are less likely to be cited.
- Both types of open access matter from the perspective of coverage in the media, which also appears to increase with journal prestige. Women, older and HE-affiliated scientists are less likely to report their work being cited in the media.
- We find some small evidence of openness on citation in legal proceedings. For this type of fora, there is a strong "expert" effect suggested by the fact that editors are more likely to have their work cited.

Table 7. Probability of document being cited, as self-reported by author, by source

Least square estimates (linear probability) with country-field strata fixed effects

	Peer-reviewed literature - major index		Peer reviewed literature - other		Working papers		Technical documents		Policy/ government documents		Practitioner's protocols		Media		Legal proceedings		Other	
	Coeff	T-stat	Coeff	T-stat	Coeff	T-stat	Coeff	T-stat	Coeff	T-stat	Coeff	T-stat	Coeff	T-stat	Coeff	T-stat	Coeff	T-stat
Journal impact	0.07	9.79	0.01	1.66	0.00	-0.07	0.00	1.05	0.01	1.20	0.00	1.04	0.02	3.79	0.00	-0.74	-0.01	-2.15
Missing journal impact	-0.13	-6.70	0.02	0.64	-0.01	-0.56	0.05	3.42	0.00	-0.08	0.01	0.84	0.00	-0.15	0.01	1.15	0.03	1.99
OA publisher	-0.03	-2.33	0.03	1.87	0.01	0.48	0.00	-0.54	0.02	2.20	0.01	1.94	0.02	1.89	0.01	1.29	0.00	0.29
OA other repository	0.05	4.13	0.01	0.74	0.05	4.59	0.01	1.32	0.01	1.29	0.00	0.42	0.02	3.03	0.01	1.15	0.00	0.36
Reviewer status																		
Remunerated and not	0.16	5.85	0.04	1.03	0.05	1.79	0.03	1.33	0.03	1.55	0.00	0.17	0.00	0.22	0.02	1.37	-0.02	-0.95
Non remunerated	0.13	7.09	0.03	1.10	0.00	-0.01	-0.01	-1.01	0.00	0.13	0.00	-0.57	0.00	0.11	0.00	0.47	-0.01	-0.43
Remunerated	0.17	6.07	0.01	0.29	0.00	0.10	-0.01	-0.74	-0.02	-0.93	0.01	1.35	-0.02	-1.03	-0.01	-0.59	-0.05	-2.78
Editor status																		
Remunerated and not	0.08	2.17	0.14	2.90	0.10	2.72	0.05	1.97	0.04	1.52	0.03	2.29	0.06	2.45	0.07	4.40	0.03	1.17
Non remunerated	0.05	3.97	0.08	4.88	0.02	2.03	0.02	2.57	0.02	2.51	0.01	2.22	0.02	2.24	0.01	2.33	-0.01	-0.77
Remunerated	0.04	1.54	0.03	0.83	0.01	0.21	0.00	-0.08	-0.01	-0.54	-0.01	-1.57	0.00	-0.24	0.01	1.21	0.02	1.31
Female	-0.01	-0.97	0.00	-0.30	-0.01	-1.19	-0.02	-2.07	0.00	-0.30	0.00	-0.04	-0.02	-1.71	0.01	1.77	0.02	1.91
Age	0.00	-1.39	0.00	-4.23	0.00	-4.72	0.00	-2.93	0.00	0.09	0.00	-0.30	0.00	-4.58	0.00	0.76	0.00	5.14
HE affiliation	0.02	1.62	0.00	-0.05	-0.01	-0.44	-0.04	-4.93	-0.05	-6.56	-0.01	-3.50	-0.02	-2.38	0.00	0.58	-0.01	-1.06
Constant	0.62	20.05	0.37	9.35	0.24	7.90	0.13	6.30	0.07	3.53	0.02	1.92	0.12	5.44	0.00	0.00	0.00	0.05
Number of observations	5090		5090		5090		5090		5090		5090		5090		5090		5090	
F(13, 4744)	24.46		4.3		5.04		5.3		5.29		2.69		5.46		3.35		4.07	
Prob > F	0		0		0		0		0		0.0009		0		0		0	
R-squared	0.19		0.11		0.14		0.09		0.08		0.10		0.09		0.09		0.09	
Adj R-squared	0.13		0.04		0.07		0.02		0.02		0.03		0.03		0.03		0.02	

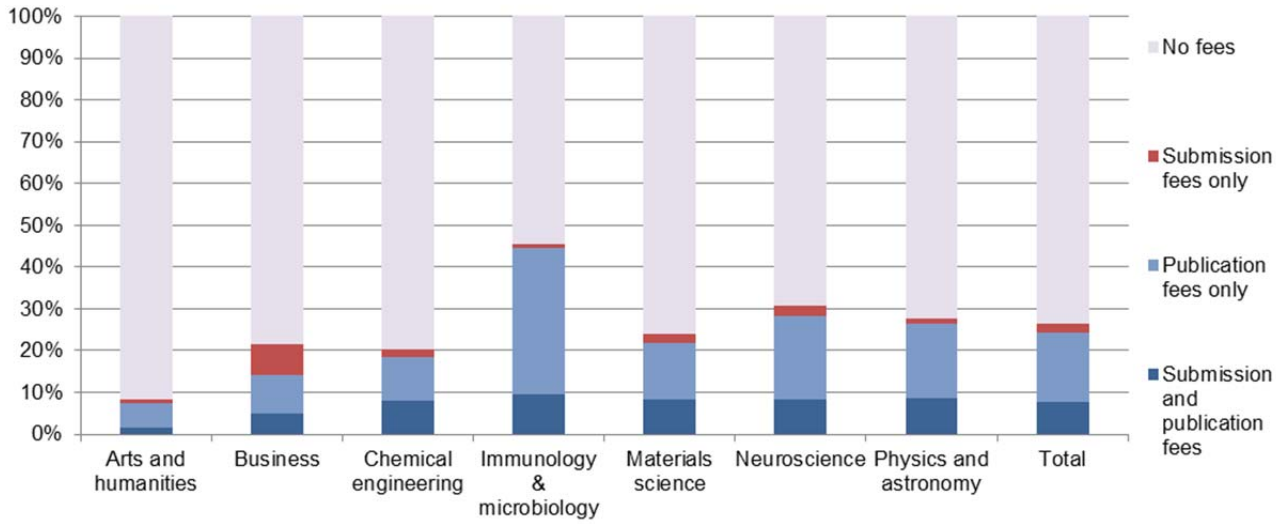
Source: OECD Pilot Survey of Scientific Authors 2014/15.

In conclusion, it appears that different open access strategies play a variety of roles in connection to promoting citation recognition in multiple vehicles, relating to types of uses and "user" communities. Overall, open access is principally associated with increased citation-use outside peer reviewed sources. This is indicative of a decoupling between dissemination and academic quality assurance functions, which is likely related to the relevance of the information conveyed with respect to its uses.

3.4. The marketplace for dissemination and quality assurance - publishing charges

Analysis of survey responses (**Figure 18**) indicates that a majority of authors report not having paid any article processing or publication charges. The incidence of fee payment is largest for articles in Immunology and microbiology journals (close to 45%) and lowest for arts and humanities at less than 10%. The practice of charging submission fees, regardless of publication (i.e. implicitly associated to the review costs), is most prevalent for business journals, at less than 10% of all articles.

Figure 18. Extent of publication and dissemination fee payment



Source: OECD Pilot Survey of Scientific Authors 2014/15.

The survey finds that for approximately 60% of fee paying authors, the payment does not ensure the publisher giving open and free of charge access to the document. The remaining 40% thus suggests that just below one third of all documents are freely available from the publisher. This was previously estimated at close to 30% of all documents.

The reported incidence of fee paying by authors is analysed in **Table 8**. The logit regression results show that authors of publications in OA journals, as well as those who report that the documents are openly available from the publisher, are more likely to have paid charges. There is not a significant relationship between paying charges and journal impact measures. Authors with HE affiliations and older authors are less likely to pay fees, but no significant differences are found by gender. The study of aspects describing the potential quality of the review process - i.e. various possible contributions by reviewers to the quality of the document - shows no link to publication speed but finds a positive relationship between fees and the number of reviews received. The evidence on the perceived quality of aspects of the review is somewhat inconclusive, potentially reflecting the high correlation across those aspects and other attributes. Capacity to pay seems to be a significant factor in paying charges, as shown by the positive and statistically significant relationship with the availability of resources for dissemination from the research funder as well as the earnings of the author (measured in 2015 and not at the time of publishing).

Table 8. Determinants of the probability of paying author publication/processing charges (APCs)

Conditional logit estimates, with country-field fixed effects (un-weighted)

	1	2	3	4	5	6
Open access journal	1.321	1.097	1.098	1.099	1.097	1.007
	(8.67)**	(7.06)**	(7.09)**	(7.06)**	(6.87)**	(5.57)**
Open access from publisher		0.420	0.424	0.421	0.407	0.334
		(4.15)**	(4.19)**	(4.15)**	(3.92)**	(2.89)**
Journal impact (SNIP)	0.071	0.097	0.103	0.097	0.084	0.076
	(1.12)	(1.43)	(1.55)	(1.42)	(1.27)	(1.12)
Missing impact information	-0.357	-0.311	-0.314	-0.305	-0.278	-0.243
	(1.79)	(1.57)	(1.61)	(1.55)	(1.37)	(1.09)
Higher education	-0.384	-0.411	-0.423	-0.417	-0.388	-0.356
	(4.20)**	(4.58)**	(4.73)**	(4.64)**	(4.23)**	(3.76)**
Female	-0.092	-0.109	-0.112	-0.108	-0.122	-0.080
	(0.86)	(0.95)	(0.97)	(0.94)	(1.03)	(0.62)
Age	-0.062	-0.051	-0.051	-0.051	-0.051	-0.074
	(2.96)**	(2.27)*	(2.28)*	(2.30)*	(2.16)*	(3.00)**
Age Squared	0.044	0.032	0.034	0.034	0.032	0.048
	(2.26)*	(1.53)	(1.59)	(1.60)	(1.41)	(2.05)*
Time to publication (months)		-0.000	-0.001	-0.001	0.001	-0.003
		(0.04)	(0.11)	(0.07)	(0.12)	(0.20)
Number of referee reviews		0.084	0.080	0.078	0.071	0.052
		(2.09)*	(2.01)*	(1.94)	(1.67)	(1.13)
Review assessment						
Thorough and prof. assessment			-0.245			
			(1.99)*			
Identify major issues			0.058			
			(0.50)			
Identify minor mistakes			0.146			
			(1.73)			
Improved scientific content			0.036			
			(0.44)			
Improved presentation			0.051			
			(0.53)			
Contributed suggestions			0.086			
			(0.85)			
Overall review score				0.040	0.025	0.009
				(1.61)	(0.98)	(0.30)
Resources provided by funder for publication					0.961	0.949
					(11.17)**	(9.45)**
Earnings (2015)						0.003
						(2.32)*
Number of observations	4 531	4 221	4 221	4 221	4 221	3 451

Source: OECD Pilot Survey of Scientific Authors 2014/15.

Information on the distribution of fees paid by respondents across disciplines is available in **Table 9**. Conditional on fees having been paid (non-payers are excluded), fees are found to be largest in the life sciences fields of Neuroscience and Immunology, exceeding one thousand dollars on average, followed by Physics and astronomy. Nearly 80% of the variation in reported fees is accounted for by variation across journals, while the remaining is variation within journals which may be due to reporting error and/or pricing practices based upon attributes such as numbers of pages, reviews or other features such as embargo length. Most fees are paid from institutional funds (50%) and project related funds (directly paid by funder or through the project's dedicated funds - nearly 50%) (**Figure 19**). The use of personal funds is very rare as it could be expected.

Table 9. Submission and publication fees, by field

Reported fee values in USD

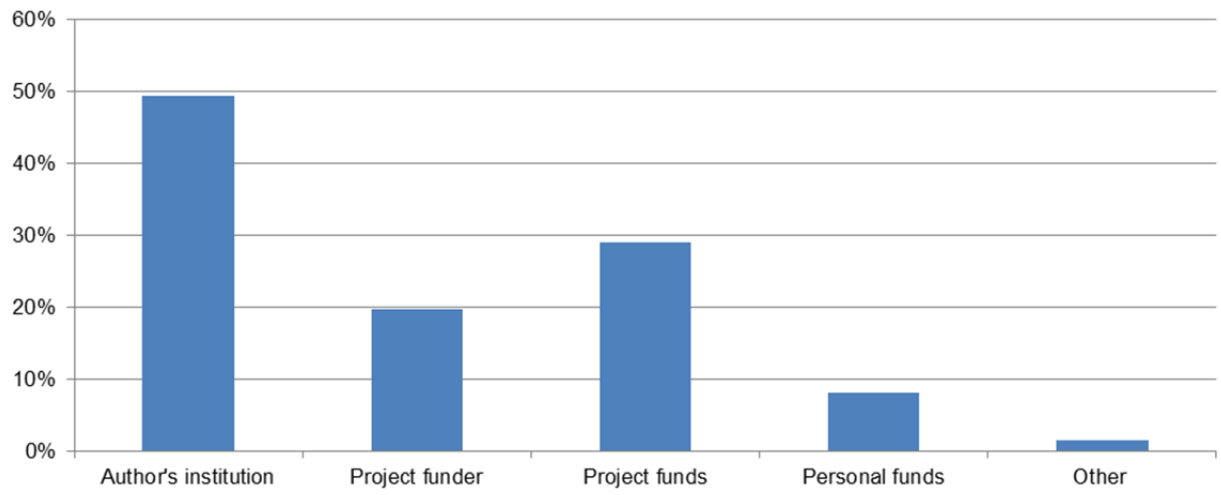
Field	Un-weighted				Weighted			
	1st quartile	Median	Mean	3rd quartile	1st quartile	Median	Mean	3rd quartile
Arts and humanities	130	300	**528**	586	100	300	**511**	716
Business	200	417	**451**	558	309	350	**434**	500
Chemical engineering	235	625	**812**	1077	139	500	**597**	800
Immunology & microbiology	500	973	**1112**	1500	400	834	**1033**	1500
Materials science	300	450	**645**	900	309	400	**581**	800
Neuroscience	975	1156	**1456**	1705	1000	1253	**1496**	1730
Physics and astronomy	417	764	**1020**	1390	309	695	**881**	1200
Total	350	700	**953**	1361	309	600	**837**	1112

Source: OECD Pilot Survey of Scientific Authors 2014/15.

Conditional on having paid fees, we observe (**Table 10**) that there is a very strong and robust relationship between their value and open access, suggesting that authors may be paying a premium for OA. This is robust to accounting for journal impact, which has a distinct significant effect on observed fees paid. Female authors appear to pay systematically less in fees than their male counterparts. We do not find any robust relationship between fees and perceived review quality measures although there is a hint that the number of reviews could matter, as it did with respect to whether fees were paid or not. There is evidence to suggest that, if an amount is paid, its value is rather insensitive with respect to benefiting from dedicated sources of funding for this purpose or earnings measures.

Figure 19. Sources of funds for submission and publication fees

Weighted results, percentage of fee-paying authors



Source: OECD Pilot Survey of Scientific Authors 2014/15.

Table 10. Determinants of variation in fees paid, conditional on having paid any fees

Least squares regression of fees paid (USD), with strata fixed effects (T-stats within parentheses)

	1	2	3	4	5	6	7
Fees assure open access	305.54	210.41	203.87	223.83	207.78	213.800	242.03
	(3.52)**	(2.20)*	(2.29)*	(2.31)*	(2.18)*	(2.23)*	(2.55)*
Authors reports OA from publisher		215.65	218.43	230.40	215.22	205.385	233.85
		(2.21)*	(2.51)*	(2.33)*	(2.21)*	(2.08)*	(2.54)*
Journal impact (SNIP)	371.81	362.96	361.93	354.66	362.01	361.219	470.15
	(3.87)**	(3.87)**	(6.21)**	(3.75)**	(3.83)**	(3.80)**	(6.17)**
Missing impact information	25.07	-14.20	-17.35	12.00	-9.72	-0.03	-3.684
	(0.16)	(0.09)	(0.11)	(0.07)	(0.06)	(0.00)	(0.02)
Higher education	2.141	-4.77	-2.03	9.068	-9.928	-4.04	-4.54
	(0.03)	(0.06)	(0.02)	(0.11)	(0.13)	(0.05)	(0.05)
Female	-209.82	-197.92	-190.86	-175.66	-192.33	-184.70	-188.88
	(2.42)*	(2.29)*	(1.94)	(2.00)*	(2.21)*	(2.09)*	(1.87)
Age	42.90	37.99	40.34	33.56	38.98	38.41	39.32
	(1.73)	(1.49)	(1.61)	(1.30)	(1.51)	(1.46)	(1.49)
Age squared	-42.50	-37.02	-39.70	-32.43	-37.63	-37.19	-34.12
	(1.81)	(1.53)	(1.63)	(1.32)	(1.54)	(1.50)	(1.34)
Submission publication lag (months)			-4.51				
			(0.47)				
Number of referee reviews				51.74			39.52
				(1.29)			(0.94)
Review assessment							
Thorough and prof. assessment						73.40	
						(0.62)	
Identify major issues						17.83	
						(0.17)	
Identify minor mistakes						66.82	
						(0.79)	
Improved scientific content						23.47	
						(0.22)	
Improved presentation						-15.33	
						(0.16)	
Contributed suggestions						-49.05	
						(0.53)	
Overall review score					16.64		
					(0.53)		
Resources provided by funder for publication							78.00
							(0.95)
Earnings (2015)							-0.88
							(0.75)
N	535	535	521	521	535	535	467
R-Squared	0.50	0.51	0.51	0.51	0.51	0.51	0.53

Source: OECD Pilot Survey of Scientific Authors 2014/15.

Further research can be carried out with the available data to study to what extent the existence of a fee is a major factor in the decision on where to publish, and how this relates to services providing by journals such as access and quality of review. We explore this subject from a different perspective in the next subsection.

3.5. Scientist demand for open access –contingent valuation estimates

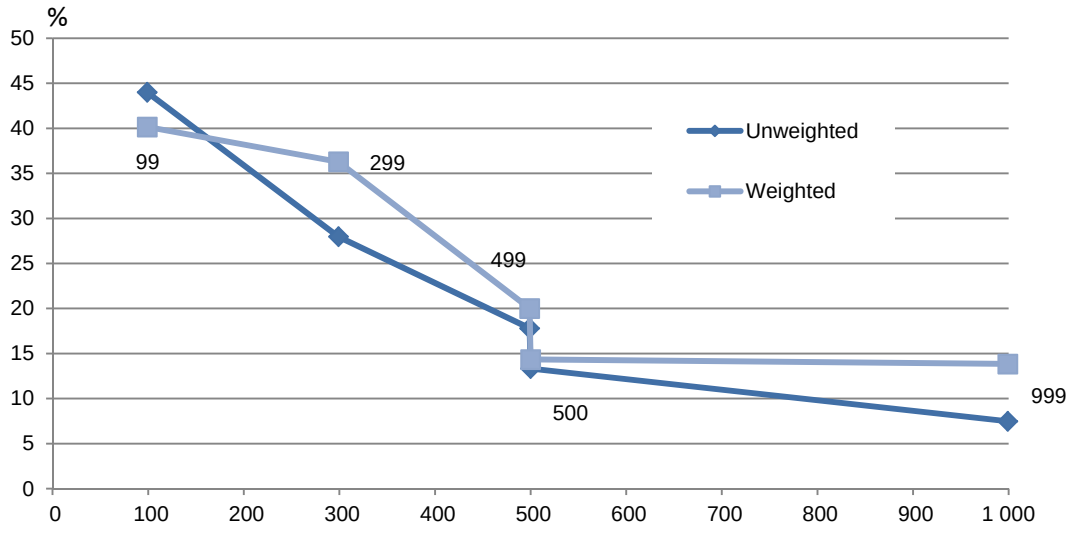
The rising interest in hybrid -on demand- models of open access raises the question of how much authors and the institutions they work for are willing to pay in order to ensure that their publications are available from publishers on an open access basis. Because of the difficulties implicit in studying demand preferences from observed data, we decided to conduct an exploratory contingent valuation exercise. In contingent valuation analysis, respondents are presented with a hypothetical scenario involving the provision of a service such as the cleaning up of a river basin, the provision of local amenities, etc. In this context, the interest is in the service represented by the publisher making the publication openly available,

in return for some form of monetary payment. Contingent valuation methods (CVM) are widely used for the valuation of private, semi private and public goods, as it allows to derive an estimate of how much the respondent values the utility of the scenario for other parties, thus reflecting potential altruism towards others. [13] CVM exercises can also involve a number of potential biases, particularly if there are strategic, warm glow, or related effects that lead respondents to reply very differently compared to how they would behave in the real world if actually presented with that scenario. Instead of asking respondents to say how much they would be willing to pay for open access, we decided to randomly propose price tags for this "service" to facilitate a prompt, accurate response, rather than asking respondents to tentatively pull some estimate out of thin air.

We presented a pay-more for openness scenario to those individuals who had paid a publication fee but that fee did not warrant open access from the publisher (Q40). We were thus interested in identifying the willingness to pay for openness over and above any publication services. The weighted and un-weighted results are available in **Figure 20**. As it can be noted, even for small amounts in the order of USD 99, the majority of respondents are not willing to pay extra for open access. We do observe a sensitivity to price increases, leading to a reduction in willingness to pay (WTP) as the proposed amount increases up to USD 500. From then on, on average, WTP is in the order of 5 to 15%, indicating a high price-elasticity. We find strong evidence of marked discontinuity in the responses to offer prices, as individuals offered to pay USD 499 are significantly more willing to pay than those offered only one more USD.

Figure 20. Willingness to pay randomly assigned prices for open access, fee-paying respondents

Percentage, by amount proposed (USD)



Source: OECD Pilot Survey of Scientific Authors 2014/15.

These findings are corroborated by econometric analysis, controlling for other characteristics likely to be related to WTP. Conditional logit regressions presented in **Table 11**, controlling for field and country strata fixed effects, show that WTP for publisher open access among this subpopulation of individuals who already pay publication fees is positively related to the citation impact of the journal and the number of publications by the individual in the given year. There is a slight positive correlation with age but it is not statistically significant.

Table 11. Determinants of the probability of accepting an extra publisher charge for OA

Conditional logit estimates, fee-paying authors

	Coefficient	Standard Error	P-value
Amount proposed (baseline=USD99)			
USD299	**-0.70**	0.33	0.035
USD499	**-1.18**	0.37	0.002
USD500	**-2.02**	0.43	0.000
USD999	**-2.14**	0.49	0.000
Female	-0.14	0.35	0.683
Age	0.02	0.01	0.133
Journal impact (SNIP)	**0.27**	0.14	0.056
Number of publications in 2011	**0.34**	0.16	0.036
Strata fixed effects (country * field)	Yes		
Log likelihood	-140.75		
Observations	401		
Pseudo R2	0.1523		
NL Wald test of equality between USD499 and USD500 (Irrelevance of rounded amount)	**Chi 2=4.72**		0.029

Source: OECD Pilot Survey of Scientific Authors 2014/15.

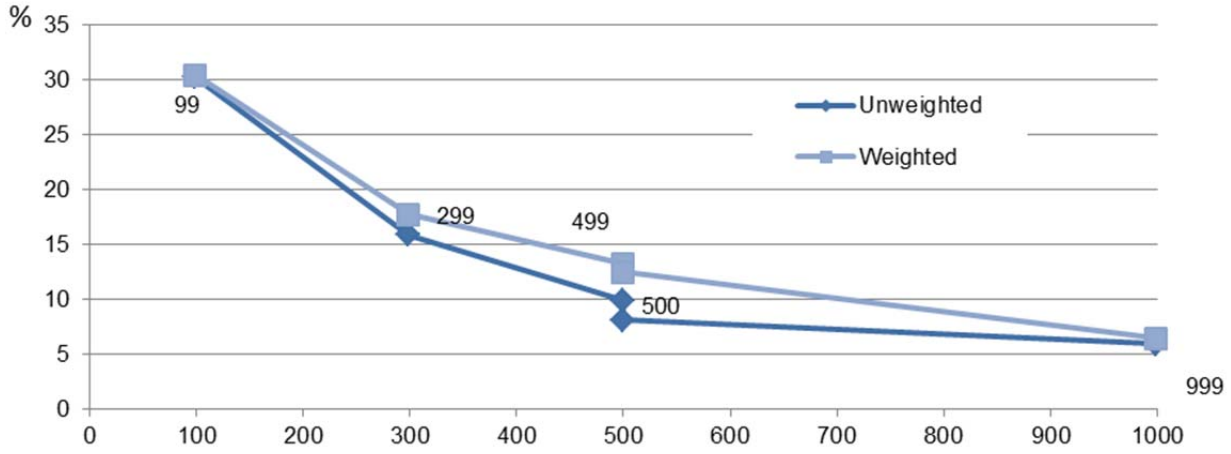
Note: Un-weighted estimates. 126 strata (196 observations) dropped because of all positive or negative values

Different reasons are provided for not being willing to pay. Only 13% of those noted as reason that the document was anyhow publicly available and 4% that it was subject to relatively brief embargo. A quarter refused to accept the "trade" offered, arguing that it should not be the author who paid for access, i.e. what is commonly defined in the CVM literature as a "protest" vote. Many authors are not directly concerned about library budgets, but in line with established CVM methodology, it was imperative to propose a payment mechanism on which the author felt that a budget constraint would apply, hence the reason for referring to own funds. About 15% of authors pointed that paying that amount for OA would not be worthwhile, while a quarter argued a lack of resources. Among those who are willing to pay the amount offered, the most common responses have to do with the wish to make knowledge available for sharing among academics and the availability of dedicated research funds for that purpose. About 8% of those accepting the offer pointed to the potential usefulness of access for non-academics.

A similar contingent valuation scenario was presented to those authors who had not paid publishing fees. The scenario in this case presented the hypothetical situation in which the paper had been accepted, and payment of the proposed one-off fee would guarantee availability to readers on an open and free of charge basis (Q44). The results for this subpopulation (**Figure 21**) point to a lower average willingness to accept the offer (15%) compared to the population of publication fee paying authors (25%). We have no clear explanation for this particular finding, other than this is different, self-selected population more inclined to paying for publication. Again, responses are sensitive to the amount offered, but we do not observe this time such a marked discontinuity when comparing responses to USD499 and USD500 offers.

Figure 21. Willingness to pay randomly-assigned prices for open access, non-fee-paying respondents

Percentage, by amount proposed (USD)



Source: OECD Pilot Survey of Scientific Authors 2014/15.

These results are confirmed by the econometric analysis (**Table 12**).[14] Unlike the fee paying group, in the case of this larger sub-population, having more publications in a given year is negatively correlated with the probability of accepting the price offered (possibly reflecting the potential need to pay for other publications). Acceptance is once more strongly correlated with the journal's citation impact as proxied by the SNIP factor. This may suggest a difficulty in identifying the valuation of open access independently of the value placed on prestige.

Table 12. Determinants of the probability of accepting an extra charge for OA

Conditional logit estimates, non-fee paying authors

	Coefficient	Std. Err.	P-value
Amount proposed (baseline=USD99)			
USD299	**-0.78**	0.14	0.000
USD499	**-1.36**	0.16	0.000
USD500	**-1.58**	0.17	0.000
USD999	**-2.04**	0.18	0.000
Female	0.08	0.13	0.550
Age	0.00	0.00	0.375
Journal impact (SNIP)	**0.18**	0.06	0.002
Number of publications	**-0.13**	0.07	0.050
Strata fixed effects (country * field)	Yes		
Log likelihood	-964.305		
Observations	3129		
Pseudo R2	0.1004		
NL Wald test of equality between USD499 and USD500	Chi-2=1.27		0.26

Un-weighted estimates.

143 strata (677 observations) dropped because of all positive or negative values

Source: OECD Pilot Survey of Scientific Authors 2014/15.

The reasons provided for rejecting the offers are very much similar to those provided in the publication fee-paying population, i.e. an alleged lack of resources and unwillingness to accept the payment vehicle (out of own pocket), which may point to a difficulty in internalising the value of openness to others. Approximately 13% refuse because the document is already openly available. Interestingly, 30% of those alluding to a lack of resources would have agreed to pay the amount proposed had it been accepted in a top-ranked journal. In total, nearly a quarter of "no" responders would reconsider their decision if the acceptance had come from a top-ranked journal in their field. This result hints again at authors principally valuing the credentials conferred by the status of a journal relative to access considerations.

Table 13. Willingness to pay randomly assigned prices for open access, pooled results

Conditional logit estimates, fee paying authors

	1	2	3
Publication fee-paying population	0.257	0.236	0.154
	(1.92)	(1.75)	(0.86)
Document is OA from publisher	0.082	0.115	0.113
	(0.65)	(0.93)	(0.89)
Document is OA from repository	-0.102	-0.108	*-0.159*
	(0.88)	(0.94)	(1.49)
Amount proposed (baseline=USD99)			
USD299	**-0.923**	**-0.939**	**-0.967**
	(7.57)**	(7.78)**	(7.85)**
USD499	**-1.384**	**-1.405**	**-1.446**
	(9.45)**	(9.39)**	(9.38)**
USD500	**-1.680**	**-1.693**	**-1.713**
	(10.36)**	(10.14)**	(9.32)**
USD999	**-2.102**	**-2.110**	**-2.128**
	(15.97)**	(16.35)**	(15.47)**
Subject field (baseline=arts and humanities)			
Business	0.218	0.198	0.127
	(1.34)	(1.20)	(0.65)
Chemical engineering	0.330	0.263	0.314
	(1.64)	(1.29)	(1.37)
Immunology	**1.040**	**1.005**	**1.024**
	(5.91)**	(5.22)**	(4.92)**
Materials science	**0.308**	0.266	0.196
	(2.20)*	(1.76)	(0.99)
Neuroscience	**0.960**	**0.917**	**0.897**
	(5.47)**	(4.89)**	(4.44)**
Physics and astronomy	**0.443**	**0.410**	0.332
	(3.08)**	(2.60)**	(1.89)
Female	0.028	0.043	0.074
	(0.35)	(0.54)	(0.67)
Age	-0.002	-0.001	-0.003
	(0.63)	(0.37)	(0.76)
Higher education	-0.019	-0.023	-0.003
	(0.18)	(0.20)	(0.03)
Journal impact (SNIP)		**0.196**	**0.179**
		(4.45)**	(3.38)**
Missing impact factor		0.035	0.001
		(0.16)	(0.01)
Number of publications 2011		-0.089	-0.082
		(1.89)	(1.58)
Earnings 2015			0.002
			(1.51)
Dedicated resources for dissemination			**0.328**
			(3.28)**
Number of observations	4,557	4,557	3,807

Source: OECD Pilot Survey of Scientific Authors 2014/15. * $p<0.05$; ** $p<0.01$

These results are confirmed by pooling the responses to the contingent valuation scenarios for both samples, which are modelled using country of affiliation as fixed effects (**Table 13**). Willingness to pay is higher for the fee paying population but this effect is imprecisely estimated. Willingness to pay for the proposed fees is also systematically higher in natural science fields, relative to social sciences and humanities.

There is no strong relationship between the document access status and the responses to the contingent valuation exercise, possibly due to available sample sizes. When the document is available from a repository, the results tentatively point to a higher reluctance to pay for access, suggesting that the availability of resources for dissemination may play a role in driving acceptance of prices offered. Price effects in demand for open access thus appear to be significant, as chiefly demonstrated by these responses to randomly assigned amounts which imply a downward sloping demand curve for open access. A dominant factor is journal impact measure (SNIP), indicating that demand for open access is highly contingent on the reputation associated to the vehicle used to disseminate the research results.

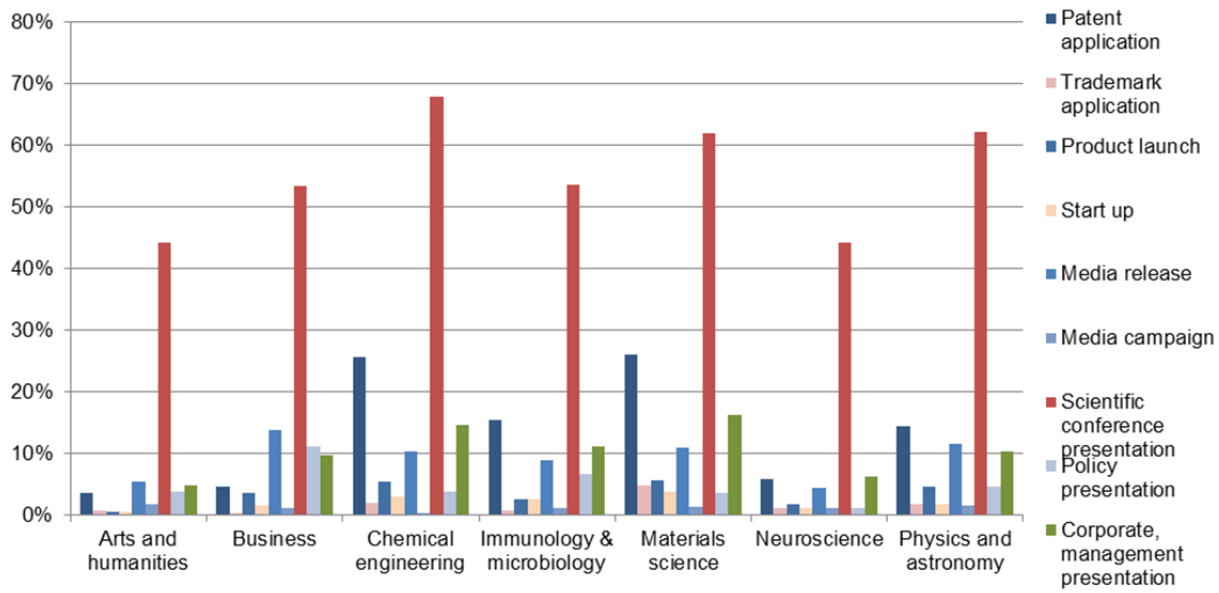
These implied demand curves can be used to derive measures of scientific author economic surplus associated with open access. Given the small marginal cost of providing access, the benefits may be relatively high even despite the higher proportion of negative responses to the hypothetical payment scenarios. This applies provided that we genuinely capture author demand for open access. To the extent that we may be capturing demand for buying into journal prestige, this would require more careful economic analysis as this comes at a cost to other authors and the overall efficiency of the system. All this is suggestive of problems in the assurance and dissemination marketplace given the challenge of internalising the benefits of access and use among those excluded by current access models.

3.6. Disclosure, access to data and research reproducibility

Disclosure

Scientific communication through peer reviewed journals such as those within the scope of this survey only account for a fraction of the overall disclosure and use of scientific research. In our survey, we explored other concurrent events to the submission and publication of the documents.

Figure 22. Journal submission aligned to other disclosure and initiatives



Source: OECD Pilot Survey of Scientific Authors 2014/15

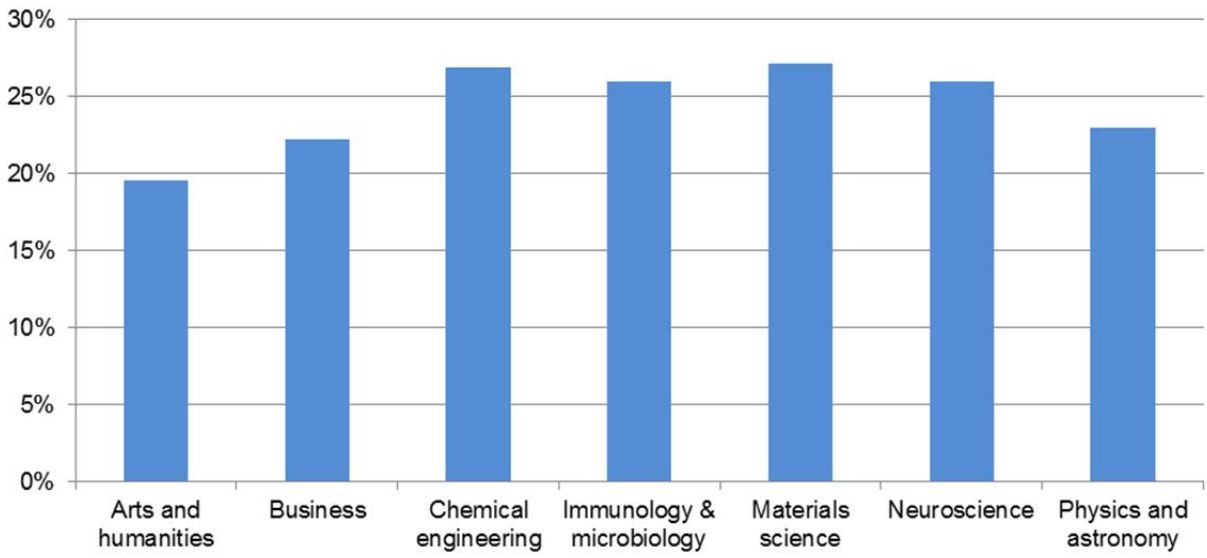
Figure 22 shows that scientific conferences are the predominant and most common form of dissemination alongside the journal submission of an article for publication. This is not surprising as this is a major mechanism for collecting feedback and generating visibility among peers, some of which may be asked to review the article. All other forms of dissemination follow by a distant margin. In the case of Chemical engineering, Materials science and Immunology it is possible to point to the significant levels of parallel patent applications. Trademark applications are indeed very rare, as are product launches or starting up a new business. Media releases and campaigns are more common in business, but less so in the arts/humanities and in materials science. 10% of business authors also report presentations for policy makers, well above other fields. Presentations to management are most common in chemical engineering and materials science.

Data

Our questionnaire inquired about different aspects of access to data and related records, although in order to make questions relevant to all domains, the study compromised on the level of detail and clarity about what was covered by this potentially broad concept. A large number of authors replied pointing out that their work did not entail any form of data. Had we chose to conduct a more data-oriented questionnaire, we would have missed out on responses concerning access to documents.[15]

Notwithstanding these limitations, our survey provides some indicators on the openness of data arising from scientific research and its relevance for research reproducibility. Nearly 30 % of respondents indicated that neither editors nor reviewers had asked for underlying data. When this happened, that was mainly for quality assurance rather than dissemination purposes.

Figure 23. Authors asked to share data or records, by field

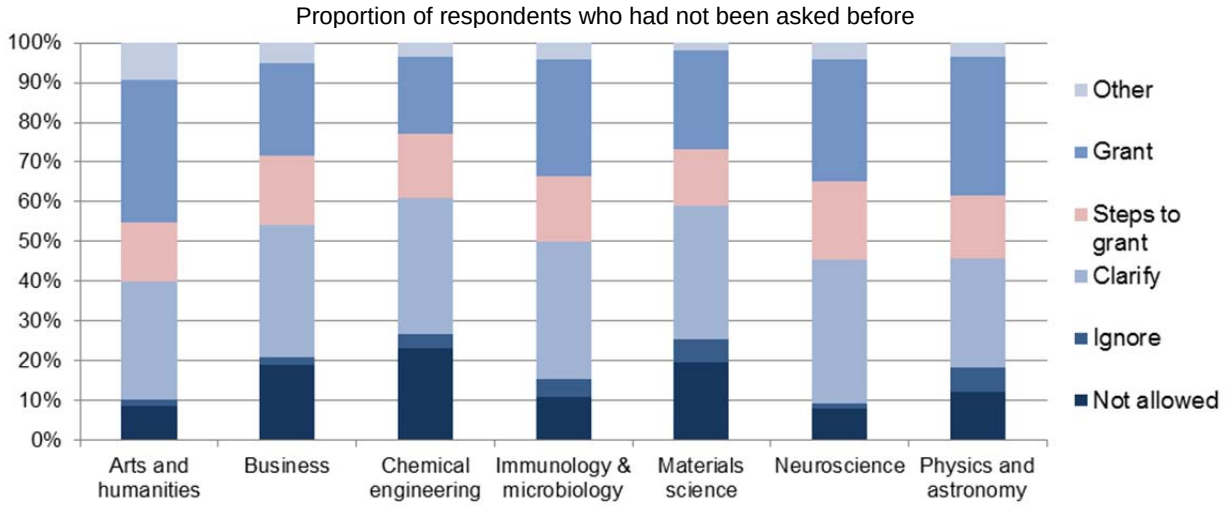


Source: OECD Pilot Survey of Scientific Authors 2014/15

Only about 25% of authors report having been asked to share data after publication (**Figure 23**). Data requests are relatively less common in the humanities and in the business domains. Neuroscience is the domain with the highest propensity among authors indicating having granted access to the data. Scientific authors working in materials science experienced the largest legal difficulties that prevented them from sharing data, possibly related to the stronger links to industry. A similar pattern is found among those who reported never having been asked to share data, in response to a question on what they would do if asked. Materials science and Chemical engineering, as fields with high proximity to the market through their

inventions, report the largest degree of legal obstacles preventing them from sharing. In a majority of cases, respondents appear to buy time by seeking to clarify the nature of the request and only relatively few appear to be ready to share.

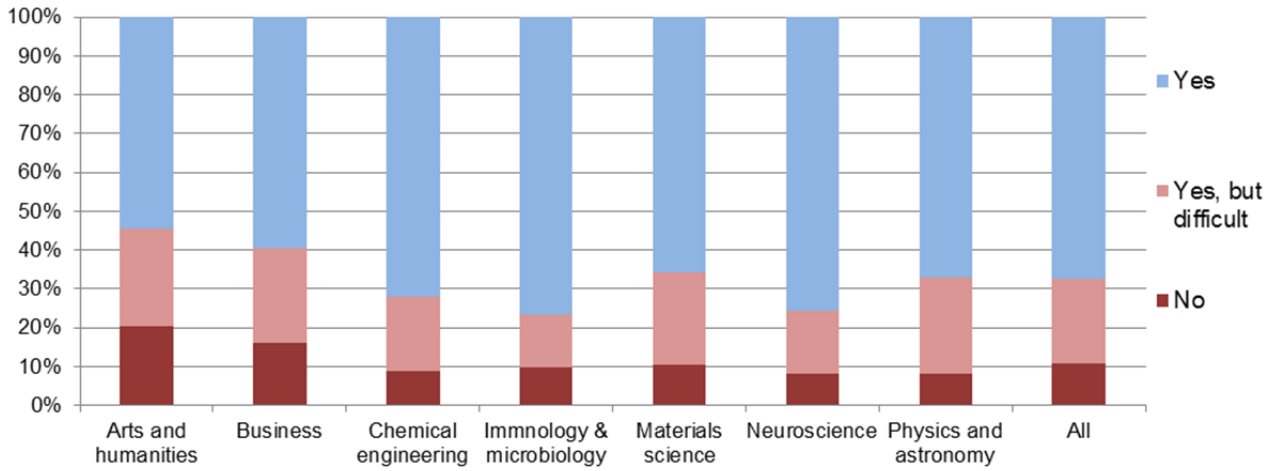
Figure 24. Author's expected response if asked to share data



Source: OECD Pilot Survey of Scientific Authors 2014/15.

A major scientific and policy concern is the extent to which research findings can be reproduced or not with available data. Our question to respondents on whether their research was easily reproducible yielded in general positive answers (**Figure 25**).[16] In immunology and neuroscience, only a quarter indicated that results could not be possibly replicated with publicly available data or with great difficulty. With the benefit of hindsight, we should have asked authors themselves if they had experienced such challenges in reproducing existing results and contrasted with this response.

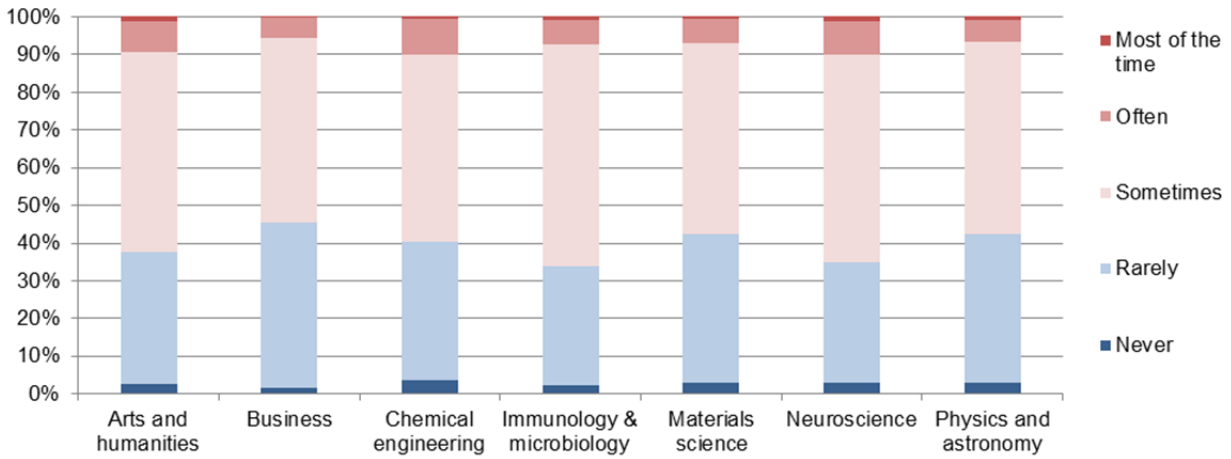
Figure 25. Is your research reproducible with publicly available information?



Source: Preliminary analysis of OECD Pilot Survey of Scientific Authors 2014/15.

Reproducibility of results is an important component of quality assurance. Our survey touched briefly upon the extent to which authors come across errors in the literature. Most respondents (Q21) indicate that errors are found rarely or sometimes in the literature. Fewer than 10% indicate finding errors on a regular basis (often or most of the time). Interestingly, responses are very similar across fields (**Figure 26**).[17]

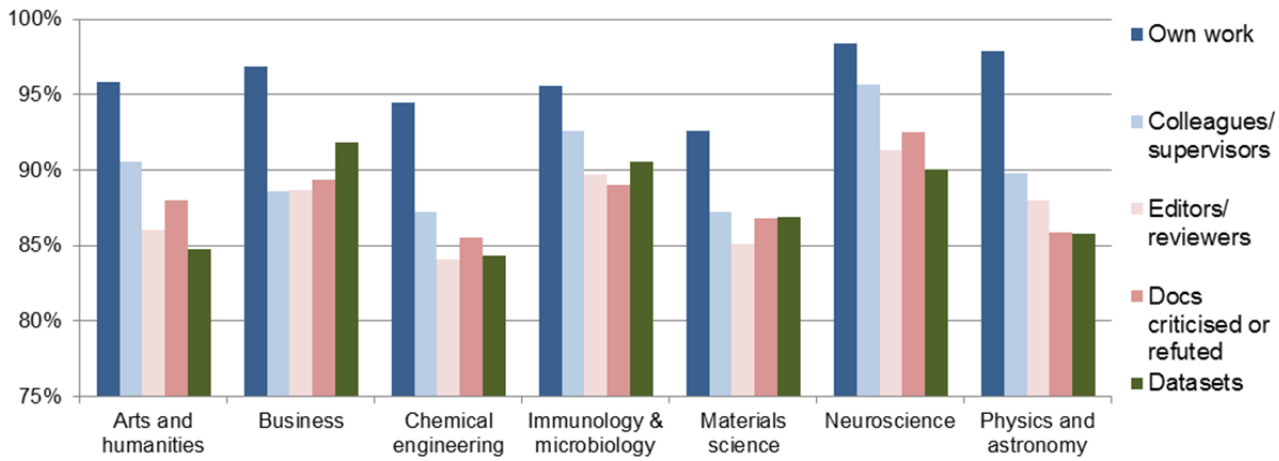
Figure 26. Reported incidence of errors in the literature



Source: OECD Pilot Survey of Scientific Authors 2014/15.

The existence of a small fraction of respondent indicating errors appears consistent with a system in which peer reviewers do their job to high standards, but may also reflect a lack of information upon which to detect such errors. In a previous chart (**Figure 14**), we documented how several authors pointed to the challenge of assessing the quality of research sources and knowing where to find relevant sources. Information on the sources cited in the documents (**Figure 27**) indicates a healthy extent of scientific criticism, judging by the large fraction of authors who refer to documents whose methods, findings and conclusions are criticised or even refuted the author's work. This figure is nonetheless dwarfed by references to the authors themselves, their colleagues, advisors, and even reviewers and editors.

Figure 27. Proportion of authors citing different sources as bibliographic references



Source: OECD Pilot Survey of Scientific Authors 2014/15.

References to databases are fairly modest in scale. These are most common in business, immunology and neuroscience. Lack of awareness concerning citation standards for databases can be a major hindrance. This could be highly relevant for credit attribution to those scientists that specialise in developing, curating and disseminating databases, which in turn may impact on data sharing behaviour.

4. Perspectives on the profiles of scientific authors

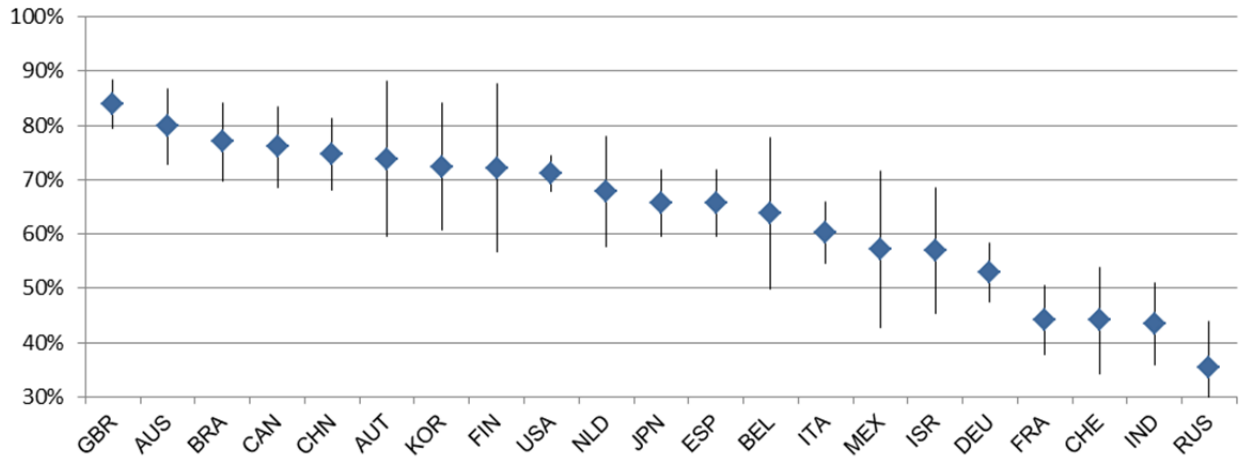
This section provides a basic overview of some relevant features of the population captured by this study, drawing upon the wealth of information collected. It is important to note that this is a very peculiar group of individuals. Not all researchers publish in scientific journals, let alone in journals listed in a major global index, and not all of them feature as corresponding authors. This should be borne in mind when considering results globally or for specific countries.

4.1. Affiliation of scientific authors

Based on the analysis of the affiliation names in the Scopus database and the occurrence of key words such as college, university and other variations, authors were tagged as being part of the Higher Education (HE) sector.[18] **Figure 28** shows a significant degree of variation in the proportion of authors whose main affiliation corresponds to a higher education institution, with a global average around 70%. Less than 50% of authors in France, Switzerland, India and the Russian Federation are in HE. In these countries, government institutions, business and international organisations appear to play an important role. These HE proportions are largest for authors publishing in the fields of Business and Arts and humanities.

Figure 28. Proportion of scientific authors affiliated to higher education institutions

Weighted percentage, for selected countries



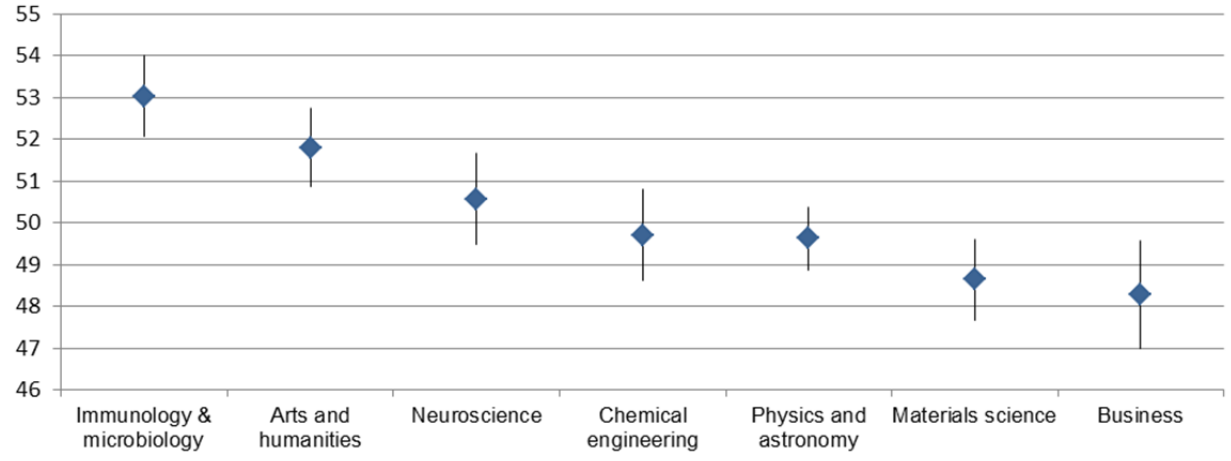
Source: Preliminary analysis of OECD Pilot Survey of Scientific Authors 2014/15.

4.2. Age and gender profile of scientific authors

Figures 29, 30 and 31 report on the age profiles of scientific authors by field and country. The results show the group of corresponding authors is a relatively mature population, with significant differences by field (youngest authors in business and materials science, older in immunology) and by country.

Figure 29. Corresponding authors' age, by subject field

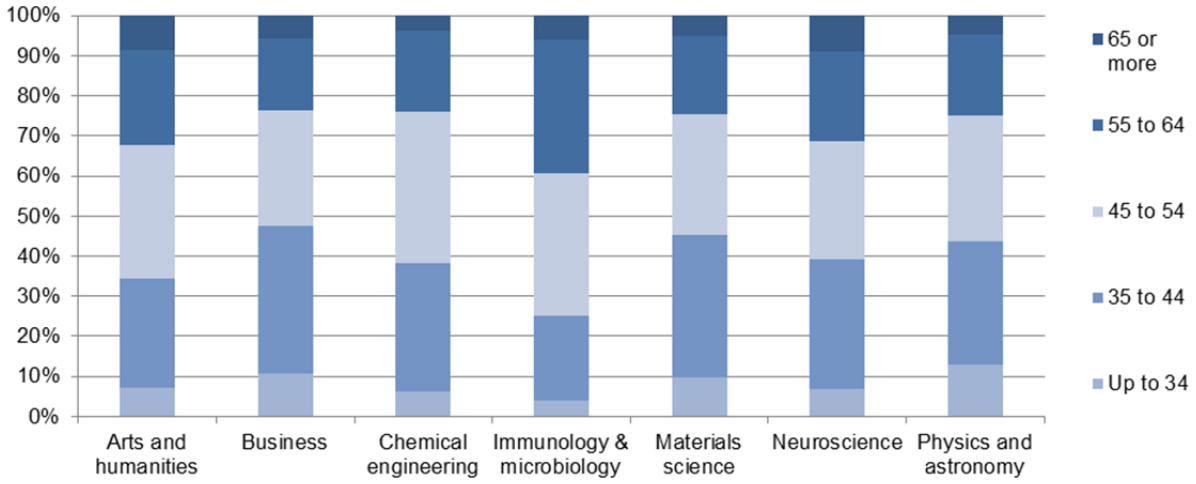
Years, weighted average and 95% confidence intervals



Source: OECD Pilot Survey of Scientific Authors 2014/15.

Figure 30. Corresponding authors' age distribution, by subject field

Weighted percentages for age groups

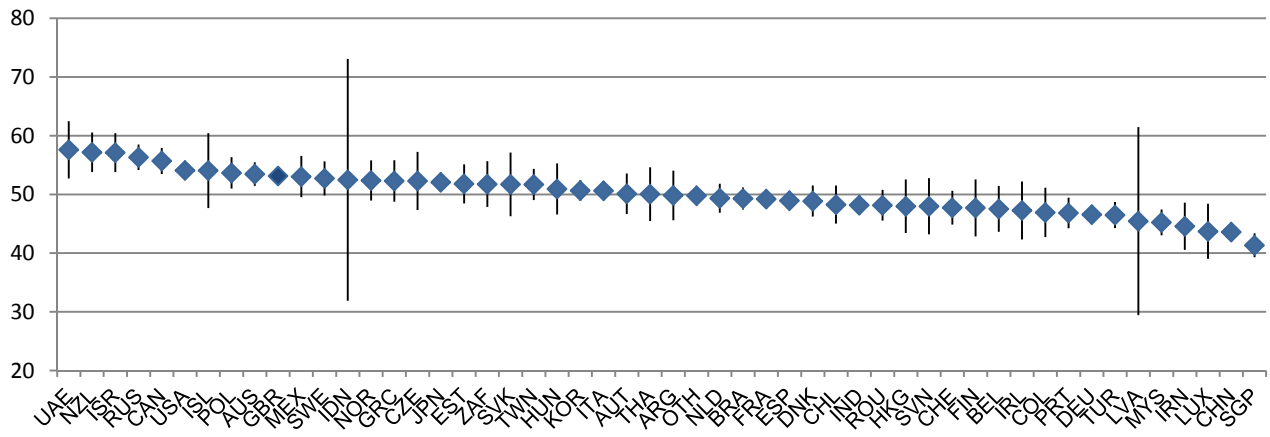


Source: OECD Pilot Survey of Scientific Authors 2014/15.

There are also significant age differences across the range of countries and economies covered in this survey, as seen in **Figure 31**. Chinese authors participating in this study are among the youngest, but this may reveal a participation bias among younger cohorts more likely to have learned English.

Figure 31. Corresponding authors' age, by country or economy of authorship affiliation

Weighted average and 95% confidence intervals

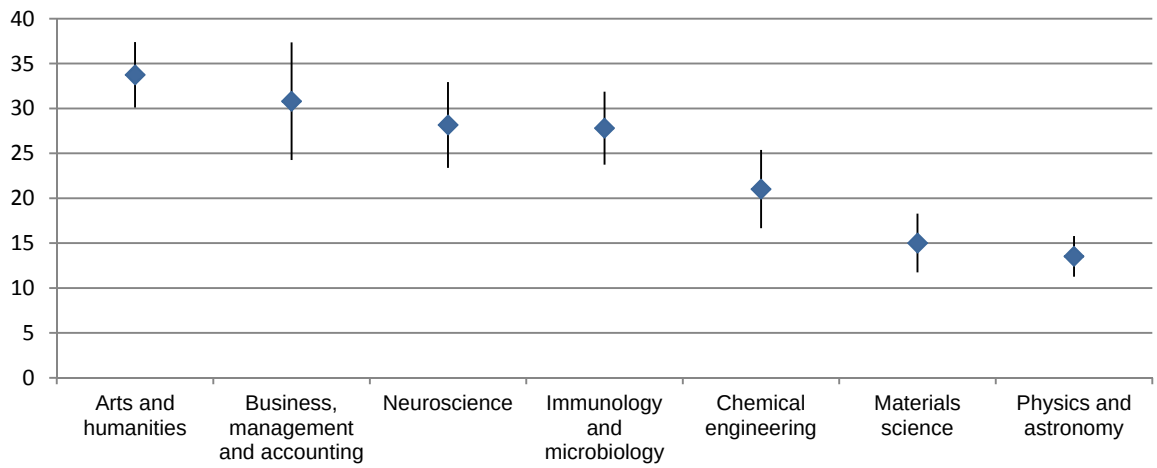


Source: OECD Pilot Survey of Scientific Authors 2014/15.

Compared to the general population, women are underepresened in the population of corresponding authors, a feature that the literature appears to confirm. Our figures present a slightly more male centric gender distribution than that available for doctorate holders or Frascati Manual-based measures of research workforce, possibly reflecting additional challenges for women to feature as corresponding authors for indexed publications. The distribution of women authors across fields also reveals some interesting patterns. Representation is highest in the social sciences, especially in the Arts (above 30%), while it is lowest in Physics, followed by Materials cience and Chemical engineering at 15% or less (**Figure 32**).

Figure 32. Female corresponding authors, by publication subject field

Weighted percentage of all authors within field and 95% confidence intervals

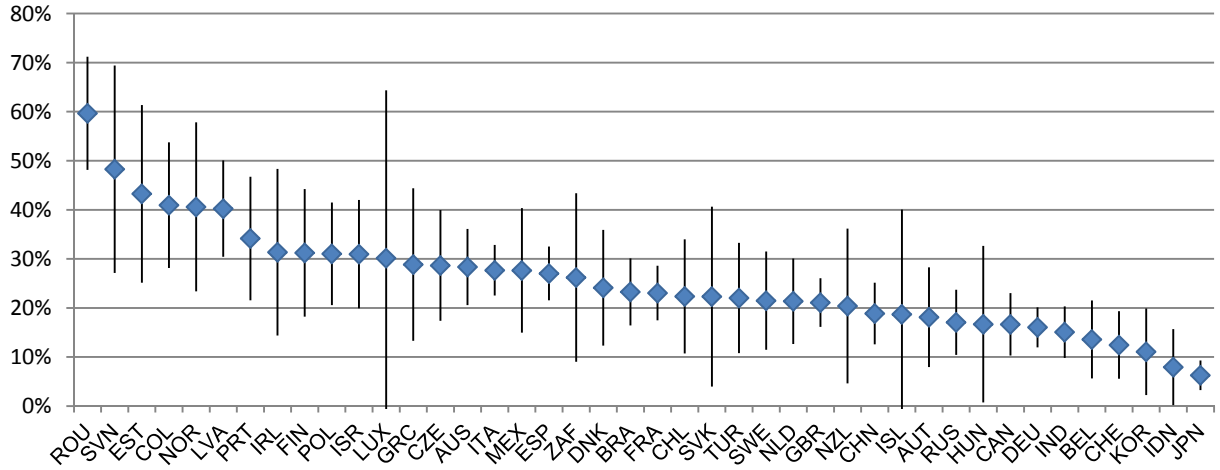


Source: OECD Pilot Survey of Scientific Authors 2014/15.

Across countries, it is possible to note some sizeable differences in the female share among corresponding authors (**Figure 33**). The largest rates, although imprecisely estimated, are found among European transition economies. Low female shares are found in Japan, Indonesia and Korea.

Figure 33. Female corresponding authors, by country / economy

Weighted percentage of all authors within field and 95% confidence intervals



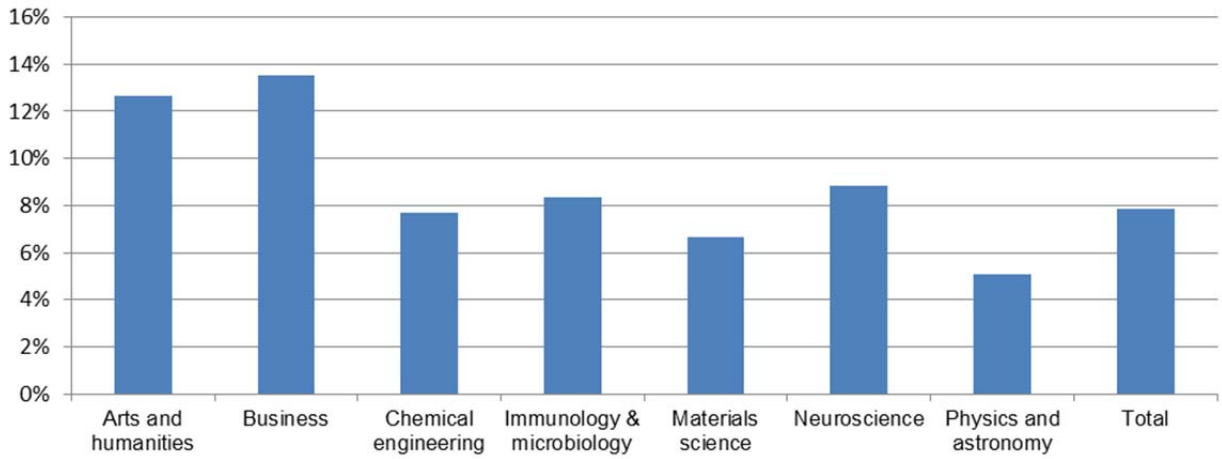
Source: Preliminary analysis of OECD Pilot Survey of Scientific Authors 2014/15.

These results are fairly consistent with recent findings from the EU SHE Figures study carried out by ScienceMetrix (European Commission, 2015), which inferred the gender of scientific authors based on their names with a country-level correlation of nearly 65% notwithstanding the use of different bibliometric databases and the differences in field coverage across studies.

4.3. Research status

We investigated the current status of scientific authors in relation to research. On average, more than 90% of scientific authors report being employed as researchers, using the OECD definition contained in the Frascati Manual (**Figure 34**). This figure may overstate the actual figure to the extent that survey completion is likely to be lower among individuals who have left their institutions after they submitted their papers.

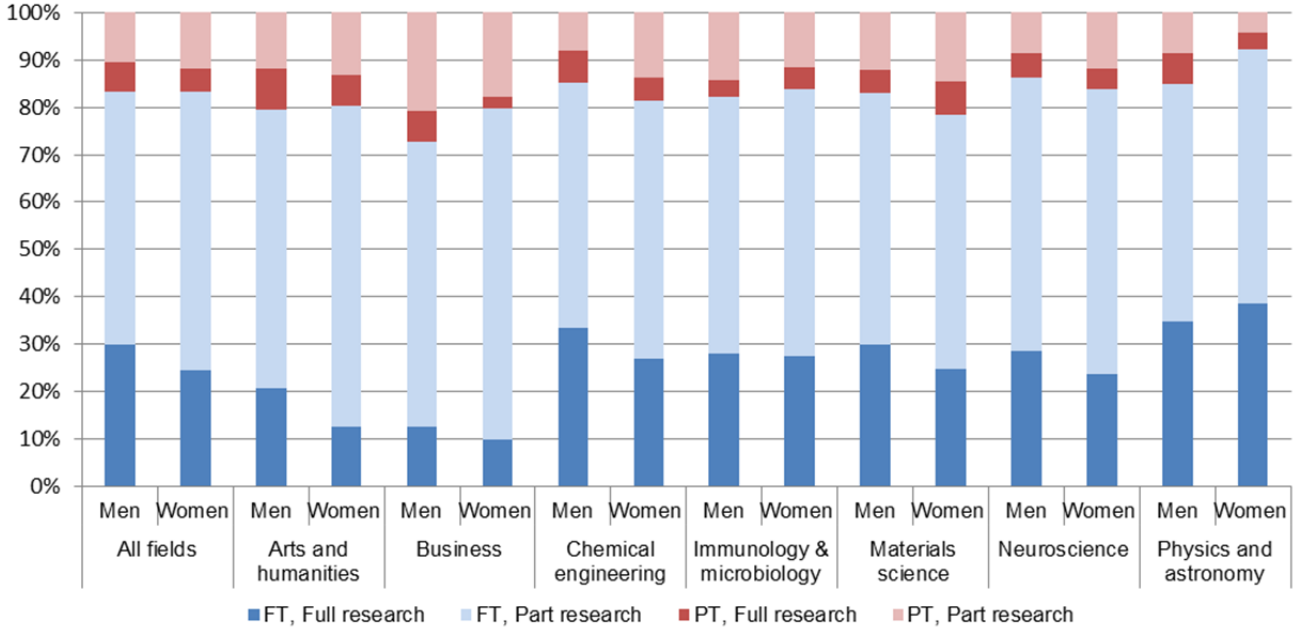
Figure 34. Scientific authors not identifying themselves as researchers



Source: OECD Pilot Survey of Scientific Authors 2014/15.

Among those who report doing research, different modes of engagement are identified (**Figure 35**). Most researchers are working full time, but within those, complete dedication to research is not the norm. These full time researchers dedicate part of their time to other activities. A similar split is found among those working on a part time basis. Complete dedication to research is the exception. To our surprise, it is impossible to find systematic differences by gender among corresponding authors within a given field. This may reflect the very selective nature of this population.

Figure 35. Modes of research engagement among corresponding authors



Source: OECD Pilot Survey of Scientific Authors 2014/15.

4.4. Educational background of scientific authors

Table 14 maps for each article's subject field the distribution of the most frequently reported highest qualification subject fields attained by the authors. In the vast majority of cases, these qualifications correspond to doctorate level degrees.

A very high percentage of authors report to have received formal qualifications in interdisciplinary subjects. This is the case of articles in business, immunology/microbiology, neuroscience and physics.

In contrast, a highly interdisciplinary domain like Materials science reflects a low share of individuals who have attained their highest qualifications in inter or multidisciplinary domains. We observe instead that physics and chemistry are the most frequently reported qualifications.

It is also interesting to note that the Arts and Humanities are quite heterogeneous, reflecting the diversity of this grouping into which very distinct subfields are combined given its small size. Economics is the most frequent qualification among authors of papers classified into Business, despite the existence of a separate Economics domain.[19]

Table 14. Main educational background field, by publication subject field

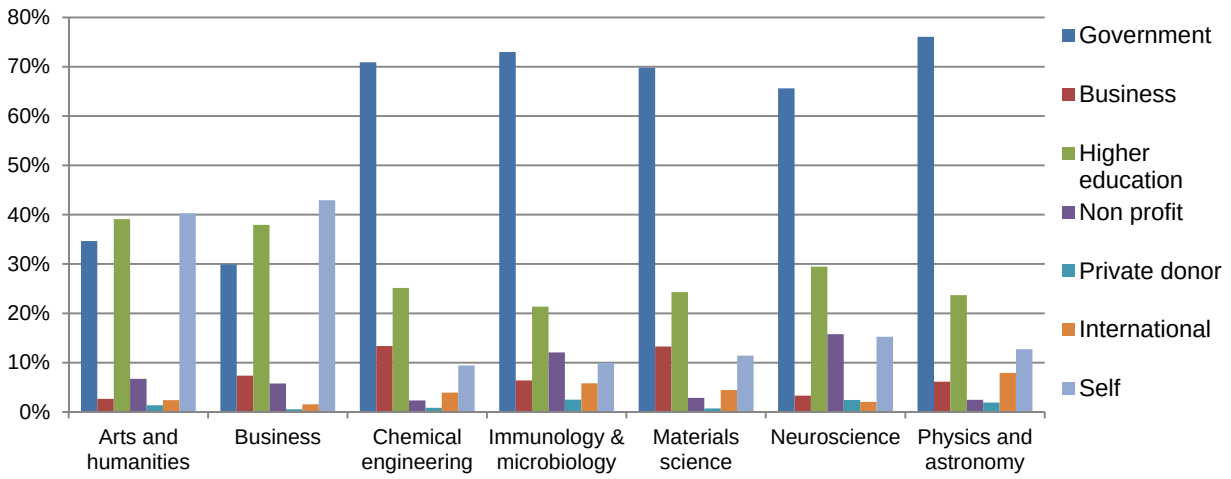
Un-weighted shares within subject field

Arts and humanities		Business, management and accounting		Chemical engineering		Immunology and microbiology		Materials science		Neuroscience		Physics and astronomy	
History and archaeology	18.6	Economics	21.2	Chemistry	33.2	Biology	27.1	Physics	23.3	Medicine	29.1	Physics	51.6
Literature and linguistics	13.8	Management and administration	16.4	Chemical engineering & processes	20.6	Medicine	24.4	Chemistry	22.7	Psychology	19.0	Chemistry	11.8
Philosophy and ethics	12.1	*Inter-disciplinary, social and behavioural*	5.3	Physics	5.6	Inter-disciplinary, biosciences	11.6	Mechanics and metal trades	8.3	Biology	16.2	*Inter-disciplinary, physical sciences*	4.4
Sociology and cultural studies	6.4	Marketing and advertising	5.1	Engineering and engineering trades NEC	4.9	Biochemistry	9.4	Chemical engineering and processes	6.0	*Inter-disciplinary, biosciences*	10.8	Mathematics	4.2
Psychology	5.7	Political sciences and civics	4.0	Biology	3.3	Veterinary	5.2	Engineering & engineering trades NEC	5.6	Biochemistry	5.1	Electronics and automation	4.1
Inter-disciplinary, humanities	4.1	Engineering and engineering trades NEC	4.0	Biochemistry	2.9	Agriculture	3.8	Electronics and automation	4.9	Health NEC	2.5	Engineering and engineering trades NEC	3.5
Education science	3.7	Psychology	3.4	Mechanics and metal trades	2.7	Health NEC	2.7	*Inter-disciplinary, physical sciences*	3.9	Physics	2.1	Mechanics and metal trades	3.2
Religion	3.1	Accounting and taxation	3.2	Pharmacy	2.7	Pharmacy	1.9	Building and civil engineering	2.8			Electricity and energy	2.5
Inter-disciplinary, social and behavioural	2.8	Finance, banking and insurance	3.1	*Inter-disciplinary, physical sciences*	2.5	Chemistry	1.7	Electricity and energy	2.6			Earth sciences	1.8
Inter-disciplinary, languages	2.7	Inter-disciplinary, business	3.1					Materials	2.5			Chemical engineering and processes	1.6
Other	27.2	Other	31.2	Other	21.6	Other	12.2	Other	17.4	Other	3.1	Other	11.2

4.5. Sources of funding for research

Government is the most often reported source of funding for the work that gave rise to the document covered within the study (**Figure 36**). Only in the case of arts and humanities and business do authors report HE and own resource-based funding for their work at a level parallel to that of government. Non-profit funding is less common and concentrated among the life/health science fields. Funding from business is concentrated in the fields of chemical engineering and materials science. This is consistent with some of the openness patterns discussed earlier in this report.

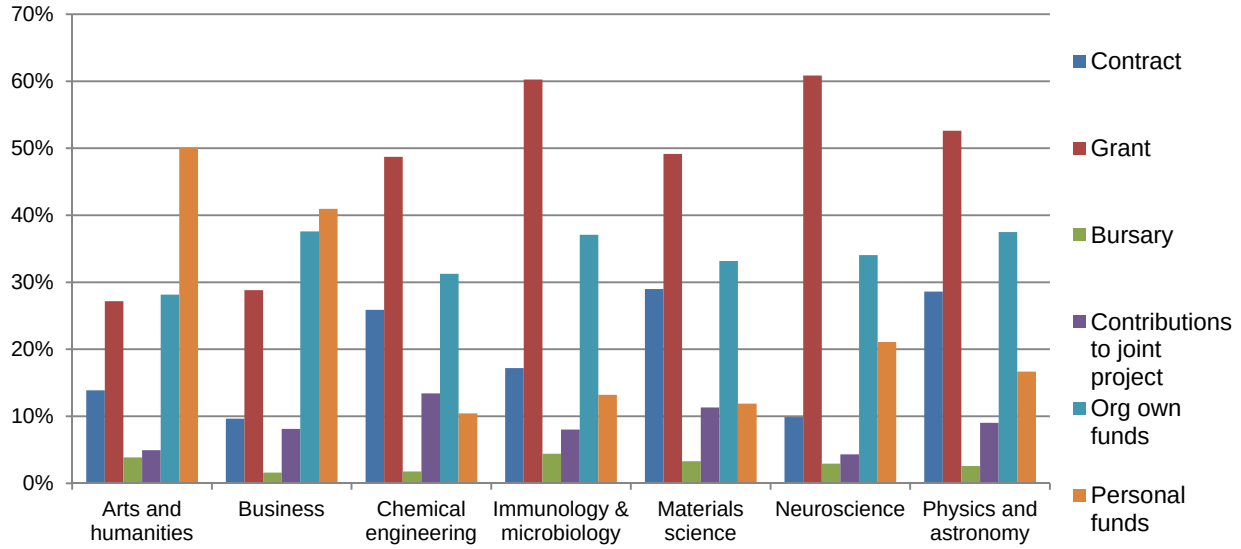
Figure 36. Reported sources of funding, by field



Source: Preliminary analysis of OECD Pilot Survey of Scientific Authors 2014/15.

Further detail on the type of funding received is available in **Figure 37**.

Figure 37. Types of sources of funds used, by field

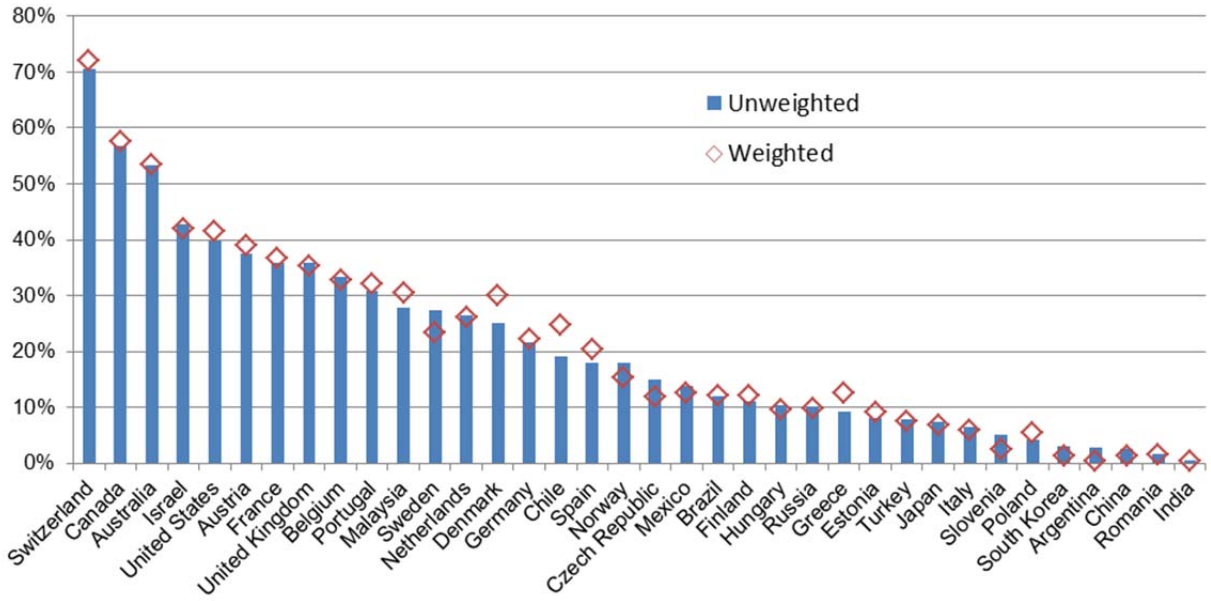


Source: OECD Pilot Survey of Scientific Authors 2014/15.

4.6. Mobility patterns of scientific authors

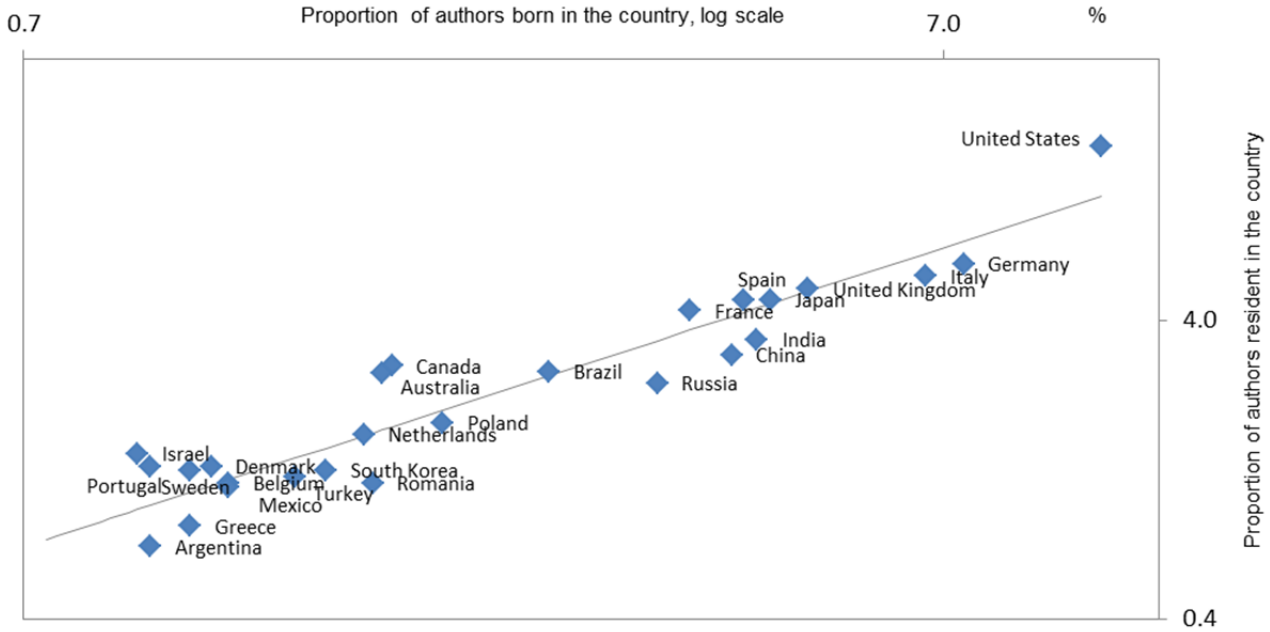
Information on the country of birth and usual residence from this survey allows us to identify which countries have the largest share of authors born abroad (**Figure 38**). In the case of Switzerland, Canada and Australia, more than 50% of corresponding authors were born abroad. These are very significant mobility rates.

Figure 38. Proportion of scientific authors born in another country
By country of residence, weighted and un-weighted results (countries with more than 35 observations)



Source: OECD Pilot Survey of Scientific Authors 2014/15.

Figure 39. Percentage (un-weighted) distribution of scientific authors, by country of birth and residence



Source: OECD Pilot Survey of Scientific Authors 2014/15.

Figure 39 compares the un-weighted distribution of survey respondents by country of birth with that by country of residence. The United States, Australia, Canada and Israel are the main net recipients, in relative terms, as implied by the vertical distance with respect to the line that indicates the hypothetical situation in which the proportion of resident scientists is equal to that of those born in the country. In this study, we have not collected information on the timing of mobility, the analysis of which would warrant a survey of its own. **Table 15** provides the underlying data, highlighting those countries or economies where the estimated number of residents is higher than that of native authors**.**

Table 15. Country of birth and residence

As a percentage of all respondents, 35 countries with highest number of respondents resident within country

	Country of birth		Country of residence		
	Unweighted	Weighted	Unweighted	Weighted	# Resident >># Born authors
United States	10.4	11.8	15.3	18.6	X
Germany	7.4	6.7	6.2	5.8	
Italy	6.7	4.0	5.6	3.1	
United Kingdom	5.0	4.7	5.1	5.1	X
Japan	4.5	5.3	4.7	5.3	
India	4.4	4.6	3.5	3.6	
Spain	4.3	2.5	4.7	2.7	X
China	4.1	17.1	3.1	15.7	
France	3.7	3.3	4.3	3.9	X
Russia	3.4	3.1	2.5	2.3	
Brazil	2.6	2.0	2.7	2.0	
Poland	2.0	1.5	1.8	1.4	
Canada	1.8	1.7	2.8	2.7	X
Australia	1.7	1.3	2.7	1.8	X
Romania	1.7	1.0	1.1	0.6	
Netherlands	1.6	1.4	1.7	1.4	
South Korea	1.5	2.9	1.3	2.5	
Turkey	1.4	1.5	1.2	1.2	
Belgium	1.2	1.0	1.1	1.0	
Mexico	1.2	0.9	1.1	0.7	
Denmark	1.1	0.6	1.3	0.7	
Greece	1.1	0.7	0.8	0.5	
Sweden	1.1	0.9	1.3	1.0	X
Argentina	1.0	0.7	0.7	0.5	
Portugal	1.0	0.5	1.3	0.7	X
Israel	0.9	0.5	1.4	0.6	X
Iran	0.9	1.4	0.6	1.0	
Colombia	0.9	0.4	0.7	0.2	
Finland	0.9	0.6	0.9	0.6	
Hungary	0.9	0.5	0.7	0.3	
Switzerland	0.9	0.7	1.7	1.1	X
Ukraine	0.9	0.6	0.6	0.4	
Austria	0.9	0.6	0.9	0.6	
Slovenia	0.8	0.2	0.8	0.2	
Czech Republic	0.7	0.6	0.8	0.6	

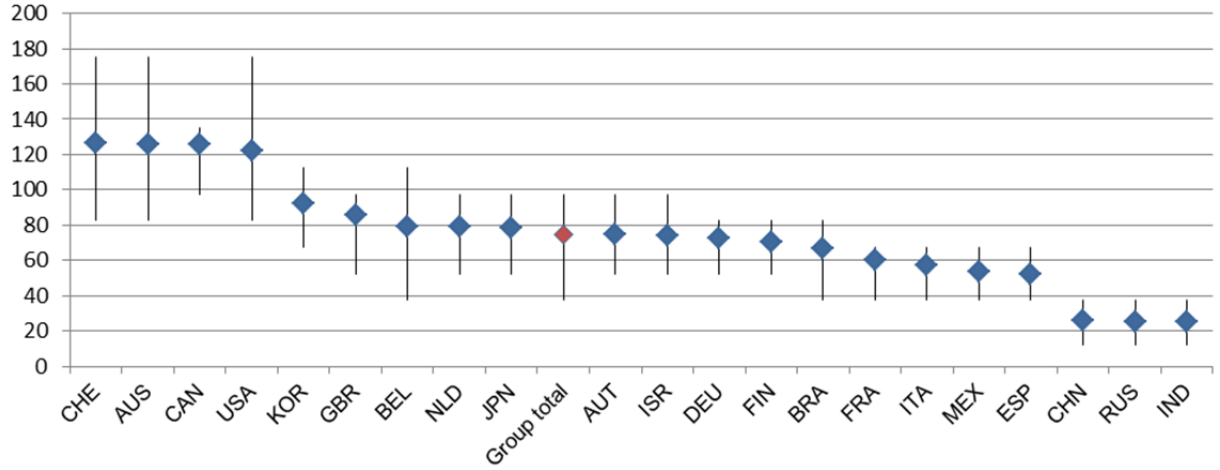
Source: OECD Pilot Survey of Scientific Authors 2014/15.

4.7. Scientist earnings

Because of the sensitivity of the information and its position towards the end of the questionnaire, our measurement of earnings (**Figure 40**) is subject to a significant item non-response rate of nearly 20% among survey respondents. Data are available for 4391 individuals with complete response records. Estimates are based on mid points for the income bands, with an adjustment for lower and upper bands. The un-weighted measure of average annual earnings for the entire the sample is USD 72 thousand, compared with USD72.9K with sampling weights, and USD70K with weights partly adjusted for unit non response.

Figure 40. Annual earnings of scientific authors, by country

Average and interquartile range, thousand USD, selected countries



Source: OECD Pilot Survey of Scientific Authors 2014/15. Not adjusted for differences in living standards.

It is not possible to compute country-level standard errors for countries without further adjustments because of lack of variation within some strata. Results indicate a cluster of highly paid (on average) scientific authors in Switzerland, Australia, Canada and the United States. Earnings data are explored in more detail through regression analyses reported in **Table 16** that control for country-field fixed effects:

- Women earn approximately 17% less than their male counterparts, regardless of accounting for several types of controls, including part time status and measures of scientific reputation.
- There is a very significant age/experience premium, although it decreases with age, which compensates at least in part for the long education spell for those going into research careers.
- Authors born in different countries earn more than their "national" counterparts. This may reflect unobserved "quality" that is not captured by other indicators.
- There is an earnings premium from publishing in highly cited journals, over and above the premium conferred by the actual citations, which is also positive. This confirms the signalling effect implied by the status of a journal and its valuation by authors in the CVM experiment. There is also some evidence of a premium from being cited in policy documents.
- Authors who are reviewers and editors earn a significant premium, especially if they are remunerated for such activities. This may simply reflect their reputation and experience.
- Scientists in HE institutions earn systematically less than their counterparts in other sectors, which may reflect other non-observable traits such as greater job stability, pensions or freedom to conduct research. Full time dedication to research is also associated on average with lower earnings.

- There is no visible impact of open access on earnings. It is, however, important to note that a single publication is not an accurate measure of the accessibility of an author's publications. Ideally one should look at the entire publication profile. This would be possible using bibliometric information on the OA status of the journals where individuals publish.

Table 16. Determinants of scientific authors' earnings

Least square regression results on log earnings, with country*field strata fixed effects

	1	2	3	4	5
Female	-0.187	-0.174	-0.171	-0.166	-0.163
	(8.87)**	(8.40)**	(8.27)**	(8.05)**	(7.90)**
Age	0.095	0.074	0.074	0.067	0.066
	(17.62)**	(12.83)**	(12.83)**	(11.63)**	(11.41)**
Age squared	-0.080	-0.059	-0.059	-0.053	-0.052
	(15.81)**	(10.84)**	(10.82)**	(9.86)**	(9.62)**
Born other country	0.090	0.087	0.085	0.079	0.080
	(4.24)**	(4.14)**	(4.09)**	(3.83)**	(3.85)**
Number of publications		0.028	0.026	0.020	0.020
		(3.19)**	(2.99)**	(2.34)*	(2.28)*
FT, research part time		0.127	0.129	0.128	0.134
		(6.11)**	(6.23)**	(6.22)**	(6.47)**
PT, research FT		-0.237	-0.233	-0.226	-0.222
		(5.78)**	(5.70)**	(5.53)**	(5.46)**
PT, research PT		-0.092	-0.082	-0.074	-0.071
		(2.94)**	(2.61)**	(2.35)*	(2.25)*
Fixed term contract		-0.050	-0.050	-0.044	-0.045
		(2.48)*	(2.51)*	(2.18)*	(2.23)*
Higher education		-0.077	-0.079	-0.095	-0.090
		(3.94)**	(4.02)**	(4.84)**	(4.52)**
Journal impact (SNIP)			0.049	0.047	0.042
			(4.50)**	(4.33)**	(3.86)**
Missing journal impact			0.034	0.046	0.056
			(1.06)	(1.46)	(1.78)
Log citations			0.032	0.023	0.016
			(2.15)*	(1.58)	(1.11)
Document open access			0.009	0.001	-0.002
			(0.52)	(0.06)	(0.11)
Reviewer, paid and not				0.186	0.172
				(4.18)**	(3.85)**
Reviewer, not paid				0.129	0.120
				(4.09)**	(3.80)**
Reviewer, paid				0.117	0.103
				(2.60)**	(2.30)*
Editor, paid and not				0.143	0.136
				(2.32)*	(2.21)*
Editor, not paid				0.119	0.114
				(6.16)**	(5.92)**
Editor, paid				0.072	0.077
				(1.70)	(1.82)
Cited in (only significant results displayed)					
Peer reviewed, major index					0.075
					(3.11)**
Policy document					0.102
					(2.91)**
Constant	1.335	1.870	1.797	1.833	1.816
	(8.92)**	(11.84)**	(11.34)**	(11.51)**	(11.37)**
R^2	0.59	0.62	0.63	0.64	0.64
N	4,351	4,188	4,188	4,131	4,131

Source: OECD Pilot Survey of Scientific Authors 2014/15. * p<0.05; ** p<0.01

5. Concluding remarks and next steps

This document has reported a set of preliminary findings on a recent OECD pilot initiative aimed at collecting survey data directly from scientific authors in different countries - an international survey of scientific authors (ISSA). The purpose of this report is to promote discussion and feedback gathering by providing a detailed account of the pilot and efforts to carry out a proof of concept for this survey.

This survey's experience is a salutary reminder of the many challenges arising when collecting data through an online survey-based questionnaire and drawing upon a selected population of corresponding authors of scientific publications found in a major commercial index of publications and citations. The experience indicates there is considerable scope for improving the quality of the estimates and experimenting with potential further improvements. This survey's performance can be deemed to provide a lower bound for potential future efforts, bearing in mind the limited financial and time resources available to conduct this pilot. A number of modifications and additional control steps can be introduced in possible future editions to significantly improve survey quality.

i) implementing a more thorough cleaning of the underlying database;

ii) seeking to identify more recent email addresses for the authors;

iii) implementing a shorter questionnaire;

iv) allowing survey completion in the language of respondent's choice;

v) carrying out non response follow-up surveys/calls and analysis;

vi) referring respondents to the results of previous related studies;

vii) developing an OECD internal survey tool to increase user trust in the survey;

viii) securing national institutional endorsement; and

ix) exploring options to increase the value proposition for respondents to encourage participation and to provide truthful responses to all questions.

The potential payoff to this type of effort is considerable. The wealth of information collected and the range of analysis enabled by this survey provides a valuable addition to the toolbox for the analysis of major science and research policy questions from a global perspective in ways that other tools cannot.

The choice of open access as the topic of investigation has helped deliver numerous insights on the barriers and incentives that policy makers need to take into account when designing open access and open science policies. These findings can further support the parametrisation of cost-benefit analyses that so far had to rely on a number of untested assumptions (BIS, 2013). Our evidence suggests that authors' preferences and behaviour are fully aligned to the goals that the market in which they operate values the most, making it necessary to bring in an economic dimension into the analysis of open access and open science. In turn, the economic analysis of OA requires new sources of evidence such as those provided in this study.

The implementation of a survey linked to bibliometric information has also allowed testing the consistency of different data sources and new knowledge about dissemination mechanisms for which open access can be relevant. This posits a promising approach for considering proposals to extend measurement practice beyond currently used indicators, towards indicators that reflect the full breadth of impacts from scientific research and the factors that shape it. The ability to consider several countries in a single study is another major advantage as access to micro-data on researchers is otherwise still limited and cannot be pooled effectively for analytical purposes.

The intention behind this study is not to replace standard approaches but complement them, acting as a form of continued data collection experiments from which standard surveys can draw inspiration and bring such enquiries into the mainstream. The approach tested in this pilot exercise could be in principle applied to other key policy topics for which traditional instruments currently cannot offer appropriate solutions on a sufficiently timely basis. The process leading to the identification and prioritisation of those topics can also provide a unique opportunity to strengthen the communication between data users and providers. As noted in the Daejeon Declaration on Science, Technology, and Innovation Policies for the Global and Digital Age (OECD, 2015c), science ministers from all over the world noted that science, technology and innovation are being revolutionised by the rapid evolution of digital technologies, changing the way scientists work, collaborate and publish; increasing the reliance on access to scientific data and publications ("open science"); opening new avenues for public engagement and participation in science and innovation ("citizen science"); facilitating the development of research co-operation between businesses and the public sector; and contributing to the transformation of how innovation occurs ("open innovation"). In light of this, ministers have invited the OECD to support the improvement of policies for "open science", to continue to develop methods and indicators for assessing the impact of research and innovation policies, and to continue improving statistics and measurement systems to better capture the key features of science, technology and innovation

The direction set out by OECD science ministers paves the way for refining and re-evaluating the ISSA initiative. The OECD is developing a proposal to carry out a second pilot collection of statistical evidence on selected aspects of the digitalisation of scientific research, to be implemented through a centralised, global, online survey of scientific authors. Three possible elements have been suggested as potential objects of analytical focus for the proposed survey as starting point for discussion.

- Documenting patterns of use of "data-intensive" processes of scientific research and discovery.
- Identifying implications and, whenever possible, impacts of different modalities of access to data arising from scientific research.
- Documenting challenges to open data and data sharing in scientific research.
- Document in more detail mechanisms for scientific production, dissemination and societal engagement and impact.

In this effort, the input of recognised experts in the fields of science studies and survey research will be vital in order to develop further the ISSA concept into what may one day become a regular component of the OECD statistical infrastructure on science, technology and innovation. Interested researchers are invited to request access to an anonymised version of the survey microdata for exploration by following the guidance provided

ANNEX 1. SURVEY METHODOLOGY

Table A1. All Science Journals classification (2–digit)

Selected fields for sampling in bold characters

General	10
Agricultural and Biological Sciences(all)	11
Arts and Humanities(all)	**12**
Biochemistry, Genetics and Molecular Biology(all)	13
Business, Management and Accounting(all)	**14**
Chemical Engineering(all)	**15**
Chemistry(all)	16
Computer Science(all)	17
Decision Sciences(all)	18
Earth and Planetary Sciences(all)	19
Economics, Econometrics and Finance(all)	20
Energy(all)	21
Engineering(all)	22
Environmental Science(all)	23
Immunology and Microbiology(all)	**24**
Materials Science(all)	**25**
Mathematics(all)	26
Medicine(all)	27
Neuroscience(all)	**28**
Nursing(all)	29
Pharmacology, Toxicology and Pharmaceutics(all)	30
Physics and Astronomy(all)	**31**
Psychology(all)	32
Social Sciences(all)	33
Veterinary(all)	34
Dentistry(all)	35
Health Professions(all)	36

Source: OECD, based on ASJC classification. http://www.elsevier.com/online-tools/scopus/content-overview

Table A2. Summary of sample frame derivation

Step	Observations
-> Extraction of 2011 documents raw database (Document* fields published 2011, with corresponding author with author id and email address)	
Number of observations	4 333 306
Corresponding to:	
Unique authors	1 234 754
Unique documents	1 965 377
->Cleaning for missing country and patently wrong emails	
Number of observations.	4 327 277
Corresponding to:	
Unique authors	1 233 166
Unique documents	1 962 395
Unique email addresses	1 243 410
->Selection of documents corresponding to chosen fields. Arts and humanities (11) ; Business, management and accounting (14); Chemical engineering (15); Immunology and microbiology (24); Materials science (25); Neuroscience (28) and Physics and astronomy (31).	These account for 19.8% of all documents capture in database
Number of observations	856 263
-> Adoption of reference field per document in order to make documents unique. (A document could have been listed as been in two same subfields of covered field or on subfields of different covered fields).	
Unique documents	554 256
(of which not id. by source id in Scopus journal titles)	35,291 (6.4%)
Unique authors	379 218
Unique email addresses (same authors can have more than one email, but also possible, realised later, that diff. author ids have same email, revealing incomplete disambiguation).	380 297
-> Selection of reference publication per author, at random	(Dropped 175 038)
Target population: Number of observations=number of unique documents=number of unique author IDs	379 218
Stratification and sampling -> stratum1=100000*eco3+field*100 -> randomly select within stratum the max of { stratsize N(i) /(N/planned target of 60K) ; mincell of 30}	
Targeted sample	61 361
Non targeted sample	317 857

Source: OECD Pilot Survey of Scientific Authors 2014/15.

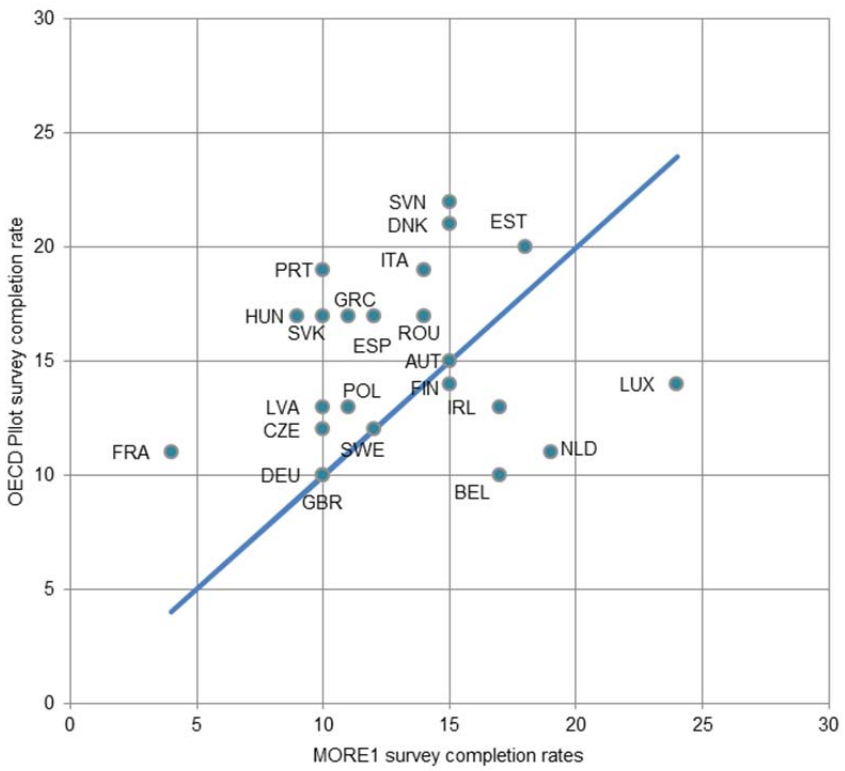
Table A3. Response rate by location of author's affiliation

Country/ economy	Targeted authors (Number)	Bounced (%)	No click (%)	Click, non-response (%)	Partial response (%)	Complete, response submitted (%)
USA	10962	11	76	2	3	8
CHN	9643	15	81	1	2	*2*
DEU	3243	18	66	2	3	**10**
JPN	3122	13	72	2	4	8
GBR	3032	17	69	2	2	10
Other	2413	10	67	4	6	**13**
FRA	2247	15	67	3	4	**11**
IND	2118	7	76	3	5	9
ITA	1756	6	67	3	5	**19**
ESP	1587	7	66	4	5	**17**
KOR	1567	6	85	1	4	*5*
CAN	1462	9	75	3	3	**10**
RUS	1394	12	68	5	5	**10**
BRA	1122	6	75	2	4	**13**
AUS	1094	11	69	3	3	**14**
TWN	989	9	79	3	5	*5*
NLD	895	16	66	4	3	**11**
IRN	875	11	75	2	7	5
POL	750	11	69	3	5	**13**
TUR	730	5	72	4	9	9
SWE	594	15	66	3	4	**12**
CHE	577	21	56	3	3	**16**
BEL	529	10	74	2	3	**10**
MYS	492	7	75	2	6	**11**
ROU	444	9	62	5	8	**17**
MEX	427	12	66	2	6	**14**
CZE	361	12	67	3	6	**12**
ISR	358	12	56	3	6	**23**
FIN	355	22	59	**	**	**14**
PRT	355	17	55	**	**	**19**
DNK	339	11	63	**	**	**21**
AUT	338	15	62	**	**	**15**
UAE	321	14	63	4	6	**13**
SGP	314	4	89	**	**	*4*
ARG	311	7	75	**	**	**13**
ZAF	292	9	73	**	**	**12**
GRC	286	6	70	**	**	**17**
NOR	267	19	58	**	**	**17**
HKG	254	14	77	**	**	6
THA	251	8	74	**	**	**12**
CHL	238	5	69	**	**	**20**
HUN	238	6	68	**	**	**17**
IRL	235	6	76	**	**	**13**
NZL	230	12	71	**	**	**16**
SVN	218	8	56	6	9	**22**
SVK	203	7	61	5	10	**17**
COL	202	5	69	**	**	**16**

Country/ economy	Targeted authors (Number)	Bounced (%)	No click (%)	Click, non-response (%)	Partial response (%)	Complete, response submitted (%)
EST	188	**	68	**	**	**20**
IDN	153	14	71	**	**	8
LVA	128	9	70	**	**	**13**
LUX	108	20	62	**	**	**14**
ISL	90	**	62	**	**	**29**

** Information in cell supressed for confidentiality
Source: OECD Pilot Survey of Scientific Authors 2014/15.

Figure A1. Response rates comparison between OECD pilot survey and MORE1



Source: OECD Pilot Survey of Scientific Authors 2014/15; and MORE 1 study (IDEA Consult, 2011).

Table A4. Multinomial logistic regression estimates of response outcome

	Bounce		No click	Click, no response		Partial or full response	
Field: Baseline=Arts and humanities	Coeff.	P-value	(base outcome)	Coeff.	P-value	Coeff	P-value
Business, management and accounting	0.051	0.419		**-0.503**	0.000	**-0.207**	0.000
Chemical engineering	0.089	0.149		**-0.479**	0.000	**-0.491**	0.000
Immunology and microbiology	-0.010	0.872		**-0.318**	0.003	**-0.561**	0.000
Materials science	**0.166**	0.002		**-0.684**	0.000	**-0.735**	0.000
Neuroscience	-0.075	0.231		**-0.257**	0.019	**-0.657**	0.000
Physics and astronomy	**0.258**	0.000		**-0.454**	0.000	**-0.510**	0.000
OA journal	-0.054	0.228		0.143	0.104	**-0.083**	0.053
Author with more than one publication in 2011	**-0.196**	0.000		-0.039	0.548	**-0.050**	0.092
Journal impact: Baseline=no info							
Low tertile	-0.015	0.784		0.058	0.613	0.070	0.191
Middle tertile	**-0.128**	0.030		0.111	0.360	**0.117**	0.040
Higher tertile	**-0.105**	0.073		0.089	0.462	0.085	0.132
Affiliation outside OECD or partner country	**-0.261**	0.000		**0.286**	0.000	**0.215**	0.000
Amount proposed in OA contingent valuation	0.000	0.440		0.000	0.894	0.000	0.870
Intercept	-1.790	0.000		-3.121	0.000	-1.315	0.000
Diagnostics	Value						
Number of observations	60697						
Chi Squared (39)	636.21						
Pseudo R-Squared	0.0062						
Log likelihood	-50718.6						

Source: OECD Pilot Survey of Scientific Authors 2014/15.

Table A5. Probability of survey completion, conditional on being targeted

	Coefficient	Standard error	P-value
Author with more than one publication in 2011	0.0492	0.0344	0.153
Journal impact: Baseline=no info			
Low tertile	0.0120	0.0638	0.851
Middle tertile	**0.1311**	0.0671	0.051
Higher tertile	0.0968	0.0667	0.147
OA journal	**-0.1289**	0.0511	0.012
Strata affiliation country * field fixed effects	**Yes**		0.000
Diagnostics	Value		
Number of strata fixed effects (country * field)	354		
Number of strata omitted for perfect prediction (observations)	7 (109)		
Number of observations	60588		
Log likelihood	-17737.44		
Pseudo R-Squared	0.0628		

Source: OECD Pilot Survey of Scientific Authors 2014/15.

Table A6. Drop out behaviour and characteristics among respondents, by wave

	Completers	Dropout rate	Higher education affiliation			Average open access journal			Average number of publications in 2011		
Wave	(Number)	(%)	Dropout	Compl	All	Dropout	Compl	All	Dropout	Compl	All
1	2,087	**4.3**	0.67	0.69	0.69	0.14	0.12	**0.12**	1.39	1.39	**1.39**
2	2,026	**7.8**	0.68	0.68	0.68	0.17	0.14	**0.14**	1.33	1.42	**1.41**
3	1,600	**55.0**	0.70	0.68	0.69	0.18	0.14	**0.16**	1.37	1.35	**1.36**

	Completers	Dropout rate	Average journal impact (SNIP)			Average number of citations		
Wave	(Number)	(%)	Dropout	Compl	All	Dropout	Compl	All
1	2,087	**4.3**	1.09	1.21	**1.21**	0.84	0.90	**0.90**
2	2,026	**7.8**	1.14	1.12	**1.12**	0.92	0.86	**0.86**
3	1,600	**55.0**	1.03	1.11	**1.07**	0.63	0.81	**0.71**

Source: OECD Pilot Survey of Scientific Authors 2014/15.

Table A7. Changes in sample composition through additional reminders

Selected countries, as a proportion of all respondents in each wave

Country	Wave 1	Wave 2	Wave 3
USA	17.25	13.03	14.75
DEU	6.27	4.81	4.81
ITA	5.22	6.27	5.56
ESP	5.08	4.64	4.5
FRA	4.55	4.49	3.69
JPN	4.07	5.23	4.56
IND	2.54	3.55	4.44
BRA	1.92	2.27	3.94
CHN	1.29	4.39	3.75
…	…	…	…
All countries	100	100	100

Source: OECD Pilot Survey of Scientific Authors 2014/15.

Table A8. Respondent features by country and completion status

Averages, selected countries

Affiliation country	Average journal impact (SNIP)					Average number of author's publications (2011)				
	Dropouts	Completers				Dropouts	Completers			
		All	Wave 1	Wave 2	Wave 3		All	Wave 1	Wave 2	Wave 3
BRA	0.86	0.90	0.81	0.98	0.88	1.25	1.43	1.40	1.37	1.39
CHN	0.97	1.05	0.88	1.05	1.02	1.71	1.52	1.35	1.76	1.57
DEU	1.12	1.26	1.22	1.28	1.19	1.29	1.47	1.34	1.58	1.36
ESP	1.06	1.23	1.26	1.22	1.11	1.25	1.41	1.39	1.49	1.28
FRA	1.08	1.22	1.23	1.21	1.13	1.21	1.34	1.28	1.39	1.26
GBR	1.08	1.25	1.29	1.19	1.14	1.40	1.33	1.36	1.32	1.34
IND	0.82	0.93	0.89	1.00	0.84	1.67	1.79	1.98	1.53	1.76
ITA	1.13	1.21	1.19	1.17	1.22	1.32	1.37	1.34	1.43	1.32
JPN	1.04	1.23	1.30	1.19	1.11	1.53	1.47	1.30	1.66	1.48
RUS	0.58	0.75	0.73	0.70	0.68	1.30	1.48	1.43	1.49	1.38
USA	1.33	1.37	1.45	1.31	1.33	1.34	1.35	1.39	1.30	1.35

Source: OECD Pilot Survey of Scientific Authors 2014/15.

REFERENCES

Archambault E., D. Amyot, P. Deschamps, A. Nicol, F. Provencher, L. Rebout, G. Roberge (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 available at http://science-metrix.com/files/science-metrix/publications/d_1.8_sm_ec_dg-rtd_proportion_oa_1996-2013_v11p.pdf

Bartelsman, E., Hagsten, E. and M. Polder (2013). "Cross-country analysis of ICT impact using firm-level data: Micro Moments Database and Research Infrastructure", accessed from http://www.scb.se/Grupp/OmSCB/Internationellt/Dokument/esslait-mmd-final.pdf

BIS (2013). "Open Access: Economic Analysis of Alternative Options for the UK Science and Research System", UK Department for Business, Innovation and Skills. Accessed from www.gov.uk/government/uploads/system/uploads/attachment_data/file/142814/bis-13-689-open-access-economic-analysis-of-alternative-options-for-the-uk-science-and-research-system.pdf

Björk B.-C., M. Laakso, P. Wlling and P. Paetau (2013), "Anatomy of Green Open Access", *Journal of the American Society for Information Science and Technology*, 65, 2, pp. 237-250 DOI:10.1002/asi.22963.

Björk, B.-C., Solomon D., (2014), "Article processing charges in OA journals –relationship between price and quality", Working paper, April 2014. Accessed from http://www.openaccesspublishing.org/oa12/DOI%2010.1007s11192-015-1556-z.pdf

Björk B.-C., Welling P., Laakso M., Majlender P., Hedlund T. et al. (2010), "Open Access to the scientific journal literature: Situation 2009", PloS ONE 5(6). http://dx.doi.org/10.1371/journal.pone.0011273

Carson, R. (2012). "Contingent Valuation: A Practical Alternative When Prices Aren't Available." *Journal of Economic Perspectives*, 26(4): 27-42.

Dallmeier-Tiessen, S. et al., (2011), "Highlights from the SOAP project survey. What Scientists Think about Open Access Publishing", ARXIV. http://arxiv.org/abs/1101.5260.

DAMVAD (2014), "Sharing and archiving of publicly funded research data", Report to the Research Council of Norway. Accessed through www.forskningsradet.no/en/Newsarticle/Norwegian_researchers_willing_to_share_research_data/1253996461347/p1177315753918.

European Commission (2015), *SHE Figures 2015*, Directorate-General for Research and Innovation. Accessed from https://ec.europa.eu/research/swafs/pdf/pub_gender_equality/she_figures_2015-final.pdf.

Franzoni, C., G. Scellato and P. Stephan (2012), "Foreign-born scientists: mobility patterns for 16 countries", *Nature Biotechnology*, 30, pp. 1250–1253. http://dx.doi.org/10.1038/nbt.2449.

Gargouri Y., C. Hajjem, V. Larivière, Y. Gingras, L. Carr, T. Brody, S. Harnad (2010), "Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research", PlosONE, 5, 10, http://dx.doi.org/10.1371/journal.pone.0013636.

Gargouri Y., V. Larivière, Y. Gingras, L. Carr and S. Harnad (2012), "Green and Gold Open Access Percentages and Growth, by Discipline", Paper presented at 17th International Conference on Science and Technology Indicators (STI), Montreal, Canada. http://arxiv.org/abs/1206.3664.

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", accessed from https://arxiv.org/ftp/arxiv/papers/0906/0906.5418.pdf.

Hardisty D. J. and D. A. F. Haaga, (2008), "Diffusion of treatment research: does open access matter?", *Journal of Clinical Psychology* 67 (7). DOI:10.1002/jclp.20492.

IDEA Consult et al. (2011). Study on mobility patterns and career paths of EU researchers. MORE1 study. Report to the European Commission. Accessed from http://ec.europa.eu/euraxess/pdf/research_policies/MORE_HEI_report_final_version.pdf

Laakso, M. (2014), "Green open access policies of scholarly journal publishers: a study of what, when, and where self-archiving is allowed", *Scientometrics*, 99(2), 475– 494. DOI:10.1007/s11192-013-1205-3

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

McCabe M. and C. M. Snyder (2014), "Identifying the effect of open access on citations using a panel of science journals", *Economic Inquiry*, 42, 4, 1284-1300.

Miguel S. and Z. Chichilla-Rodrígues and Félix 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*, 62 (6), 1130-1145.

Nagaoka, S., Igami, M., Walsh, J., and Ijichi, T. (2011). "Knowledge Creation Process in Science: Key Comparative Findings from the Hitotsubashi-NISTEP Georgia Tech Scientists' Survey in Japan and the US". IIR Working Paper 11-09. Accessed from http://core.ac.uk/download/pdf/6374651.pdf.

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

OECD (2015b), *The Innovation Imperative: Contributing to Productivity, Growth and Well-Being*, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/9789264239814-en.

OECD (2015c). Daejeon Declaration on Science, Technology, and Innovation Policies for the Global and Digital Age. www.oecd.org/sti/daejeon-declaration-2015.htm

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

OECD (2010), *Measuring Innovation: A New Perspective*, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/9789264059474-en.

OECD (2007), *Science, Technology and Innovation Indicators in a Changing World: Responding to Policy Needs*, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/9789264039667-en.

Taylor and Francis (2014). Open Access Survey 2014. Accessed from http://www.tandf.co.uk/journals/explore/open-access-survey-june2014.pdf

UNESCO (2012), Policy Guidelines for the development and promotion of Open Access, UNESCO Publishing. Accessed from http://en.unesco.org/open-access/sites/open-access/files/215863e.pdf.

Zinner, D., Pham-Canter, G., and Campbell, E. (2016), "The Changing Nature of Scientific Sharing and Withholding in Academic Life Sciences Research: Trends From National Surveys in 2000 and 2013", *Acad Med*., 91(3):433-40. DOI: 10.1097/ACM.0000000000001028.

NOTES

1. See for example the analysis to model the costs and benefits of open access conducted by the UK Department for Business, Innovation and Skills (BIS, 2013).
2. See for example the discussion in Bartelsman, Hagsten and Polder (2013).
3. For more information, including methodological guidelines, please visit www.oecd.org/sti/cdh
4. www.researchersmobility.eu and www.more-2.eu
5. See https://hub.wiley.com/community/exchanges/discover/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont?referrer=exchanges
6. For example, we might choose to highlight the institutional endorsements in country A to authors with an affiliation to an institution in that country as indicated in our database or to those responding from such country. Neither solution is ideal.
7. A hard bounce is an email that has been returned by the recipient's mail server as permanently undeliverable. This can be caused by invalid addresses (domain name doesn't exist, typos, changed address, etc.). Soft bounce is an email that gets as far as the recipient's mail server (it recognises the address) but is bounced back undelivered before it gets to the recipient. A soft bounce might occur because the recipient's mailbox is full, the server is down or swamped with messages, or the message is too large. Soft bounces can also include auto-replies to your email.
8. Because of the lack of scientific author population statistics for demographic variables such as gender and age, it was not possible to implement weight adjustments based on such characteristics.
9. Based on personal communication with the responsible official in Statistics Canada.
10. The MORE1 study completion rates are based on a population of substantially cleaned and edited email addresses. In our study, the only editing of email addresses was aimed at ensuring a structure of the form AAA@BBB.{2 or 3digit domain code}. A later editions of the MORE study appears to have attained improved response rates.
11. 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 a publication fee (also called an article processing charge - APC) to the publisher
12. Our results are consistent with those in Gentil-Beccot, Mele and Brooks (2009) who report that free and immediate online dissemination of preprints creates a considerable citation advantage in High-Energy Physics (HEP), whereas publication in open access journals presents no discernible advantage.
13. See for example https://stats.oecd.org/glossary/detail.asp?ID=438 and Carson (2012).
14. Since the amounts are randomly allocated, the econometric results that control for other factors do no yield substantially different results for the response to amounts offered.
15. For example, our question aimed at identifying if supporting data were publicly available included the term digital material to make it more humanities-friendly. This appears to have been interpreted as if comprising online only annexe documents, etc…
16. With additional resources, it might be possible to conduct a confidential, expert assessment of whether these responses support the validity of the survey instrument. There is a possible risk that responses overstate true replicability.
17. In this pilot, it was decided not to ask about episodes of potential scientific misconduct.
18. It is left for future work to attempt to implement a more comprehensive tagging of all sectors.
19. It should be noted that this analysis is based on the reference ASJC field per paper selected at random. A number of journals may be classified simultaneously into Business and Economics, as the ASJC classification is not implemented on a mutually exclusive basis.