



OECD Science, Technology and Industry Working Papers 2020/04

The 2018 OECD International Survey of Scientific Authors

Michela Bello,

Fernando Galindo-Rueda

https://dx.doi.org/10.1787/18d3bf19-en



Note to Delegations:

This document is also available on O.N.E under the reference code: DSTI/STP/NESTI(2019)2/ANN1/FINAL

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the authors. Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcomed, and may be sent to the Directorate of Science, Technology and Innovation, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

This document, as well as any data and 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 Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Note by Turkey:

The information in this document with reference to 'Cyprus' relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the 'Cyprus issue'.

Note by all the European Union Member States of the OECD and the European Union:

The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

© OECD (2020)

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.

Foreword

This paper presents the main results of the second pilot of the OECD International Survey of Scientific Authors (ISSA2), an initiative carried out as part of work of the OECD's Working Party of National Experts in Science and Technology (NESTI) of the Committee for Scientific and Technology Policy (CSTP). The paper was approved and declassified by written procedure by CSTP on 24 January 2020 and prepared for publication by the OECD Secretariat.

The ISSA2 project has benefited from the advice and contributions of the members of the ISSA2 advisory group: Eric Hauet (France), Tomohiro Ijichi and Masatura Igami (Japan), Janet Metcalfe (VITAE), Alexandre Paredes (Portugal), Elisabeth Pastor and Tobias Philipp (Switzerland), Roberto De Pinho (Brazil), Luis Sanz-Menéndez (Spain), Carthage Smith (OECD), Paula Stephan (Georgia State University) and Maria Victoria Tignino (Argentina).

The authors also wish to thank Alessandra Colecchia (OECD), Alice Meadows, Josh Brown and Laurel L. Haak (ORCID), the participants of the OECD Applied Economics Work-in-Progress Seminar, CSTP, Global Science Forum (GSF) and NESTI delegates for comments and advice. Pierpaolo Perez designed and kindly made available the ISSA logo featured on the project webpage and related material.

Within the OECD, Arnaud Atoch, Brunella Boselli, Ríkharður Egilsson, Stefan Plizga, and Samuel Pinto Ribeiro provided support with the online survey implementation. Brigitte Van Beuzekom provided assistance with the bibliometric data. The authors are particularly grateful to all survey participants, including those who contributed to testing the questionnaire, without whom this study would not have been possible. Any errors or omissions are the authors' sole responsibility.

Table of Contents

Foreword	
The 2018 OECD International Survey of Scientific Authors	6
Executive summary	7
1. Introduction	9
2. Study objectives	
3. Survey design and implementation	
4. Survey data processing and adjustments for non-response	
5. The profile of scientific authors in the ISSA study	
6. Using the ISSA2 data	
7. Concluding remarks	
References	
Annex A. Sampling frame and target sample	59
Annex B. Non-response bias analysis	
Annex C. Nonresponse adjustment estimation	
Annex D. Summary descriptive statistics	

Tables

Table 2.1. Key features for testing in the ISSA2 "business case"	16
Table 3.1. Comparison to the population of researchers, selected economies, 2017	20
Table 4.1. ISSA2 survey unit response rate, by authors' country/economy of affiliation	27
Table 4.2. Unweighted and weighted response rate, by science field	29
Table 4.3. Response rate by country/economy of authors' affiliation	31
Table 4.4. Tests for non-response bias, overall sample	33
Table 4.5. Comparison of item respondents and non-respondents, by questionnaire section	34
Table 4.6. Comparison of key item respondents and non-respondents, selected questions	35
Table 4.7. Multinomial logistic regression estimates of response outcome	36
Table 5.1. Scientific research field codes	40
Table 5.2. Women in Higher Education - a comparison of R&D statistics, Scopus and ISSA2	42
Table 5.3. Researchers and corresponding authors compared by sector of employment	48
Table 5.4. Scientific authors' job security by age and occupation	51
Table A A.1. Number of corresponding authors in the sampling frame and target sample, by country/economy of affiliation	59
Table A A.2. Number of corresponding authors in the sampling frame and target sample, by	
coinnon field	61

science field	. 61
Table A B.1. Comparison of unit survey respondents and non-respondents, selected economies	. 62
Table A B.2. Comparison of partial respondents and survey completers, selected economies	. 64
Table A B.3. Comparison of early and late respondents, selected economies	. 66
Table A C.1. Survey nonresponse adjustment estimation	. 68
Table A C.2. Item nonresponse adjustment estimation	. 69
Table A D.1. Summary descriptive statistics of ISSA2 selected variables	. 70

Figures

Figure 3.1. Steps involved in rendering the target population operational for a survey study	. 17
Figure 3.2. Sample selection procedure	. 23
Figure 4.1. (Unweighted) response rate by question item	. 32
Figure 5.1. Female authors, by country/economy of affiliation	. 41
Figure 5.2. Female scientific authors, by field	. 41
Figure 5.3. Authors' average age, by country/economy of affiliation	. 43
Figure 5.4. Authors' average age, by science field	. 44
Figure 5.5. Authors with a doctoral degree, by country/economy of affiliation	. 44
Figure 5.6. Authors with a doctoral degree, by science field	. 45
Figure 5.7. Identification of author's main institutional sector of employment	. 46
Figure 5.8. Authors' employment sector, by country/economy of residence	. 47
Figure 5.9. Scientific authors' employment sector, by science field	. 48
Figure 5.10. Authors' main occupation by science field	. 49
Figure 5.11. Authors' job security, by country/economy of residence	. 50
Figure 5.12. Authors that graduated abroad (vs current residence), by country/economy of	
residence	. 52
Figure 5.13. Authors that graduated abroad (vs current residence), by science field	. 52
Figure 5.14. Proportion of authors that plan to move to another country in the next year, by	
country/economy of residence	. 53
Figure 5.15. Proportion of authors that plan to move to another country in the next year, by science	
field	. 53
Figure 5.16. Authors' gross annual earnings	. 55

Boxes

Box 2.1. Main research hypotheses for the ISSA study on science going digital 12

The 2018 OECD International Survey of Scientific Authors

Michela Bello and Fernando Galindo-Rueda

This technical paper describes the methodology and main features of the second pilot of the OECD International Survey of Scientific Authors (ISSA2). ISSA2 was carried out in 2018 to provide evidence on the nature and effects of digitalization in science. This paper describes the key design and implementation feature of the ISSA2 online survey. It also provides an overview of the data collected focusing on the profiles of the survey respondents in relation to the broader population they represent. This aims to guide thirdparty research use of the data and code made openly available under two different access regimes.

Executive summary

This paper provides a technical description of the 2018 OECD International Survey of Scientific Authors (ISSA2), a global study that gathered statistical evidence on the nature and effects of digitalization in research. The study is part of the broader ISSA initiative set within the Programme of Work and Budget of the OECD Committee for Scientific and Technological Policy (CSTP) to develop and test new measurement instruments aimed at informing the design and assessment of science policies. It targets authors of peer-reviewed publications and is based on evidence collected by a means of an online, global survey directly carried out by the OECD. A first ISSA pilot was implemented in 2015 and focused on identified issues of scientific publishing and open access (Boselli and Galindo-Rueda, 2016). As a response to the increasing use of digital tools in research and its important implications for science policy, the second round of ISSA focused on the new features of science going digital. By doing so, it also contributes to the OECD Going Digital project, a horizontal OECD effort to build a coherent and comprehensive policy approach on the societal and economy-wide process of digital transformation and is in line with the priorities set in the OECD Blue Sky Agenda (OECD, 2018) adopted by the CSTP's Working Party of National Experts in Science and Technology Indicators (NESTI), which has responsibility for the delivery of the project.

This document describes the survey design and implementation processes and provides an overview of the survey data collected. The final ISSA2 dataset contains data on approximately twelve thousand scientific authors of publications released in 2017 and listed in a global bibliographic index and covers all science fields. Data were collected in the last quarter of 2018 through an online and by-invitation only survey, which restricts the study to scientific authors that are designated as corresponding authors. In addition to presenting the project and the survey data, this paper also assesses the representativity of the achieved sample and the potential of the ISSA as a tool to examine different features of scientific work. The following key points emerge from the analysis reported in the document.

- The population of scientific authors and that of researchers differ considerably in terms of sectoral composition. Business researchers are underrepresented in the population of indexed scientific authors.
- The use of an online survey results in differences in response rates by country and other personal characteristics. For instance, ISSA2 respondents seem to have a longer academic history but have a less number of publications than non-respondents and are slightly more likely to be male. The survey's final results are adjusted to take into account such response patterns.
- Survey results suggest that only 30% of corresponding authors are women and this share largely varies across countries. The majority of corresponding authors seem to be between 35 and 55 years old and hold a doctoral degree. Most of them are employed as higher education teaching professionals and report different levels of job security according to their country of residence. Overall, 22% of scientific authors reside in a country that differs from the place where they attained their highest qualification, compared to 29% of authors who reside in a country that differs from the country where they were born.

This paper also highlights a number of takeaways that will be taken into consideration in the development of future rounds of ISSA and include the following.

- The procedure for invitation and authentication needs to be streamlined to overcome the trust / participation gap, and has to be supported by state of the art tools. For this type of population, an alternative approach might involve respondent authentication by means of widely used identifiers such as ORCID. This would, in principle, allow to avoid the use of invasive email campaigns.
- The survey needs to be somewhat shorter and deal with simpler constructs. As already appreciated in other studies of ICT adoption, it can be difficult to secure accurate responses concerning the adoption of advanced digital tools and practices. This technical information may often reside in other team members.
- The initiative would considerably benefit from explicit support or at least endorsement by national authorities. In order to achieve sufficiently precise national estimates, a higher order of magnitude in the number of participants is required.

Overall, there is scope for the OECD to capitalize on its reputation as a trusted provider of statistical data, proximity to policy makers, and engagement with national experts. The bull of methodological efforts has to concentrate on developing the value proposition for scientists as an ultimate information providers.

Anonymised responses to ISSA2 can be downloaded from the OECD dedicated page (<u>http://oe.cs/issa</u>). This page also provides instructions on how to apply for access to additional public use files containing derived indicators and code for data analysis.

1. Introduction

This report describes the main defining and technical features of the OECD International Survey of Scientific Authors (ISSA) conducted in 2018. This pilot initiative –the second of its kind, i.e. ISSA2–aims to collect and report statistical data on the nature and effects of digitalisation within scientific research, contributing to the OECD Going Digital project¹, a horizontal OECD effort to build a coherent and comprehensive policy approach on the societal and economy-wide process of digital transformation. The ISSA2 study, set within the Programme of work of the OECD Committee for Scientific and Technological Policy (CSTP), specifically aims to contribute to the evidence on the digital transformation of scientific research and its broader implications.

The ISSA2 study aims to provide new insights into how digital tools are changing the features and impact of scientific research, help study the responses of the scientific community to these changes, and contribute to the discussion on the measurement of scientific activity and its impact. It is part of the broader ISSA initiative undertaken under the aegis of the CSTP's Working Party of National Experts in Science and Technology Indicators (NESTI) to develop and test new measurement instruments to inform the design and assessment of science policies. Since science policy priorities are under constant change and target a plurality of actors (e.g. researchers, institutions, agencies, publishers, government, citizens), there are growing demands for more timely, granular and comprehensive information. By developing a new, flexible survey instrument, ISSA represents an experimental addition to the set of tools and initiatives deployed by the OECD to address policy user needs and work with data at the micro level. In contrast with other tools that are focused on organisational behaviour and data, this instrument places individuals at the centre of measurement efforts.

As an experimental study, ISSA targets the global scientific research community by means of a dedicated, online survey directly administered by the OECD. This relatively novel – from an OECD perspective – concept was first implemented in early 2015 when a first pilot survey (ISSA1) was carried out with a focus on identified issues of scientific publishing and open access (Boselli and Galindo-Rueda, 2016). This first pilot collected responses from corresponding authors of publications listed in a major global scientific publication index across seven diverse, hand-picked science domains. Nearly 61 thousand scientific authors were targeted, with approximately 10% complete responses collected from them. Anonymised survey results were subsequently made available for use by researchers.²

With some technical differences, ISSA2 builds on a similar approach to the ISSA1 model to develop a statistical view of key aspects of scientific research. While the statistical standards attained by this study may not be at the level attained by surveys conducted by National Statistical Offices (NSOs), ISSA provides a number of insights and facilitates a range of analyses that would otherwise not be available or feasible. ISSA is also intended

¹ See <u>https://www.oecd.org/going-digital/</u>

² Available at <u>http://oe.cd/issa</u>

OECD SCIENCE, TECHNOLOGY AND INDUSTRY WORKING PAPERS

as an open resource for analysis by the research community and a mechanism for dialogue between the OECD and the scientific community.

This technical document describes the key design and implementation processes adopted through the ISSA2 study and provides a basic introduction to the data collected. The paper is structured as follows: Section 2 presents the study objectives, highlighting the key user needs and the survey concept on which the business case is built. Section 3 illustrates the survey design and implementation processes. Section 4 describes the survey data processing procedures and the adjustments made to minimise the potential impact of unit and item non-response bias. The profiles of survey respondents is presented in Section 5, with a focus on gender, age and careers information. Section 7 concludes with a number of methodological considerations for potential future work.

2. Study objectives

User needs

CSTP policy user needs define the objectives of the ISSA study. These needs have been identified through repeated consultation with the CSTP and its working parties, as well as the international expert community. A series of internal presentations and discussions initiated in 2014 secured internal user feedback from the CSTP and two of its working parties, namely the Global Science Forum and the Working Party for Innovation and Technological Policy. Exposure to the international Experts on Science and Technology Indicators (NESTI), which has responsibility for the delivery of the project as well as at the OECD Blue Sky forum held in 2016. Based on those consultations, two major types of user needs were identified as key drivers for ISSA: thematic and instrumental.

Thematic user needs: digitalisation of science

The main factor driving the ISSA2 study corresponds to the policy priority of understanding the key features of the digitalisation of scientific research - how is science going digital and what are the potential implications for policy. The development and adoption of new information and communication technologies (ICTs) has the potential to transform science in multiple ways. Firstly, it can change the ways in which data is collected, generated, accessed, used and analysed, enabling a faster and more extensive use of data in scientific research and making it possible to adopt new research paradigms based on deeper analysis of larger and more complex data. Digitalisation is also thought to be transforming the scientific process in terms of how scientists interact among themselves and beyond the scientific community. Furthermore, while affecting the way research is carried out, digitalisation can also create new opportunities for measuring scientific outputs and its impact, especially through Internet-based data. These include new metrics such as usage-based data and new measures of scientific outputs such as data and software. The use of metrics for research assessment and related purposes in turn transforms incentives at the level of individuals, teams and institutions, possibly in ways that may not have been fully intended or anticipated.

Policy makers have a strong interest in understanding how digital tools are used by scientists and how they are affecting research activity as these changes have important implications for policies aimed to promote open science, assure data quality, stimulate skills development as well as for legal and ethical frameworks, business models and international coordination. The intensive use of data and the larger access to data and knowledge are, for instance, raising concerns on data standards and integrity (e.g. quality), ethics (e.g. privacy), research incentives, and Intellectual Property (IP) rights among others, which have impacts on IP and science policies. Likewise, the use of digital tools and the collection and analysis of a large quantity of data that is becoming more popular in most scientific domains requires scientists and researchers to acquire new skills. This drives policy interest about emerging skill needs and pathways to promote their development.

The key research questions of the study have been informed by a review of the literature and previous OECD committee discussions within CSTP, NESTI and the project's steering group. The main questions relate to the types of digital tools used in science and their impact on research activities; the challenges faced by the scientific community and brought about by digitalisation; the measurement of scientific activity and its impact in the digital era for statistical and administrative purposes. Box 2.1 outlines the main research hypotheses that this project sets out to investigate. It should be noted that when it comes to questions about causes and impacts of digitalisation, it may not always be possible to give a definite causal interpretation to the results of this study. However, the analysis may be able to discard a number of competing hypotheses. A number of questions eliciting in an explicit fashion subjective views on barriers, enablers and impacts as perceived by respondents can also provide additional evidence on causes and impact of digitalisation as perceived by key actors.



• The use of digital tools in research is changing how research output is assessed.

Questions on the enablers of scientific digitalisation

Barriers and enablers of digitalisation can include socio-demographic factors, awareness and trust, skills, technical infrastructures, ethical and legal norms, data quality assurance practices, data sharing policies, and funding and research assessment systems. There are competing hypotheses as to what are the most important enablers or barriers to science going digital:

- Digitalisation requires individuals to have new skill sets.
- Access to infrastructure and tools has become essential in the digital era.
- Trust and legal issues are of increasing importance in the digital era.

User needs for new instruments mapping the state of play of scientific research

In addition to the specific research questions prioritised in the CSTP policy context, most recently the digital transformation of science, there is a plethora of science-related topics that call for new evidence, especially around the nature and orientation of scientific research and its broad ranging impacts. Taking stock of the available sources, the conclusion is an overarching user need for new measurement instruments that are suitable for capturing rapidly evolving features of science across different multiple countries. The fundamental challenge in this context is to have access to a timely source of statistical information on international research policy questions that evolve very rapidly and cannot wait for a complete cycle of traditional distributed data development to be complete. The

quantitative data that informs many of the policy discussions in an OECD context is based on data collected on a distributed fashion, through NSOs, after going through a process of consensus-based standard-setting on what data should be collected and on what basis. This is then normally followed by data collection by NSO, leading to publication and pooling of aggregated, non disclosive results. Such a model underpins data such the OECD R&D and innovation statistics (<u>http://oe.cd/rds</u>) and the indicators on the Careers of Doctorate Holders (<u>http://oe.cd/rds</u>). A complete data cycle can in the best of cases take about 5 years without a guarantee that at the end there will be an internationally comparable set of indicators as countries may not have resources to ultimately collect the data. This stands in contrast with the 2-year programming and budgeting cycle of the OECD through which specific policy questions are first prioritised and then addressed with the available resources.

Existing statistical or administrative data sources are largely unsuited to address a wide range of these policy questions, requiring the design and implementation of ad hoc surveys over a much faster development cycle. In addition to the needs for timely data collection on a multi-country scale on specific topics of interest, there are additional analytical needs that call for data resources to be inter-linkable (widening the possibilities for data analysis) and allow analysis to proceed at the most granular possible level, in order to assess the potential impact of a policy or understand the views or opinions of relevant actors regarding a specific phenomenon.

In recognition of this, participants at the 2016 OECD Blue Sky Forum on the future of data and indicators confirmed the case for surveying and collecting information on key actors in the STI system, highlighting the role of individuals in measurement initiatives and providing a rationale for developing a survey instrument that can also be used for new CSTP priorities (OECD, 2018).

The survey concept and its business case

Based on the examination of topical and instrumental user needs, the OECD put forward the concept of a new survey instrument and proposed that it should be tested by means of a series of pilot studies, the first one of which took place in 2015.

The key features of the OECD ISSA instrument concept are as follows:

- It is a survey targeted to the global scientific research community, which as will be described later is effectively operationalised on the basis of a proxy population, namely that of authors of scientific publications.
- It is a multi-country survey directly implemented by the OECD, therefore reducing the duration of the development and implementation cycle and ensuring that data on respondents in different countries can be pooled and analysed safely in a combined and strictly identical fashion.
- It adopts an online survey tool for data collection to minimise data collection costs.

This survey concept borrows on key features of existing academic-led survey work, in particular the work by Franzoni, Scellato and Stephan (2012), an influential multi-country study of authors of scientific articles published in 2009. A literature search informed the review of different models of researcher and scientist surveying in general and in relation to various aspects of digitalisation. It is beyond the scope of this note to attempt to provide an overview of all those studies, their strengths and weaknesses. In addition to studies such

as Frazoni et al (2012)'s, or Thursby et al. (2018) involving academic partnerships across teams in different countries, several have been driven by scientific publishers³ or scientific associations⁴. Surveys of scientists focus on concrete topics, for example on climate change issues⁵, labour mobility, work conditions and earnings⁶, or questions on access to research outputs, among others. These studies can either provide open mechanisms of participation (through open links to online surveys) or invitation-based (with emails being sent to distribution lists) or combinations of the two (e.g. through snowballing techniques).

Another key reference has been the OECD work on Careers of Doctorate Holders and the basic guidelines developed within that project for surveying highly qualified individuals to collect personal and career data from them.⁷

The ISSA concept itself represents a digitally enabled innovative approach in the way the OECD collects and reports data on science and research, and complements the tools currently available for informing international analysis and comparisons. In order to achieve its objectives, the concept has to prove capable of overcoming to a sufficient degree the drawbacks that are inherent to it. Among these, one should note the limited experience of the OECD in direct data collection. There are potential limits to its ability to reach and motivate respondents across different countries to participate in the exercise and provide robust and accurate answers to the questions asked so that the results can be reliably used for the intended purposes. An experimental pilot approach is therefore an appropriate basis for testing the concept's implementation and its potential long-term sustainability.

The ISSA pilots carried out to date have been accepted and incorporated in the CSTP Programme of Work (2015-16 and 2017-18 respectively, extended into 2019-20), as well as in the OECD Statistical Programme of Work and Budget.

The ISSA2 project oversight was facilitated from its start in 2017 by an expert advisory group comprising a diverse membership with backgrounds in academic research, scientific skills development, STI statistics and policy. The group has met online at key decision points throughout the project. As a first of a kind project at the OECD, the project team requested special advice at the different steps of the project implementation. Given the

³ See, for instance, the survey carried out by Elsevier as part of the EU's Open Science Monitoring Tool in 2016 and 2018. Data sharing and open data practices were explored in the Wiley Open Science Researcher Insights Survey, which collected information on around 4,600 Wiley authors, worldwide, and in the Digital science and Springer Nature Survey on Sharing Data and Open Data, which surveyed around 2,100. Elsevier also conducted in 2018 a large-scale, future-scoping and scenario-planning study including a survey of 2,000 researchers to assess how today's trends might shape the research landscape in the decade ahead (Elsevier, 2019). Nature also carries out a biannual salary survey targeting researchers around the world and covering various topics of science from job satisfaction to mental well-being, pay packages to career progression.

⁴ For example, the 2018 Global Survey of Mathematical, Computing, and Natural Scientists funded by ICSU and conducted by the Statistical Research Center at the American Institute of Physics.

⁵ See <u>https://en.wikipedia.org/wiki/Surveys_of_scientists%27_views_on_climate_change</u>.

⁶ These surveys sometimes morph into professional surveys, for example the Inomics survey on salaries for economists that seeks out to capture professionals working in academia, public or private sector.

⁷ See <u>http://oe.cd/cdh</u>

project's processing of personal data, it also required the approval of the OECD Commission for Computerised Information and Privacy and the OECD's Executive Director. A detailed application was submitted to the Commission on 26 June 2018 for the conduct of an online statistical survey data collection, the subsequent analysis and release of statistical results. The proposal was authorised after having received careful scrutiny of what was considered a very complete application, including a thorough data privacy risk assessment. In accordance with the Decision of the OECD Secretary-General on the *Principles governing computerised personal data processing*, the Commission noted that the purpose of the processing operation was strictly statistical, which implies that the result of data processing is not personal but aggregate or anonymised data and that neither these results nor the personal data can be used in support of measures or decisions regarding any particular individual. It also agreed that the data transferred to the OECD was necessary for the purpose of the processing and took the view that the collected data were adequate, relevant and not excessive for the purposes of the project.

ISSA2 provides a complementary approach to the OECD statistical work in the area of STI as an attempt to overcome the timeliness and resource-driven limits to STI data collections for distribution across NSOs. The business case for ISSA ultimately rests on being able to collect data of reasonable quality for the intended purposes that would otherwise be impossible to secure in the available time and with minimal resource implications for either the OECD, countries and respondents (burden). To achieve that, the OECD needs to attain a basic level of engagement on the part of its target population, especially when the latter has limited knowledge about the role of the OECD. The objectives laid out in Table 2.1 involve a series of trade-offs that need to be assessed on the basis of a concrete, practical experience. OECD-level usability over and above other potential tools and sources is, for instance, a primary consideration. Potential shortcomings in terms of the validity and reliability objectives, which are without a doubt a major requirement of usability, can to some extent be compensated by other dimensions relating to the latter. Ultimately, the ISSA business case rests on the value proposition and incentive compatibility for the respondents, without whose participation the usability and core quality objectives cannot be achieved. These trade-offs are assessed by comparison to other related initiatives. In addition to these considerations, ISSA can, ultimately, be a valuable vehicle for communication between the OECD and the global scientific community.

These multiple objectives are unlikely to be simultaneously achieved on the basis of a single or few one-off pilots, but they require a progressive process of communication that builds on results from previous pilots. The remainder of this documents seeks to document the main technical features of the ISSA2 process and primary data outputs in order to inform a better understanding of the relevant trade-offs between different approaches to collecting new STI data.

Overarching objectives	Attributes	Implications
OECD-level usability and efficiency	Added value	The study provides OECD with outputs with attributes that cannot be secured otherwise
	Multi country coverage, with OECD or broader scope	The survey has an international scope
	Consistency in application of survey instrument	The survey instrument is a applied consistently across countries
	Micro data availability for centralised analysis	Microdata can be used by OECD for analysis without loss of information in aggregation procedures
	Speed of implementation and adaptability	The survey instrument can be developed and implemented in a normal OECD 2 year budget cycle
	Connectedness of survey process and data to available information sources	Data collected can be connected to other sources to reduce response burden and to improve quality (e.g. through sampling)
	Demonstration of use of digital tools for evidence gathering and analysis and their sustainability	The study provides insights to OECD on the effectiveness, efficiency and sustainability of different approaches for collecting and processing evidence in digital environments
Validity and reliability	Representativeness	Results are representative of a relevant population of interest
	Authentication of legitimate respondents	The integrity of the instrument can be ensured through authentication of legitimate respondents
	Ground quality	Measurements are valid and reliable
Incentive compatibility for individual respondents	Intrinsic value of participation	Respondents find it in their interest to participate and provide truthful and well meditated responses
	Privacy and confidentiality protection	Respondents can be reassured about the treatment of their own data
	Relevance and accessibility of research outputs (reports, data and code)	Respondents find the results of the study relevant. The data resources are relevant for the academic study of scientific research
	Institutional support in countries and organisations	National authorities and organisations support the exercise and facilitate it to the extent that it is possible for them to do so
	Understanding of OECD work on science and research policy	Respondents in the scientific community understand better the role and work of OECD in this area

Table 2.1. Key features for testing in the ISSA2 "business case"

3. Survey design and implementation

Key design features of ISSA2

In order to achieve its objectives of reaching and collecting representative data from a global population of scientists through an online survey, it is necessary to operationalise the concept into a workable solution.

Operationalising the target population

In both ISSA1 and ISSA2, the target global population of scientific researchers has been approximated by a statistical frame derived from authorship information in a global bibliographic database. ISSA requests information from authors of scientific publications (articles and conference proceedings) published in 2017 and indexed within a version of Elsevier's Scopus custom database which is used by the OECD under a licence for statistical purposes. As laid out in Figure 3.1, two main steps are involved, namely the use of scientific publishing as proxy for scientific research (1), followed by the use of Scopus data as the approximate source of data on the universe of scientific publishing (2).



Figure 3.1. Steps involved in rendering the target population operational for a survey study

The scientific peer review system and the body of scholarly publications it provides constitute both a guarantee and a window into the scientific research process. It is indeed a convenient source of statistical information as demonstrated by decades of bibliometric study of scientific research publication output. The metadata contained in publications about the documents and their authors support the analysis of multiple dimensions of scientific production and dissemination, but can also provide a basis for identifying and operationalising the target population for a survey study such as ISSA. Like other forms of administrative or commercial data sources, the primary purpose of bibliometric data is not to serve statistical purposes (OECD/Scimago, 2016). While this does not invalidate their relevance for the statistical analysis of science, it is important to keep in mind this key feature and apply a degree of caution when designing and interpreting bibliometric analysis and derived statistical activities such as ISSA.

The focus on the proxy population of scientific publication authors in step 1 reduces considerably the actual population scope compared to the broad population of researchers (which as per the OECD definition includes those in charge of conceiving and carrying out experimental development activities). While the concept of scientific research has not been formally established in guidelines like the Frascati Manual (OECD, 2015), the functional OECD definitions of basic and applied research closely approximate what several bodies define as scientific research, by allusion to the pursuit of knowledge and understanding of the natural and social world following a systematic methodology based on evidence. ⁸ This often includes notions of scientific methodology around evidence, experimentation and/or

⁸ In contrast, the OECD definition of experimental development that is part of the broader R&D concept and definition can in turn be associated to systematic activities associated to the application of such knowledge – noting also that these generate new knowledge about the world and the newly developed artefacts (products and processes).

observation as benchmarks for testing hypotheses, reasoning to establish general rules or conclusions drawn from facts or examples, critical analysis and verification and testing through critical exposure to scrutiny, peer review and assessment. The requirement of output validation through sharing of findings, data and peer review is often presented as an intrinsic element of science and scientific research, but this also lends itself to multiple possible interpretations, as many organisations, large or small, may run internal peer review systems without necessarily opening their outputs to the rest of the world.⁹ Those will be invisible to studies like ours while they may be reported under official statistical surveys.

The reliance on publicly disclosed peer-reviewed outcomes of research efforts, and the existence of specific, context-varying norms, result in implicit selection bias when attempting to extend conclusions from bibliographic databases to the underlying scientific research population. Incentives to publish are more pronounced among individuals in academia or with former or planned academic careers. As a result, this study *cannot* make robust claims about the behaviour of researchers operating in the business sector or in other domains where secrecy of research is of paramount importance or publication is not the prevalent mechanism of disclosure.

The Scopus's journal title list corresponding to the bibliographic database used in the ISSA2 study contains over 36 000 titles in total, including over 22 800 active titles, out of an estimated 80 000 to 300 000 scientific serial publications in existence worldwide (Elsevier, 2017). Providers of scientific bibliographic information intending to be global in scope adopt different procedures to determine an appropriate trade-off between ensuring an exhaustive coverage and the relevance of journal and other periodical titles, subject to resources. In the case of Scopus, a Content Selection & Advisory Board (CSAB) evaluates and determines which peer-reviewed serial titles are accepted into Scopus, and which ones are excluded.¹⁰

Scientific publications in languages other than English are not automatically excluded in Scopus. Titles from all geographical regions are covered, including non-English titles as long as English abstracts can be provided with the articles. According to Elsevier, approximately 22% of titles in Scopus are published in languages other than English (Elsevier, 2019). In terms of titles, a significant majority are based in Northern America and Europe, including those published by Elsevier itself. While this reflects the major loci of scientific production, there are implicit biases in indexing practices, for example the requirement for abstract content to be available in Latin characters, or for the information about the journal to be available in English as demonstration of its "international" scope or

⁹ The definition of R&D in the Frascati Manual requires that the R&D activity, in addition to fulfilling four other criteria, should lend itself to being codified.

¹⁰ According to Elsevier (2017), titles should meet a series of eligibility criteria to be considered for review, namely they should consist of peer-reviewed content; be published on a regular basis (have an ISSN number that has been registered with the International ISSN Centre); be relevant and readable for an international audience (e.g., have references in Roman script and English language abstracts and titles); have a publication ethics and publication malpractice statement. CSAB subject matter experts review titles using both quantitative and qualitative measures, and the selection is partly based on sample documents from the title. The criteria used in the review process are grouped in five main categories: journal policy, content, citedness, regularity and online availability: http://www.elsevier.com/solutions/scopus/content/content-policy-and-selection.

outreach. In practice, solution-oriented indexing practices can be in some cases detrimental to reflecting scientific endeavours with a local/national focus in terms of subject matter and potential application. As a result, any available global index may prove insufficient as a basis for representing the overall scientific enterprise, especially in a number of domains where the knowledge frontier is primarily of a local nature. These are some fundamental issues that lie at the heart of major ongoing debates concerning what should be the minimal requirements for inclusion in a scientific scholarly index and the types of uses, global or country-specific, that bibliometric databases may or may not be ready to support (OECD, 2016).

The OECD secretariat is familiar with the Scopus database, as it uses it under licence for statistical analysis to analyse multiple dimensions of scientific production and dissemination. ISSA2 is conceived to combine the relative strengths of bibliometric and survey data. Bibliometric data provide not only data on collaboration, citation, track publication history, and affiliation but also contains contact information that enables a controlled survey approach.¹¹ Survey questions can therefore focus on capturing data information that is impossible or very difficult to obtain from other sources. The combination of the two sources allows for far richer insights than separate analysis.

The existence of author contact information is critical for the survey implementation. This results in an additional constraint, as highlighted in step 3 in Figure 3.1. In Scopus, contact information is only systematically available for authors that are self-identified as contact authors for corresponding purposes. As a result, the ISSA2 survey has as a reference target *the population of corresponding authors in indexed peer-reviewed journals*. Corresponding authors represent a body of key actors in the STI system that are greatly relevant to a study of digitalisation of science. The implementation of ISSA on a population of corresponding authors does however imply a number of potential limitations, which constrain the ability to extrapolate to the entire population of scientific researchers, particularly in the case of contributors in more junior positions, who are less likely to be listed as co-authors or let alone be named as contacts.

Some diagnostics about coverage

A comparison between scientific author estimates and R&D human resource statistics can help shed light on the coverage of bibliometric studies and surveys like ISSA based on bibliographic data. Comparison with OECD R&D statistics on a country-by-country basis (Table 3.1) show that, measured on a headcount basis, the population of publishing authors is considerably smaller than that of researchers. This confirms the point that only a fraction of researchers (an estimate that excludes technicians and other individuals in research support positions) are active in scientific publishing.

¹¹ As noted before, it is important for data integrity reasons to be able to control who is able to submit information through the survey tool. This cannot be achieved through completely open web questionnaires for example, as the same person that stumbles on an open link can submit a response several times.

This gap shrinks once researchers in the business sector are excluded. These are individuals who are less likely to be involved in basic research.¹² Overall, the number of researchers in the Higher Education and Government sectors (where publication is most common) is considerably more similar to the number of authors in the Scopus population although a gap persists as differences can also be driven by authorship naming/listing patterns, while there may also be measurement error in R&D human resource statistics.¹³

Country	Total number of researchers	Total number of researchers in HE and GOV sectors	Number of unique authors in the Scopus population used	Number of unique corresponding authors in the ISSA2 sampling frame	Ratio of corresponding authors to total researchers	Ratio of corresponding authors to researchers in HE and GOV sectors	Ratio of corresponding authors to total authors in Scopus
USA	n.a.	n.a.	776 320	199 867			0.26
CHN	2 069 650	893 221	964 825	183 621	0.09	0.21	0.19
IND	n.a.	n.a.	181 036	56 894			0.31
DEU	586 030	333 133	180 761	51 087	0.09	0.15	0.28
GBR	510 980	363 894	168 547	50 654	0.10	0.14	0.30
JPN	930 720	363 953	219 189	46 630	0.05	0.13	0.21
FRA	383 843	153 399	130 526	33 139	0.09	0.22	0.25
ITA	185 916	106 929	127 818	31 196	0.17	0.29	0.24
BRA	n.a.	n.a.	132 563	30 511			0.23
RUS	359 793	172 194	91 354	29 073	0.08	0.17	0.32
ESP	218 680	155 545	114 011	28 467	0.13	0.18	0.25
CAN	n.a.	n.a.	97 042	27 386			0.28
AUS	n.a.	n.a.	83 065	25 746			0.31
KOR	482 796	132 610	126 809	25 222	0.05	0.19	0.20
TUR	210 769	139 937	54 849	17 101	0.08	0.12	0.31
NLD	114 589	37 838	54 527	15 558	0.14	0.41	0.29
POL	132 547	82 983	46 169	15 545	0.12	0.19	0.34
SWE	108 761	56 142	35 589	11 427	0.11	0.20	0.32
CHE	70 834	47 027	38 586	10 346	0.15	0.22	0.27
MEX	42 222	32 536	35 003	8 376	0.20	0.26	0.24
IDN	n.a.	n.a.	23 453	8 140			0.35
BEL	73 709	36 410	27 970	7 938	0.11	0.22	0.28
CZE	59 789	35 579	22 960	7 677	0.13	0.22	0.33
DNK	60 782	28 694	23 973	7 311	0.12	0.25	0.30
PRT	85 780	58 862	24 938	7 184	0.08	0.12	0.29
NOR	54 601	31 888	18 564	6 295	0.12	0.20	0.34
AUT	78 051	40 446	22 005	6 260	0.08	0.15	0.28

Table 3.1. Comparison to the population of researchers, selected economies, 2017

¹² Unlike R&D expenditure figures, conventionally reported statistics do not allow to identify human resources devoted to R&D broken down between types of research and development. On average, less than 10% of business R&D is dedicated to basic research. Arora et al (2015) have documented a marked decline in publications by company scientists over time across a range of different industries, associated to a decline in the financial incentives for companies to invest in basic research relative to investment in more market oriented R&D and innovation activities.

¹³ In economies like China, Italy and the Netherlands, the number of HE and GOV researchers is lower than that of publishing authors.

THE 2018 OECD INTERNATIONAL SURVEY OF SCIENTIFIC AUTHORS | 21

Country	Total number of researchers	Total number of researchers in HE and GOV sectors	Number of unique authors in the Scopus population used	Number of unique corresponding authors in the ISSA2 sampling frame	Ratio of corresponding authors to total researchers	Ratio of corresponding authors to researchers in HE and GOV sectors	Ratio of corresponding authors to total authors in Scopus
ISR	n.a.	n.a.	21 163	6 148			0.29
ZAF	51 877	45 284	16 489	6 128	0.12	0.14	0.37
FIN	53 752	26 104	18 729	6 060	0.11	0.23	0.32
GRC	60 736	53 221	18 194	4 891	0.08	0.09	0.27
NZL	31 000	21 700	12 581	4 118	0.13	0.19	0.33
CHL	14 200	9 848	13 541	3 676	0.26	0.37	0.27
IRL	30 612	16 438	12 303	3 618	0.12	0.22	0.29
COL	n.a.	n.a.	12 550	3 610			0.29
HUN	38 915	22 189	12 240	3 321	0.09	0.15	0.27
SVK	26 720	22 284	7 904	2 712	0.10	0.12	0.34
SVN	11 282	5 734	5 047	1 791	0.16	0.31	0.35
LTU	17 746	14 910	3 892	1 210	0.07	0.08	0.31
EST	6 845	4 942	2 719	914	0.13	0.18	0.34
LVA	7 400	6 302	1 957	661	0.09	0.10	0.34
LUX	3 134	1 968	1 121	412	0.13	0.21	0.37
ISL	3 935	2 513	1 174	304	0.08	0.12	0.26

Note: Country refers to the country/economy of affiliation for the figures relating to the authors. As an author can have more than one document and each document can report a different country of affiliation, the number of authors per country is calculated as the sum of the weights of the authors affiliated to an institution in the specific country. An author's weight is calculated as one divided by his/her total number of documents. *HE* and *GOV* stand for Higher Education and Government, respectively. Both researchers and authors are in head counts. Data on researchers refer to the latest year available.

Source: OECD calculations based on Scopus Custom Data, Elsevier, Version 2018.1; OECD International Survey of Scientific Authors (ISSA), 2018. <u>http://oe.cd/issa</u>. OECD, Research and Development Statistics Database, <u>http://oe.cd/rds. June 2019</u>

The difference between the number of total authors and the number of corresponding authors highlights marked differences in the ratio of corresponding authors to total authors per country.¹⁴ In the case of international collaborations, the affiliations of authors other than the contact person will be removed from the reference population frame for that particular publication. Some individuals will disappear entirely from the database if they never feature as corresponding authors, an effect that will vary across fields. For example, in fields where publication requires higher levels of equipment and other resources, or teams tend to be larger, it will be more likely for junior research staff to be excluded from the population of corresponding authors.

On average, there is a ratio close to 20% for the number of corresponding authors in one country on the one hand relative to the corresponding numbers of HE and GOV researchers. In conclusion, the ISSA study, as currently implemented, should be considered as a study about science *as reported by* individuals in a position of some responsibility with respect to indexed scientific publication output. For this reason, and unlike other studies, no attempt has been made to provide an extrapolation of survey estimates to national totals in terms of researchers.

¹⁴ See also the leadership indicators in OECD/Scimago (2016).

OECD SCIENCE, TECHNOLOGY AND INDUSTRY WORKING PAPERS

Sample selection

As already noted, the survey targets corresponding authors of scientific publications. The focus is on Scopus-indexed articles and conference proceedings published in 2017. The rationale for excluding reviews is that many of the questions of interest in the study do not apply to the subject matter of reviews. Other indexed documents are excluded on the basis of their low representativity and lack of indexed citations contained within the database. Such documents are customarily also excluded from the computation of scientific bibliometric indicators relying on citation analysis.

The underlying data extracted from Scopus has a network structure comprising two types of nodes, namely corresponding authors and documents. Edges link authors to documents by indicating whether an individual is the corresponding author of a given document. Personal information such as affiliations and contact details are available at the level of edges (and not author nodes) because the information is specific to a given document and may thus vary across publications for any given author.

The sample was selected based on a stratified proportional (with quota) sampling methodology. Figure 3.2 depicts the various steps of the sample selection process. After the initial extraction of the documents from Scopus, those whose corresponding authors had a valid Scopus author ID were retained (step 2). The survey frame contains around 1.7 million author-document pairs. Because any author can have more than one single affiliation, the location indicated for correspondence purposes was selected as the author's main affiliation (step 3).

Strata were defined along the combination of two dimensions: the territory (country / economy) of the document lead author's affiliation and the broad subject areas corresponding to the journal's main field of science (All Science Journal Classification Codes, ASJC)¹⁵. A dominant subject area is identified per document based on the different 4-digit fields that can be identified simultaneously.¹⁶ As defined, document-author pairs are uniquely ascribed to strata, but this is not the case for authors, who can present in multiple strata at the same time through publications featuring different territory of affiliation or subject area. No territory of affiliation or science field was excluded from this survey pilot's frame. However, strata were merged into one single "Other" category for small non-OECD countries (in terms of number of documents). For very small OECD and EU countries, a single stratum was used per country instead of multiple sub-strata for different subject areas. Documents without contact information¹⁷ were removed from the population frame (step 4), taking into account their incidence within strata.

¹⁵ The ASJC subject areas mapping to broad science domains used for stratification purposes are: health sciences, life sciences, physical sciences, and social sciences & humanities. See https://service.elsevier.com/app/answers/detail/a_id/15181/supporthub/scopus/

¹⁶ In case of draws, one subject area was picked up at random.

¹⁷ This includes document where the contact information does not appear to apply univocally to an actual author. The Scopus has been partly cleaned for this exercise for instances where identical emails apply to different author IDs.



In order to select the target sample, a sequential process was adopted: 1) pre-selecting a target sample of documents for the survey based on the stratification; 2) selecting one document among the set of documents pre-selected that correspond to the same corresponding author. A proportional allocation formula was defined with a minimum size threshold, taking into account expected high non-response rates (step 5), and planning for targeting a reference sample size of approximately 150,000 authors. Documents were selected through random sampling within each stratum (step 6), leading to a set of 170 914 pre-selected documents that correspond to 158 488 authors (step 6). Because any given author can only be contacted on the basis of a single document, an additional step is required to identify which of the pre-selected documents is to be retained for the survey. As a result, one single document was ultimately selected per author (Step 7 in the figure), so the number of documents is reduced to the same number of authors (158 488). Annex tables Table A A.1 and Table A A.2 provide further details on sampling figures by territory and subject area.

In order to take into account an author's and a document's probabilities to be selected in the survey data analysis, survey selection weights were computed. The latter take into account a unit's selection probability at the two stages of the sampling process: the authors' selection in the stratification process (document pre-selection) and the random selection of one single document per author – as well as the availability of contact information per document. Given the two-layered nature of the survey sampling frame, where units can be

scientific documents or authors, survey results can either be applied to the population of scientific authors or to that of scientific publications depending on the intended analysis.

Two distinct types of selection weights are computed. One reflects an author's probability of being included in the sample, which is based on the probability that at least one of his/her documents is pre-selected for sampling (in one or more strata). The document level selection probabilities also take into account the probability that a particular document among an author's set of pre-selected documents is ultimately selected. The probabilities coincide for authors of single documents.

The probability of selecting a document \mathbf{D} into the ISSA2 sample is therefore the unconditional probability of the first event in the sequence (document retained based on available contact information) times the conditional probabilities of each subsequent event given the outcome of the prior events in the sequence:

$$P_i^D = P_i^{D,0} * P_i^{D,1|0} * P_i^{D,2|01}$$

where:

 $P_i^{D,0}$ = probability of pre-selecting the *i*-th document in Stage 0 (probability that a document can be used for sampling, based on availability of contact information)

 $P_i^{D,1\setminus 0}$ = conditional probability of selecting the *i*-th document in Stage 1 (i.e. preselecting), given its availability of contact information in Stage 0

 $P_i^{D,2\setminus 01}$ = conditional probability of selecting the *i*-th document in Stage 2 given its selection in Stage 0 and 1.

An author is selected into the ISSA2 target sample if at least one of his/her documents is preselected in *Stage 1*, conditional on having contact information. Hence, the probability of sampling an author **A**:

$$P_j^A = P_j^{A,0} * P_j^{A,1\setminus 0}$$

where:

 $P_i^{A,0}$ = probability of selecting author *j* in *Stage* 0

 $P_j^{A,1\setminus 0}$ = conditional probability of selecting author *j* in *Stage 1*, given its selection in *Stage 0*, which is equal to 1 minus the probability that none of his/her documents is selected.

For simplicity, the sampling weights have been calculated as if sampling effectively occurred with replacement, which is not the case in this large but finite population. This can be a reasonable approximation in case of large strata, as for a majority of cases. However, when it comes to author-level probabilities for smaller strata, failing to adjust for the lack of replacement can have a relatively larger impact on the weights.¹⁸ The calculation of full weights is, as described in the next section, expanded to take into account non-response patterns.

¹⁸ Further information on the calculation of weights is available from the authors on request.

Questionnaire design, testing and launch of the survey campaign

Online survey tool selection

Data collection was carried out through the implementation of an online survey. After reviewing several possible options, it was decided to use the OECD Checkbox online survey tool to implement the ISSA2 survey. The main reason to use this internally deployed tool was the requirement to have data uniquely stored on and delivered from OECD's own servers, ensuring high levels of data confidentiality and security. In addition, its use sought to enable the sending of invitations from an OECD e-mail address with an expected positive impact on the survey email open rate and response rate. Furthermore, the specific formulation of the questions was reviewed on several instances after examining usability in different types of devices. The choice of this tool turned out to impose a number of limitations that were not initially foreseen, as explained below when discussing the launch of the survey.

Questionnaire

The ISSA2 questionnaire (available at <u>http://oe.cd/qissa2</u>) was developed to comprise six main modules, namely: (a) basic information, (b) access to scientific outputs, (c) use of digital tools and research methods, (d) digital identity, outputs and assessment, (e) views on recent trends in science, and (f) career-related information. The concepts and items included in the questionnaire were chosen based on their relevance to the study's policy research questions, measurement properties, comparability across groups and countries, and comparability with other similar surveys and relevant studies. Members of the project advisory group, composed by OECD delegates and other experts in the field of science, also contributed to the questionnaire and survey design. A completion time of 15 minutes was aimed for although reports by respondents revealed this figure to be closer to 20 minutes.

In order to increase the response rates and reduce the risk that the results might turn be overly skewed towards English speaking authors, the survey was made available for completion in ten different languages. In addition to the English original version, Chinese Mandarin, French, German, Korean, Italian, Japanese, Russian, Portuguese and Spanish were also supported in ISSA2, relying in part on the directorate translation quota as well as volunteer input from colleagues and members of the steering group.

Testing

The survey questionnaire was initially tested in May 2018 with five authors from different countries and science fields, including economics, biology, electronic engineering and medicine. These volunteers were asked to complete the survey using the online tool and provide comments on their experience. Feedback on the questionnaire, including the survey interface, was also received from members of the project steering group, OECD delegates, as well as a number of stakeholders who were invited to share their views.

Invitation approach

Targeted respondents were to be contacted via an email automatically generated by the online survey tool. The invitation email explained the objectives and relevance of the project. It also sought to reassure recipients about the strictly confidential treatment of their responses providing a link to the project privacy statement, which is available online on the

OECD dedicated page on ISSA, and the OECD email and contact information. To reassure email recipients about the authenticity of the survey and increase their awareness of the OECD activities in the field, the invitation also pointed to the project's webpage, which provides information on the first pilot of ISSA in addition to links to other relevant work.

The cover email invitation was designed to be submitted only in English but with an option for respondents to change the language with a link to the online platform. The reason for not undertaking a full language customization of the invitation email was the lack of information about the preferred languages and the concern that the affiliation implied language might not be actually relevant to the person concerned. The cover email included a personalised reference to the title of the document selected as the first section of the survey asked questions about that specific publication.

Survey launch

An initial batch of invitations was dispatched a few weeks prior to the planned full launch of the survey at the end of September 2018. Around 7 500 authors were contacted and asked to complete the survey on that occasion. Based on the pilot batch experience, a number of survey design features were revised in order to increase click and response rates, and mitigate the risk that survey invitation emails are blocked or marked as SPAM by the email recipients. This involved sending the survey invitations in small batches to minimise throttling and SPAM alerts and performing additional cleaning of the email addresses to reduce the rate of bounced emails.¹⁹ The text in the invitation email was also revised in order to enhance their attractiveness. A few weeks after the survey pilot, the survey was manually launched in multiple batches created to match suitable response slots in the average time zone for the various countries and territories.

Right after the survey launch, a number of targeted respondents contacted the OECD through the generic contact addresses provided within the survey and the cover email. In addition to verifying the source, targeted authors also had other types of questions and comments. Common issues raised included whether the survey was relevant after the targeted respondent's retirement. A few target respondents did not fully recognise themselves as scientific researchers, highlighting the existence of indexed articles and conference proceedings of a review-like nature. In a number of instances, respondents also shared their opinion about the subject matter of the survey, providing additional and important insights into the study's topic and even sharing their original own material on the subject.

¹⁹ The OECD corporate survey tool turned out to lack core features required to manage the flow of email participation requests and track the status of such invitations through reminders. Only after the launch it became apparent that, owing to security measures, the OECD set up of Checkbox did not incorporate the integrated entire package for managing survey invitation emails and tracking responses. Due to the negative effect of the bulk invitation emails on OECD email servers for regular purposes such as management of visitor request confirmations, a number of *ad hoc* and time-consuming procedures had to be implemented by the OECD IT support team and the survey project team in STI. The OECD is currently in the process of adopting a new online survey tool and set up that can cater towards different types of data collection efforts while providing the necessary stability and security to other Internet-based information exchange operations.

Three reminders spaced out in time were sent to the recipients that did not complete the survey shortly following the first request. Given the observed low response rate in the People's Republic of China (hereafter 'China'), reminders sent to authors based in this country included both an English and a Chinese text. In order to reduce the risk that the survey email was blocked by the recipient's email server, the last reminder did not contain the survey link²⁰, but it pointed to the previous survey invitation emails or invited the email recipients to contact the OECD to request the link.

The survey was effectively closed in December 2018, significantly later than originally anticipated, owing to the need to slow down the flow of submitted emails. The final, collected sample included fully and partially complete responses from 11,963 respondents. Response patterns and subsequent adjustments are described in the following section.

4. Survey data processing and adjustments for non-response

Response patterns

Response rates are commonly used indicators of the effectiveness of a survey instrument to reach its intended target population, one of the key drivers of data quality. The results from ISSA2 are reflective of the general response challenge faced by online surveys. Unit survey response rates are reported on the basis of territory (Table 4.1) and field of science (Table 4.2), both on an unweighted and weighted basis.²¹

Economy	Unweighted response rate (%)	Weighted response rate (%)
CHN	0.70	0.74
KOR	4.47	5.22
USA	5.01	5.41
SGP	5.77	5.37
RUS	7.25	7.41
FRA	7.42	7.56
AUS	7.61	7.72
EGY	7.80	6.48
CAN	7.82	8.14
GBR	7.84	8.10
POL	8.20	7.55
IND	8.61	9.40

Table 4.1. ISSA2 survey unit response rate, by authors' country/economy of affiliation

²⁰ The presence of hyperlinks is a known trigger factor for anti-spam software and related defence mechanisms. The rapid accumulation of several emails sent to a same institution or email client is another trigger.

²¹ The weighted estimates are reflective of the sampling and response probabilities. The basis on which the weighted estimates are calculated is described towards the end of this section.

28 | THE 2018 OECD INTERNATIONAL SURVEY OF SCIENTIFIC AUTHORS

Economy	Unweighted response rate (%)	Weighted response rate (%)
DEU	8.63	8.98
JPN	8.82	9.13
THA	8.95	8.81
PAK	9.41	8.18
MYS	9.60	8.80
NLD	9.61	10.28
BEL	9.66	9.78
DNK	9.84	9.75
CZE	9.88	9.42
ESP	10.45	10.99
SWE	10.62	10.48
ITA	10.80	11.39
TUR	10.89	11.11
BRA	10.89	11.69
IDN	10.91	10.05
CHE	11.12	11.02
ISR	11.41	11.51
Other economies	11.53	10.84
FIN	12.13	12.05
ROU	13.57	12.20
GRC	14.48	13.09
AUT	14.72	13.80
MEX	15.71	14.95
PRT	21.35	21.37
Total	7.55	7.50
Total excluding China	8.94	8.80

Note: Unweighted response rate is the total number of authors that have started, but not necessarily completed, the survey divided by the total number of authors in the target sample. W*eighted response rate* is calculated by dividing the weighted number of authors that have started the survey and provided at least a partial response by the weighted number of authors in the target sample. Sampling weights are applied in the calculation of the weighted response rate. Both unweighted and weighted response rates are expressed in the form of a percentage. *Source:* OECD International Survey of Scientific Authors (ISSA), 2018. <u>http://oe.cd/issa</u>.

Overall, these admittedly low rates are in line with the literature and recent trends pointing to a worsening of response rates in voluntary online surveys (Beullens et al., 2018). They lie towards the lower range of previous purely academic exercises such as Thursby et al (2018) (12% in 2014) but compare favourably with more recent studies such as that carried out by Elsevier (2019), with academic partners, as part of the EU sponsored European Open Science monitor (2.5% for EU member states and G8 countries). Survey response rates are highest in the cases of Portugal (21%), Mexico (15%) and Austria (15%), whereas it is worth noticing in particular the low response rates in China (0.7%) and Korea (5%). Challenges related to conducting online surveys in China are well documented (Franzoni et al. 2012; Boselli and Galindo-Rueda, 2016), in particular in relation to the so called "Great FireWall", which restricts Chinese access to the Internet outside China and makes it virtually impossible to use non indigenous platforms for online data collection (Mei and Brown, 2018).

Science field	Unweighted response rate (%)	Weighted response rate (%)
Chemistry	4.61	5.09
Chemical Engineering	4.62	5.07
Dentistry	4.74	5.15
Biochemistry, Genetics and Molecular Biology	4.76	4.45
Materials Science	4.90	5.71
Medicine	5.04	4.86
Immunology and Microbiology	5.17	5.03
Pharmacology, Toxicology and Pharmaceutics	5.86	5.33
Engineering	5.89	6.30
Neuroscience	6.04	5.73
Health Professions	6.38	6.44
Nursing	6.91	6.82
General	6.97	7.36
Physics and Astronomy	7.04	7.65
Computer Science	7.58	7.94
Mathematics	7.75	7.82
Earth and Planetary Sciences	7.79	8.16
Veterinary	8.16	8.00
Energy	8.44	8.82
Environmental Science	8.76	9.15
Agricultural and Biological Sciences	8.95	8.79
Psychology	9.64	9.27
Decision Sciences	11.64	10.44
Arts and Humanities	13.41	12.05
Social Sciences	15.86	14.52
Business, Management and Accounting	16.18	15.21
Economics, Econometrics and Finance	18.71	17.13
Total	7.55	7.50

Table 4.2. Unweighted and weighted response rate, by science field

Note: See details in Table 4.1.

Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Response patterns also differ by field (Table 4.2). Unweighted response rates are highest in the fields in the subject areas of social science and the humanities, with rates between 13 to 19%. Participation was the lowest in chemical engineering (5%), chemistry (5%), biochemistry, genetics and molecular biology (5%), material science (5%) and dentistry (5%).

In order to better understand the patterns of nonresponse, Table 4.3 reports survey more detailed patterns for selected countries. Responses are divided in partial and complete (fully submitted responses). The former are answers from authors that have partially but not fully completed the survey, dropping out before reaching the last question and submitting the response. About 24% of authors that started the survey did not complete it and submit their response. This drop out behaviour and item non response will be analysed further below.

Unit non-responses are further categorised into bounces, no clicks, and clicked nonresponses. Bounces refer to undelivered invitations due to incorrect email address, retirement, mobility or death where the email does not make it through the recipient server and the OECD server gets notified of such status. Bounce rates on this survey are surprisingly low in comparison with the experience of the previous ISSA study, although this appears to be related to limitations in the survey tool (limiting for example the possibility to identify bounced emails due to wrong addresses). The no-click category is the largest by far, accounting for 90% of the outcomes. This is a hybrid category on which the survey tool does not provide further insights, making it impossible to discriminate between instances in which the target respondent has not been aware of the email and those in which he or she has refused to view and read the email. Several factors can contribute to the former. Spam filtering is the prime suspected reason, as phishing attacks increase in intensity and sophistication in recent years has driven a strengthening of protective measures on the part of email server administrators. Furthermore, target respondents with personal or generic email accounts provided by large ISPs are more likely to see invitation emails withheld or delayed, or simply placed in different folders or tabs related to promotions.

Several out-of-office-messages were received informing circumstances that prevented response, such as retirement, institutional mobility and paternity or other forms of leave. This information was not processed due to limited resources so these instances count as nonresponses. Author mobility or change of personal circumstances can be important aspects that should be taken into consideration in any study of this kind for it is a potential source of systematic non-response bias. A mobile author is likely to have characteristics that may distinguish them from the rest of the population with respect to the phenomenon under study.

Despite the assurances provided, individual caution about the provenance of the email is another major consideration behind nonresponse rates. Unsolicited emails are naturally treated with caution, as reflected by the fact that several target respondents followed up by emailing the OECD team for confirmation about the provenance of the survey invitation.²² The high frequency with which this occurred²³ suggests that the associated transaction costs were a significant factor preventing a larger response rate, as many researchers potentially interested in participation might not have had the sufficient motivation to request direct confirmation about the survey's provenance. A non-negligible number of respondents wrote after the survey had closed, once they had inspected their SPAM folders. Out of all invitations sent, about 1.6% of target respondents clicked and opened the survey but did not provide any response. This accounts for approximately 10% of all individuals who clicked into the system to find out more about the exercise.

Overall, response rates across countries appear to have been highly influenced by the Internet security and protection maturity factors, as well as knowledge of OECD activity as an economic and development co-operation organization. This explains why response rates are also higher in the social sciences, as well as in countries recently engaged in activities such as those related to OECD accession, reviews of national science systems (Austria and Portugal), among others.

²² The invitation email and reminders highlighted clearly the importance for survey respondents to assure themselves about the safety of opening links.

²³ In all these instances, it was possible to quickly corroborate that the survey has indeed generated by the OECD and sent to the relevant person. This process however imposed a significant burden on both the OECD secretariat and the survey participants.

Country	Targeted authors (Number)	Bounces (%)	No click (%)	Click, non- response (%)	Partial response (%)	Response submitted (%)	Total number of responses
CHN	26 837	0.50	98.50	0.31	0.27	0.42	187
USA	24 052	0.61	93.63	0.76	1.12	3.88	1 204
IND	6 972	0.72	88.63	2.07	3.89	4.72	600
DEU	6 214	1.01	88.75	1.61	1.54	7.08	536
GBR	5 954	0.47	90.73	0.97	1.70	6.15	467
JPN	5 647	1.08	87.48	2.62	1.31	7.51	498
ITA	4 150	0.36	87.64	1.20	2.10	8.70	448
KOR	4 047	0.59	93.20	1.75	1.43	3.04	181
FRA	3 907	1.20	89.89	1.48	1.38	6.04	290
BRA	3 589	0.61	85.82	2.67	2.34	8.55	391
ESP	3 549	0.42	87.07	2.06	2.14	8.31	371
CAN	3 543	0.56	90.38	1.27	1.55	6.27	277
RUS	3 478	0.40	87.95	4.40	1.64	5.61	252
AUS	3 312	0.54	91.36	0.48	1.18	6.43	252
TUR	2 167	0.55	85.93	2.63	4.34	6.55	236
POL	2 086	0.34	89.31	2.16	1.68	6.52	171
NLD	1 883	0.64	88.85	0.90	1.81	7.81	181
MYS	1 584	1.33	86.36	2.71	2.97	6.63	152
SWE	1 422	0.28	87.97	1.13	1.55	9.07	151
CHE	1 385	1.23	85.70	1.95	1.81	9.31	154
IDN	1 347	0.30	85.00	3.79	5.12	5.79	147
CZE	1 255	0.40	87.25	2.47	2.95	6.93	124
MEX	1 197	0.84	79.95	3.51	2.42	13.28	188
PRT	1 157	0.69	76.06	1.90	4.06	17.29	247
BEL	1 118	0.45	88.55	1.34	1.79	7.87	108
ROU	1 105	1.00	83.08	2.35	3.17	10.41	150
AUT	1 087	0.92	82.61	1.75	2.48	12.24	160
DNK	1 077	0.56	88.21	1.39	2.32	7.52	106
SGP	1 058	0.66	92.82	0.76	1.13	4.63	61
EGY	1 025	0.88	88.88	2.44	2.93	4.88	80
PAK	1 020	0.49	87.25	2.84	4.80	4.61	96
GRC	1 008	0.69	82.94	1.88	2.98	11.51	146
ISR	1 008	0.79	85.02	2.78	3.27	8.13	115
FIN	1 006	0.50	85.69	1.69	1.69	10.44	122
THA	1 006	0.40	88.07	2.58	2.39	6.56	90
Other countries	26 236	0.15	10.18	0.38	0.20	0.26	3 024
Total	158 488	0.73	90.11	1.62	1.87	5.67	11 963

Table 4.3. Response rate by country/economy of authors' affiliation

Unweighted results, selected economies

Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

In order to better understand the extent of partial survey completion rate and its potential impact on the data analysis, Figure 4.1 reports the response rates by question item and sections so that it is possible to follow at which points respondents are more likely to drop out. The largest fall in the item response rate (from 100 to 80%) occurs at the end of the second section and the start of the third section of the questionnaire which includes questions on the use of digital tools.

The item response rate decreased further in the subsequent section relating to research methods and reached a response rate of around 70%. Beyond this point, item response rates did not change further suggesting that respondents that completed the first half of the questionnaire tended to finalise the survey.

An analysis of the response rate by question item also provides important insights into the complexity of the questions. Our results suggest that respondents may have encountered some problems in understanding and/or answering two core questions of the survey: one relating to the features of data and code dissemination (Q12) and one relating to the use of more advanced digital tools or methods (e.g. big data analytics) (Q16). Item response rates for these questions are 72% and 68%, respectively.





Sample of survey respondents, adjusted for question filters and condition

Note: Questions 8 to 13 were asked exclusively to a selected group of respondents depending on their answers to previous questions. Such conditions are taken into account in the calculation of the response rates to these questions.

Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa

Assessment of potential non-response bias

Tests for non-response bias

More than low absolute response rates, bias is the primary concern for this type of study. This may occurs when respondents to a survey differ systematically from the ensemble of those that were invited to participate but did not participate. A common approach to assessing the potential presence and degree of non-response bias is to compare the characteristics of respondents to those of non-respondents. To do so, we consider a series of author characteristics, obtained from the Scopus database, that are known for both respondents and non-respondents: the number of publications listed in Scopus in the period 1996-2017; the average, field-based normalised number of citations received per year by the author's publications listed in Scopus over the period 1996-2017; the number of publications in open access (OA) journals listed in Scopus in the period 1996-2017; and an

indicator of whether the survey's underlying publication is related to artificial intelligence (AI).²⁴ Tests for equality of means are performed for each pair by country²⁵ and for the overall sample groups. Mean differences for the overall sample are reported in Table 4.4, whereas figures on a country-by-country basis are reported in Table A B.1.

We find significant differences between respondents and non-respondents for the overall sample of authors. Unit survey respondents tend to have slightly lower citations per document and exhibit lower cumulative publication numbers. They are more likely to publish in open access journals. T-test results also suggest that authors of AI-related publications were less likely to answer the survey than other authors, in contrast to the expectation that authors that are more familiar with the topic of the survey should be interested in participating. This suggests that the effective sample appears to exhibit a slight bias towards under-representing most experienced and successful authors. In addition to potential composition effects, for example those related to the Internet security aspects correlated with these measures, this might be due to these individuals being on average busier and less available to engage in the survey.

Table 4.4. Tests for non-response bias, overall sample

Comparison groups: Null hypothesis mean A-mean B=0 Variable used for comparison	Non-respondents versus respondents	Providers of incomplete responses versus complete responses	Early respondents versus late respondents
Author's average of field-normalised citations per	0.085***	-0.059	0.044
document (1996-2017)	(0.000)	(0.093)	(0.215)
Authors' number of publications (1996-2017)	10.08***	1.634	1.634
	(0.000)	(0.157)	(0.157)
Al-related reference document	0.007***	0.005	-0.001
	(0.000)	(0.085)	(0.775)
Author's share of publications in OA journals	-0.014***	0.007	0.003
(1996-2017)	(0.000)	(0.167)	(0.673)

T-tests for equality of average characteristics of different groups in the target sample

Note: P-values in parentheses. * refers to p<0.05; ** refers to p<0.01; *** refers to p<0.001. A document's relatedness to AI was derived through a text mining exercise of keywords in scientific publications based on Scopus data (see OECD, 2019).

How to read: Non-respondents have a higher average rate of field-normalised citations per document than survey unit respondents (8.5% more). The are no significant differences among respondents between those completing the full survey and those providing partial responses.

Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Survey response bias can also be analysed by comparing the characteristics of authors that started the survey and dropped out before reaching the last question and those that finished the survey for systematic differences between these two groups of respondents may also

²⁴ This indicator is based on OECD semantic analysis of the content of abstracts for Scopus-indexed publications (OECD, 2019).

²⁵ Mean differences by country are not studied in terms of AI-related publications given the low number of AI-related publications in each group.

affect the survey results.²⁶ The results show no significant differences between these two groups in terms of citations, AI-related publications, and publications in OA journals for the whole sample. The same lack of systematic differences applies for comparisons of early and late respondents most likely prompted by reminders.

In order to shed further light on nonresponse bias and its possible implications for data analysis, we compare the characteristics of the authors according to whether they completed specific sections of the questionnaire. The results provided in Table 4.5.

As was highlighted previously, authors with a higher number of publications are more likely not to drop out of the survey and complete subsequent sections of the survey. Contrarily to our expectations, authors of AI-related publications do not seem to be more likely to respond to the core sections of the questionnaire on the use of digital tools. In contrast, they seem to be less likely to complete the section on digital identity and views of the phenomenon of digitalisation.

Table 4.5. Comparison of item respondents and non-respondents, by questionnaire section

Comparison groups: Null hypothesis: mean A (non resp)-mean B (resp)=0 Variable used for comparison	Section on use of digital tools and Research methods	Section on digital identity and views of digitalisation	Section on career information
Author's average of field-normalised citations per document (1996-2017)	-0.081	-0.062	-0.055
	(0.058)	(0.089)	(0.126)
Observations	11 936	11 936	11 936
Authors' number of publications (1996-2017)	-6.689***	-6.506***	-6.433***
	(0.000)	(0.000)	(0.000)
Observations	11 936	11 936	11 936
Al-related reference document	0.002	0.006*	0.005
	(0.565)	(0.049)	(0.083)
Observations	11 963	11 963	11 963
Author's share of publications in OA journals (1996-2017)	-0.000	0.008	0.007
	(0.979)	(0.178)	(0.168)
Observations	11 686	11 686	11 686

T-tests for equality of average characteristics of different groups in the target sample by item response status

Note: P-values in parentheses. * refers to p<0.05; ** refers to p<0.01; *** refers to p<0.001. A document's relatedness to AI was derived through a text mining exercise of keywords in scientific publications based on Scopus data (see OECD, 2019).

Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Additionally, the study of response rate patterns by item suggests that an author's familiarity with certain survey topics or background may have affected his/her participation in the survey. More specifically, two questions stand out as being more prone to potential bias, namely questions relating to the dissemination of data and code arising from the reference document and those on the use of advanced digital tools. Table 4.6 compares the characteristics of the respondents and non-respondents to these questions suggesting that there are significant differences between these two groups.

²⁶ T-test results for this comparison on a country by country basis are shown in Table A B.2

Comparison groups Null hypothesis: mean A (non resp)-mean B (resp)=0 Variable used for comparison	Features of data and code sharing	Use and/or development of advanced digital tools or methods
Number of publications (1996-2017)	-3.953***	-7.563***
	(0.000)	(0.000)
Observations	11 936	11 936
Normalised, mean citations (1996-2017)	-0.089**	-0.143***
	(0.004)	(0.000)
Observations	11936	11 936
Share of publications in OA journals (1996-	-0.02***	0.000
2017)	(0.000)	(0.945)
Observations	11 686	11 686
AI-related reference document	-0.014***	-0.001
	(0.000)	(0.699)
Observations	11 963	11 963
Use and development of data	-0.643***	0.005
	(0.000)	(0.641)
Observations	11 052	11 052
Gathering, digitising or curating information	-0.049***	-0.069***
	(0.000)	(0.000)
Observations	9 633	9 633
Use of computational modelling and	-0.181***	-0.183***
simulation methods	(0.000)	(0.000)
Observations	9 633	9 633

Table 4.6. Comparison of key item respondents and non-respondents, selected questions

T-tests for equality of average characteristics of different groups in the target sample by item response status

Note: P-values in parentheses. * refers to p<0.05; ** refers to p<0.01; *** refers to p<0.001. A document's relatedness to AI was derived through a text mining exercise of keywords in scientific publications based on Scopus data (see OECD, 2019). Use and development of data, Features of data or code sharing, Gathering, digitising or curating information, Use of computational modelling and simulation methods, Use and/or development of advanced digital tools or method refer to responses to Questions 8, 12, 15, and 16 of the ISSA2 questionnaire (see http://oe.cd/qissa2), respectively.

Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

The question on use and/or development of advanced digital tools or methods targets all survey respondents to assess the overall level of use or development of more advanced, digital tools. However, t-test statistics suggest that responses to this question may underestimate the level of no use or development of the tools. T-test results suggest that authors that responded to this question tend to be more productive, have more publications, adopt computational modelling and simulation methods in their research, and be involved in data collection or curation. Survey participants that are not familiar with these tools may thus have intentionally not answered the question.

Finally, we compare the characteristics of respondents to different survey wave cohorts to predict whether subsequent waves would potentially have created differences among respondents. T-test results for the whole sample suggest that there are no significant differences between late and early respondents (Table 4.4). Early respondents include authors that participated in the first wave of the survey, where late respondents include those that responded to the survey in the last, fourth wave. In contrast, a number of significant differences are found at the country level (Table A B.3). Early respondents in

Belgium are likely to have lower-cited papers, whereas those in Austria and India have more publications. Early respondents in the Netherlands have most-cited publications. Finally, there are significant differences at the 5 percent confidence level between early and late respondents in Spain and Sweden in terms of number of publications in OAjournals.

Survey participation modelling

In light of the results highlighted above, author's propensity to participate in the survey is assessed in a more comprehensive approach, which takes into consideration a number of authors' and reference document's characteristics at the same time. Table 4.7 reports the estimates of a multinomial logit considering different types of response outcomes, which include bounce, no click, clicked non-response, and partial or complete response. Survey respondents seem to have a longer academic history but have a lower number of publications and are more likely to be male. Additionally, results also indicate a higher propensity to click or respond to the survey among authors in social, physical sciences and engineering, and, to a less extent, life sciences, compared to health sciences. Furthermore, authors of publications that are available in an OA journal are more likely to click and complete the survey than their counterparts.

	Baseline = no click				
	Bounces	Click, no response	Click, partial response	Click, submitted	
Baseline = Health sciences					
Life sciences	-0.138	0.157*	-0.022	0.341***	
	(0.213)	(0.033)	(0.748)	(0.000)	
Physical sciences and engineering	0.029	0.427***	0.165**	0.526***	
	(0.744)	(0.000)	(0.004)	(0.000)	
Social sciences and humanities	0.093	0.672***	0.691***	1.138***	
	(0.424)	(0.000)	(0.000)	(0.000)	
Affiliation with a higher education institution in 2017	-0.07	-0.045	0.014	-0.130***	
	(0.387)	(0.374)	(0.771)	(0.000)	
Information on gender not available	-0.022	0.025	-0.030	-0.224***	
	(0.796)	(0.634)	(0.554)	(0.000)	
Female	-0.288**	-0.296***	-0.069	-0.233***	
	(0.003)	(0.000)	(0.175)	(0.000)	
Academic age (1996-2017)	0.069**	0.054***	0.014	0.069***	
	(0.004)	(0.000)	(0.335)	(0.000)	
Academic age squared (1996-2017)	-0.002*	-0.001	0.000	-0.001**	
	(0.012)	(0.340)	(0.320)	(0.001)	
Long-term mobility (1996-2017)	0.092	-0.029	0.223***	0.064	
	(0.382)	(0.683)	(0.000)	(0.060)	
Average of field-normalised citations per document (1996-2017)	-0.059	-0.057**	0.017*	-0.01	
	(0.062)	(0.006)	(0.027)	(0.308)	
Number of publications (1996-2017, in log)	-0.143**	-0.311***	-0.250***	-0.215***	
	(0.002)	(0.000)	(0.000)	(0.000)	

Table 4.7. Multinomial logistic regression estimates of response outcome
THE 2018 OECD INTERNATIONAL SURVEY OF SCIENTIFIC AUTHORS | 37

Reference publication in an OA journal	-0.073	0.232***	0.007	0.071*
	(0.433)	(0.000)	(0.897)	(0.021)
Al-related reference publication	0.401*	0.230	-0.223	-0.310**
	(0.045)	(0.107)	(0.175)	(0.003)
International collaboration	-0.073	-0.164**	0.051	0.075**
	(0.387)	(0.003)	(0.289)	(0.006)
Constant	-4.319***	-3.499***	-3.337***	-2.859***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	146 595			
Pseudo R-squared	0.075			

Note: P-values in parentheses. Country-fixed effects included in the regression. * refers to p<0.05; ** refers to p<0.01; *** refers to p<0.001. *Academic age* is calculated as the number of years between the first and last publications in Scopus. Information on *gender* was derived by linking the scientific author's name in Scopus to the worldwide gender-name dictionary compiled by WIPO. *Long-term mobility* is a dummy variable equal to 1 when an author's country of affiliation reported in his/her first publication indexed in Scopus is different from that reported in his/her last publication available in Scopus, and 0 otherwise. *International collaboration* is a dummy variable equal to 1 if an author's reference document involves institutional affiliations with other countries or economies, and 0 otherwise. *A1-related reference publication* is a dummy variable, which is equal to one if the publication is AI-related, and 0 otherwise. A document's relatedness to AI was derived through a text mining exercise of keywords in scientific publications based on Scopus data (see OECD, 2019). *Source*: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Weight adjustments for non-response

In order to make the results of the ISSA2 data analysis as representative as possible of its target population, survey weights are adjusted for unit nonresponse. A respondent's likelihood to respond to the survey is estimated through a logistic regression model using the following characteristics:

- gender, based on the likely gender of the respondent based on her/his first name²⁷;
- specific document-level measures, including:
 - o research AI-relatedness
 - Open Access journal publication;
 - o normalised citation ranking of the survey underling document;
- affiliation to a higher education institution in 2017, derived from the name of the institution of affiliation;
- career related indicators:
 - academic age, measured as number of years between the time of the first and last publication listed in Scopus;
 - o number of publications listed in Scopus in the period 1996-2017;

²⁷ This is an approximate variable derived by linking the scientific author's name in Scopus to the worldwide gender-name dictionary compiled by WIPO. Please see Lax-Martinez et al. (2016) for more information on the WIPO worldwide gender-name dictionary data.

- average normalised number of citations received per publication over the period 1996-2017;
- average science journal ranking of the publications listed in Scopus in the period 1996-2017;
- long-term mobility, based on whether an author's country of affiliation reported in his/her first publication indexed in Scopus is different from that reported in his/her last publication available in Scopus;
- collaboration in the period 1996-2017, measured by the proportion of an author's documents involving co-authorship and indexed in Scopus in the period 1996-2017; and
- international collaboration in the period 1996-2017, measured by the proportion of an author's documents indexed in Scopus in the period 1996-2017 and coauthored among institutions in different countries.

The model additionally controls for country-science field combinations fixed effects²⁸. The non-response weights are calculated as the inverse of the predicted response probabilities derived from the logistic regression model. The non-response weights are then combined with the selection weights described in the sampling description to create the document-and author-level final weights used for analysis.

Additionally, item non-response adjustments were also estimated using a logit model predicting a respondent's probability to complete and finalise the survey²⁹. This model was based exclusively on the sample of survey respondents.³⁰ The resulting set of weights are used to correct for potential non-response bias at the questionnaire item-level only in the cases in which the effect of survey drop-out is significant.³¹ Unless otherwise indicated, the results in this and the main report are based on the use of sampling weights adjusted by unit and item nonresponses, calculated as follows:

$$FW_j = \frac{1}{(P_j^S * P_j^{SR} * P_j^{CR})}$$

²⁸ The estimation results are provided in Table A C.1.

²⁹ The estimation results are provided in Table A C.2.

³⁰ The list of variables used comprises: gender, age, AI-relatedness of the reference document, development of new data or code in the context of the reference document, number of publications listed on Scopus in the period 1996-2017, affiliation to a higher education institution in 2017 (based on the information provided for the reference document), normalised mean number of citations received by the publications listed in Scopus in the period 1996-2017, and long-term mobility (defined as described above).

³¹ To further correct for the potential non-response bias stemming from non-responses to the question on use and/or development of advanced digital tools or methods, it is assumed that the survey participants that did not answer such a question but completed the survey item preceding it are not involved in any of the activities listed in that question. This results in around 700 responses being imputed as "neither use nor development".

Where P_j^S is the sampling probability of author (or document) *j*, P_j^{SR} is the probability to participate in (start) the survey of author (or of the author of document) *j*, $P_j^{A,CR}$ is the probability to complete and finalise the survey of author (or of the author of document) *j*.

1. Finally, while the selection and non-response adjustments correct for potential imbalances in the achieved sample that are due to "observable" features of the sample selection and response processes, the reweighted results may still fail to reproduce key features of the underlying population. Potential factors include selection based on unobservable characteristics, failure to account for the lack of replacement in sampling. Because detailed demographic statistics (other than those implied by bibliometric analysis and already used in the model) are not available, not attempt is made to undertake additional calibration reweighting. The characterisation of respondents based on the responses collected can be compared to official statistics on human resources for R&D with the major caveat that these two populations do not necessarily match for the reasons already described earlier in this document.

5. The profile of scientific authors in the ISSA study

As the final element of this technical note, this section provides a general description of the demographic and career profile of scientific authors as captured in the ISSA2 study, with a main focus on attributes such as gender, age, educational background and careers. This description aims to characterise the diversity of survey participants and points at how this information can shed light on the questions about digitalisation that populate the core of the questionnaire.

Additional descriptive statistics (i.e. proportions, standard errors, confidence internals and design effects) for the near full set of variables collected through the survey are presented in Table A D.1, giving an indication of the precision of the estimates. This Annex table shows that most questions have a 95% confidence interval whose extremes are around plus or minus one to three percentage points of the survey percentages. Variables with the highest variation are those relating to the factors influencing decisions to share scientific outputs and those on the use of digital tools or methods. The design effects, which compare complex sampling design to its simple random equivalent, are also higher for these variables as well as for the questions on gender and age. The comparison between unweighted and weighted descriptive statistics provided in this section sheds some light on the potential impact of response behaviour and the nature of the statistical corrections introduced based on information on response patterns.

The following codes were used for the fields captured in the All Science Journal Classification used in the Scopus and ISSA2 database.

Code	Field
AGRBIO	Agricultural and Biological Sciences
ALL	All fields
ARTHUM	Arts and Humanities
BIOCHEM	Biochemistry, Genetics, Molecular Biology, Immunology and Microbiology
BUSMAN	Business, Management and Accounting
CHEMENG	Chemical engineering and Chemistry
COMPSCI	Computer Science
ECODEC	Economics, Finance and Decision sci.
ENERENV	Energy and Environmental Science
ENG	Engineering
EARPLAN	Earth and Planetary Sciences
MATHS	Mathematics
MEDHEAL	Medicine and Health Professions
MATSCI	Materials Science
MULTIDIS	Multidisciplinary
PHARNEU	Pharmacology, Toxicology, Pharmaceutics and Neuroscience
PHYSAST	Physics and Astronomy
SOCPSY	Social Sciences and Psychology

 Table 5.1. Scientific research field codes

Scientific authors' gender, age and educational attainment

Survey results suggest that there are large differences in the share of women among scientific corresponding authors (Figure 5.1). As also documented in the ISSA1 findings (Boselli and Galindo-Rueda, 2016), the lowest shares of female authors are found in Asian countries (Korea, Japan and China) and Saudi Arabia, whereas the largest shares are found in some European transition economies and Argentina. As previously noted, the results for both China and Korea should be treated with extreme caution as revealed by the very large nonresponse rates in those countries and the very large difference between weighted and unweighted estimates. Overall, only 30% of corresponding authors are women.



Figure 5.1. Female authors, by country/economy of affiliation

Percentage of women respondents within each country/economy

Note: Only countries/economies with more than 75 responses are displayed in the chart. Author-level sampling weights adjusted for unit nonresponses are used. *Source:* OECD International Survey of Scientific Authors (ISSA), 2018. <u>http://oe.cd/issa</u>.

There are systematic differences in the incidence of female corresponding authors across science fields, from 45 percent in the Social sciences and psychology to less than 15 percent in Physics and astronomy, Engineering and Multidisciplinary journals (Figure 5.2).

Figure 5.2. Female scientific authors, by field



Percentage of all respondents within each field

Note: Author-level sampling weights adjusted for unit nonresponses are used. *Source:* OECD International Survey of Scientific Authors (ISSA), 2018. <u>http://oe.cd/issa</u>.

Although the population of corresponding authors does not necessarily match that of researchers, it can be instructive to compare the share of women in higher education³² according to whether they are researchers (R&D statistics) or corresponding authors, in the sampling frame or the achieved sample (weighted and unweighted). The results in Table 5.2 show that weighted survey responses are close to those in the sampling frame in most countries, indicating that the response adjustments help restore the balance of results. In general, the share of female corresponding authors seems to be lower than that of female researchers, especially in Austria, Germany, the UK, Japan and Russian Federation. These figures may point to difficulties faced by female researchers to feature as corresponding authors when they publish. This may be ultimately related to issues of access to scientific leadership positions.

Table 5.2. Women in Higher Education – a comparison of R&D statistics, Scopus and ISSA2

	Researchers	Scopus sampling frame of corresponding authors	ISSA2 corresponding authors respondents (unweighted)	ISSA2 corresponding authors respondents (weighted)
Source of HE affiliation info	OECD R&D statistics, 2017 or nearest	Implied from institution name in Scopus	Self-reported by respondent	Self-reported by respondent
Country/ Source of gender information	OECD R&D statistics	Implied from name in Scopus	Self-reported by respondent	Self-reported by respondent
AUT	39.9	26.6	27.2	26.4
CHE	38.8	30.6	27.3	31.9
CHL	32.7	28.1	31.5	33.3
CZE	34.5	23.8	17.3	15.3
DEU	38.9	25.4	19.9	23.2
DNK	44.5	33.8	27.8	28.7
ESP	42.1	35.1	33.8	37.7
EST	47.7	40.2	49.5	43.3
FIN	48.4	38.1	28.7	30.6
FRA	36.4	28.5	29.2	32.9
GBR	45.5	32.1	28.3	34.3
GRC	37.7	25.3	22.8	20.0
IRL	45.1	36.7	42.9	45.2
ITA	41.0	34.6	31.0	33.0
JPN	27.1	12.5	10.6	11.9
KOR	31.7	20.5	15.6	15.0
LTU	55.8	39.6	50.6	46.9
MEX	34.6	27.7	35.3	32.1
NLD	42.8	35.4	35.3	42.8
NOR	48.2	36.0	34.1	40.7
POL	43.7	38.2	27.7	26.9
PRT	48.8	42.7	43.3	47.6
RUS	46.3	22.6	26.8	33.7

Share of women in the higher-education sector for the relevant category, 2017

³² The comparison is restricted to the Higher Education sector to minimise the coverage bias of researchers in bibliometric data for business sector.

	Researchers	Scopus sampling frame of corresponding authors	ISSA2 corresponding authors respondents (unweighted)	ISSA2 corresponding authors respondents (weighted)
SVN	41.7	38.1	46.8	48.8
SWE	44.9	36.2	29.7	30.0
TUR	43.3	33.8	34.0	38.6
ZAF	46.0	25.8	31.9	36.6

Note: For the figures relating to the share of authors, *country* refers to the country/economy of affiliation. As an author can have more than one document and each document can report a different country of affiliation, the number of authors per country is calculated as the sum of the weights of the authors affiliated to institutions in the specific country. An author's weight is calculated as the inverse of her total number of documents. Information on gender of the authors in the survey sampling frame is derived linking the scientific author's name in Scopus to the worldwide gender-name dictionary compiled by WIPO. Researchers are in head counts. Figures on researchers refer to the latest year available.

Source: OECD International Survey of Scientific Authors (ISSA), 2018. <u>http://oe.cd/issa</u>. OECD, Research and Development Statistics Database, <u>http://oe.cd/rds</u>. June 2019.

Figure 5.3 reports on the age profile of scientific authors by country of affiliation. It is worth noting that response patterns by age are only taken into account in the weighting through the estimated effect of the proxy variable on academic publishing age, so that comparisons may reflect varying propensity to participate in online surveys. The results suggest that the average age of corresponding authors is above 40 with more than 50 percent of them in 35-to-55 age bracket (Table A D.1). Authors in Israel, Bulgaria, and Australia are among the oldest on average at over 50, whereas those in Pakistan and India are the youngest with less than 40. Significant differences are also found by field (Figure 5.4). Authors are youngest in Computer sciences, whereas they tend to be older in Biochemistry, genetics, molecular biology, and immunology.





Note: Only countries/economies with more than 75 responses are shown in the chart. Author-level sampling weights adjusted for unit nonresponses are used. *Source:* OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.



Figure 5.4. Authors' average age, by science field

Note: Author-level sampling weights adjusted for unit nonresponses are used. *Source:* OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Around 85% of corresponding authors hold a doctoral degree. This percentage is highest at around 95% in Switzerland, France and Israel, whereas it is lowest at around 55% in Colombia, followed by India. At less than 80%, Computer science, Engineering and Health/medicine are the fields where holding a doctoral degree is least common.



Figure 5.5. Authors with a doctoral degree, by country/economy of affiliation

Note: Only countries/economies with more than 75 responses are shown in the chart. Author-level sampling weights adjusted for nonresponses are used. *Source:* OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.



Figure 5.6. Authors with a doctoral degree, by science field

Percentage of respondents within each field

Note: Author-level sampling weights adjusted for nonresponses are used. *Source:* OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Sector, occupation and employment security status

The survey collects information on an author's sector of employment. It does so by asking the respondents to indicate whether his/her employer is in the public sector, whether it is a for-profit organisation, and the main area of the employer's economic activity. Responses to these questions make it possible to classify an author's employment sector into one of the following institutional sectors (OECD, 2015): business enterprise, government, higher education, and private non-profit. A unit's allocation to one of these categories is described in Figure 5.7.

46 | THE 2018 OECD INTERNATIONAL SURVEY OF SCIENTIFIC AUTHORS



Figure 5.7. Identification of author's main institutional sector of employment

Approximation to Frascati Manual guidelines

Source: Adapted from OECD (2015b), Frascati Manual 2015: Guidelines for Collecting and Reporting on Research and Experimental Development, <u>http://oe.cd.frascati</u>.

Figure 5.8 reports on the institutional sector of employment of survey respondents according to their country of residence. In the OECD area, Higher Education accounts for nearly 60% of all corresponding authors, followed by Government at approximately 25%. The results are reflective of the limited role of business-based research when it comes to scientific publishing and the organisation of national scientific research systems. More than 80 percent of corresponding authors are part of the higher education sector in Chile, Colombia, and Portugal. The share of authors employed in government institutions outside the HE sector is highest in Argentina, France and Italy at over 40%.





Note: Only countries/economies with more than 75 responses are shown in the chart. Author-level sampling weights adjusted for nonresponses are used.

Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Large differences in authors' sector of employment are also found by science field (Figure 5.9). Whereas the highest share of authors in all sectors are employed in HE, this share is largest and above 70% in Arts and humanities, and Social sciences and psychology, whereas it is lowest in Medicine. Scientific authors in the business sector account for nearly 20% of the total number of authors in computer sciences, just as many as government institutions, suggesting that scholarly publishing is not as uncommon among business researchers working in this area.

Figures related to an author's sector of employment can be compared to statistics on the overall population of researchers. As shown in Table 5.3, the populations of researchers and corresponding authors, proxied by the weighted share of authors in the ISSA2 achieved sample, differ considerably in terms of sectoral composition. Business researchers are under-represented in the population of scientific authors.



Figure 5.9. Scientific authors' employment sector, by science field

Weighted percentage of respondents within each field

Note: Author-level sampling weights adjusted for nonresponses are used. *Source:* OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.



Country	Weighted share of a	authors in the ac	hieved ISSA2 sa	mple	Share of	researchers i	n R&D statistic	s
	GOV	PNP	BE	HE	GOV	PNP	BE	HE
AUS	12.8	6.6	4.4	76.2	8.9	3.3	29.9	57.8
AUT	29.9	4.9	19.2	46.0	7.5	0.8	62.2	29.5
BEL	21.8	5.9	13.4	58.9	7.5	0.6	54.4	37.5
CAN	27.6	5.8	7.0	59.6	4.8	0.4	56.7	38.2
CHL	11.3	3.3	1.3	84.1	14.4	7.6	29.5	48.5
CZE	39.8	0.8	10.2	49.3	20.2	0.5	51.6	27.8
DEU	49.6	3.2	10.8	36.4	13.0		59.7	27.3
DNK	17.5	4.1	4.5	73.9	3.1	0.5	60.7	35.7
ESP	28.2	1.6	5.3	65.0	15.6	0.2	37.2	47.0
EST	24.4	0.2	4.3	71.2	11.7	1.9	33.9	52.5
FIN	23.8	5.0	8.7	62.5	9.0	1.2	55.5	34.4
FRA	46.9	3.2	10.4	39.4	9.9	1.6	60.3	28.3
GBR	10.2	5.0	3.9	80.8	2.3	1.5	37.9	58.2
GRC	36.6	8.0	5.9	49.5	18.1	0.8	30.3	50.8
ISR	16.4	5.8	0.1	77.7	0.8	0.7	83.7	14.9
ITA	38.6	4.0	3.0	54.4	16.0	3.3	42.6	38.1
JPN	23.5	3.8	11.2	61.4	4.5	1.2	73.7	20.5
KOR	27.9	5.6	14.6	51.8	7.1	1.4	81.3	10.1
MEX	30.6	2.7	3.9	62.8	24.3	3.0	24.5	48.2
NLD	34.8	8.1	10.0	47.1	10.7		61.8	27.4
POL	19.6	0.9	8.6	70.9	2.4	0.4	52.0	45.2
PRT	8.7	2.7	8.4	80.1	3.3	1.1	33.7	61.9
RUS	24.0	1.6	27.8	46.6	34.0	0.4	47.1	18.5

Selected economies

Country	Weighted share of authors in the achieved ISSA2 sample				Sha	are of researche	ers in R&D stat	istics
	GOV	PNP	BE	HE	GOV	PNP	BE	HE
SVN	35.3	2.1	4.2	58.5	16.2	0.3	61.8	21.7
SWE	32.5	1.1	6.7	59.7	4.8	0.2	67.0	28.0
USA	17.4	8.1	9.7	64.8	n/a	n/a	70.9	n/a
ZAF	11.1	3.1	5.0	80.8	10.6	1.5	17.7	70.2

Note: For the figures relating to the share of authors, *country* refers to the country/economy of affiliation. Researchers and authors are presented in full-time equivalents. Data on researchers refer to the latest year available. ISSA2 figures are weighted by the author-level final weights.

Source: OECD International Survey of Scientific Authors (ISSA), 2018. <u>http://oe.cd/issa</u>. OECD, Research and Development Statistics Database, <u>http://oe.cd/rds</u>. June 2019.

Survey respondents were also asked to identify their main occupation according to the International Standard Classification of Occupations (ISCO). Their responses can be grouped into the three main broad categories: HE teaching professionals, science professionals, and other occupations. According to results presented in Figure 5.10, more than 60% of authors in social sciences, economics, finance, decision sciences, and business, management and accounting are HE teaching professionals. In contrast, the self-reported "science professionals" in the ISCO classification are mostly found among authors in Agricultural and biological sciences and Earth and planetary sciences. The group of "other occupations" is considerably more homogeneous, comprising for example health professionals (as reflected by the large share for the field of Medicine and health), engineers and software specialists (as visible by the also large share for Computer science).





Note: Author-level sampling weights adjusted for nonresponses are used. *Source:* OECD International Survey of Scientific Authors (ISSA), 2018. <u>http://oe.cd/issa</u>.

The survey also provides information on the level of participating authors' job security. Respondents were asked to indicate a) whether they hold an indefinite contract; and b) whether they can be dismissed for reasons other than gross misconduct. This information allows distinguishing between authors with a permanent, highly protected contract with those with also indefinite contracts but relatively low levels of protection. The pursuit of

science careers appears to be associated with different levels of job security according to country of residence (Figure 5.11). For example, while scientific authors in France and United Kingdom have similar probabilities of being subject to indefinite contracts, nearly half of those in the United Kingdom have contracts offering high levels of protection compared to nearly all in the case of France. Participating scientific authors in the Russian Federation face the highest rates of fixed term contracts. These results may not fully extrapolate to the entire scientific population if participation is related to seniority levels and the reweighting scheme does not fully account for that.

Figure 5.11. Authors' job security, by country/economy of residence



Weighted percentage of employed respondents within each group

Note: Only countries/economies with more than 75 responses are shown in the chart. Author-level sampling weights adjusted for nonresponses are used.

Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

When considering the author's science field, there is evidence of lower job security in Agricultural and biological sciences and Engineering. Furthermore, survey results also seem to point to lower job security among authors that are less than 35-years-old and for associate professionals (Table 5.4).

Table 5.4. Scientific authors' job security by age and occupation

Weighted percentage of employed respondents in each security status, by group

	Indefinite protected employment contract	Other indefinite	Fixed term employment contract
Age groups			
Up to 34	16.45	8.88	74.67
35 to 44	34.37	14.38	51.25
45 to 54	52.69	18.08	29.22
55 to 64	61.63	12.94	25.43
65 or more	49.12	15.63	35.25
Occupation groups			
University and Higher Education Teaching Professionals	52.84	11.15	36.02
Senior managers and legislators	48.49	23.23	28.28
Science Professionals	35.4	15.09	49.51
Other Professionals	34.85	17.46	47.69
Associate professionals	21.36	13.66	64.98
Other occupations	35.24	6.33	58.43

Note: Author-level sampling weights adjusted for nonresponses are used.

Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Mobility

A number of geographical mobility measures can be derived based on the data collected in the ISSA2 study. In addition to the bibliographic information about changes in affiliation over time, survey responses allow to compare where authors are resident at the point of being interviewed, where they achieved their highest qualifications, and where they were born.

Results in Figure 5.12 report mobility rates based on the comparison of higher qualification location and current residence. Globally, 22% of authors reside in a country that differs from the place where they attained their highest qualification, compared to 29% of authors who reside in a country that differs from the country they were born in. The "study-to-reside" mobility rate is largest for residents in Malaysia, Colombia and Chile, with values close to 50%. In contrast, less than 5% of residents in the Russian Federation and Ukraine attained their highest degree abroad.

These "study-to-reside" mobility rates vary by science field (Figure 5.13), with the share of authors who studied abroad ranging from 30% in environmental science to less than 20% percent in medicine and computer sciences, suggesting that study is associated to work opportunities in that domain.

52 | THE 2018 OECD INTERNATIONAL SURVEY OF SCIENTIFIC AUTHORS



Figure 5.12. Authors that graduated abroad (vs current residence), by country/economy of residence

Percentage of respondents within each field

Note: Only countries/economies with more than 75 responses are shown in the chart. Author-level sampling weights adjusted for nonresponses are used.

Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.





Percentage of respondents within each field

Note: Author-level sampling weights adjusted for nonresponses are used. *Source:* OECD International Survey of Scientific Authors (ISSA), 2018. <u>http://oe.cd/issa</u>.

Figure 5.14 and Figure 5.15 report on the proportion of authors that plan to move to another country in the next year (approximately 9% on average). This share seems to be the highest in Indonesia and Turkey (with rates over 35%) whereas it is the lowest in the cases of Australia, New Zealand and Norway (in the order of 1 to 2%). Authors working in the field

of Computer science are those more inclined to move abroad (over 20%), while authors in the fields of Medicine and Biochemistry, genetics, molecular biology, immunology and microbiology are the least keen to do, probably reflecting opportunities and challenges associated to the transferability of knowledge abroad.

Figure 5.14. Proportion of authors that plan to move to another country in the next year, by country/economy of residence



Note: Only countries/economies with more than 75 responses are shown in the chart. Author-level sampling weights adjusted for nonresponses are used. *Source:* OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure 5.15. Proportion of authors that plan to move to another country in the next year, by science field



Note: Author-level sampling weights adjusted for nonresponses are used. *Source:* OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Research funding and scientists' earnings

The ISSA2 questionnaire's module on careers also includes questions on earnings and sources of research funding. Although it is beyond the scope of this technical note to explore these data in detail, it is worth noting that basic descriptive statistics in the Annex section show that in 82% of cases research work is part of the respondents' core responsibilities and their remuneration. Authors in medicine and other health disciplines are among those more likely not to be expected to engage in research.

A very significant number of authors have to compete for resources to conduct research work (either to fund themselves or their teams): 65% report competing for resources controlled by their organisation, while 71% report directly competing for external resources. This information, which suggests that project and institutional sources of funding are highly intertwined, can facilitate the study of research funding systems from a bottom-up perspective. For example, while institutions may count on stable and predictable sources of funding, researchers may find themselves in competition for either jobs or positions on the one hand, or the ability to use such resources for their own projects on the other hand.

There are currently no authoritative sources on scientist earnings worldwide across different positions and roles. This is an important hindrance to a wide range of analyses as well as potential personal career decisions, as pecuniary factors can also come into play in addition to other motivations to carry out research in different contexts. Survey results indicate considerable income dispersion across scientists worldwide. The ISSA2 results show that close to 15% of publishing scientists (corresponding authors) earn less than USD 15 000 per year, compared to nearly 6% earning USD 150 000 or more (Figure 5.16). These data on incomes are used more extensively in the main paper on the survey's results (Bello and Galindo-Rueda, 2020) to assess the extent to which the scientific labour market rewards a range of digital skills and practices.



Percentage of authors whose annual earnings belong to a specific range



Note: Author-level sampling weights adjusted for nonresponses are used. Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

6. Using the ISSA2 data

The ISSA2 public use files are made available to interested users for research and policy purposes on the OECD dedicated page: <u>http://oe.cd/issa</u>. These include:

- 1. Anonymised responses
- 2. Anonymised, micro dataset, containing survey responses and derived indicators generated for the data analysis.
- 3. Code to replicate the data analysis in Bello and Galindo-Rueda (2020) focusing on various aspects of the digitalisation of science.

ISSA2 responses can be downloaded directly from the above OECD page, whereas access to the micro dataset and code is exclusively granted to approved researchers. Interested users can apply for access by filling out the relevant form and submitting a research proposal to the ISSA2 team. Instructions are provided on the same page.

7. Concluding remarks

This document has outlined the main features of the ISSA2 project and has provided an assessment of the data collected. It accompanies the analysis of the study's main results in relation to the subject on how science is going digital (Bello and Galindo-Rueda, 2020). ISSA2 gathered data on approximately 12,000 authors worldwide. Although response rates are rather low, they are in line with comparable studies, and outperform those achieved by commercial parties. The wealth of information collected makes it possible to explore key aspects of digitalization and science from different perspectives, complementing standard approaches already existing in this field. A number of survey data adjustments have been developed to correct for possible non-response bias and improve the quality and usability of the data collected. The implementation of a survey linked to bibliometric information has also allowed assessing the representativeness of the data and reweighting it to reproduce the features of the population of corresponding authors. Conclusions drawn from this study can still provide a valuable tool to investigate the new facets of science in the digital era and can be representative of a larger population, although not necessarily the full researcher population as business-based researchers are, by and large, rather unlikely to publish in scholarly journals or serialised conference proceedings.

There are a number of key methodological takeaways from this study as part of the development and testing of the business case for a potential, regular OECD survey or "barometer" of science:

- The procedure for invitation and authentication needs to be streamlined to overcome the trust / participation gap, and has to be supported by state of the art tools. For this type of population, an alternative approach might involve respondent authentication by means of widely used identifiers such as ORCID. This would allow in principle to avoid the use of invasive email campaigns.
- Future surveys need to be somewhat shorter and face respondents with simpler constructs. As already appreciated in other studies of ICT adoption, it can be difficult to secure accurate responses concerning the adoption of advanced digital tools and practices. This technical information may often reside in other team members.
- The initiative would considerably benefit from explicit support or at least endorsement by national authorities. In order to achieve sufficiently precise national estimates, a higher order of magnitude in the number of participants is required.

ISSA2 is the second pilot of the OECD initiative to develop complementary data infrastructures aimed to inform the analysis of key and rapidly changing science policy topics. Despite a number of challenges, a majority of the criteria laid out in the business case have been demonstrated, while potential avenues to address others have been identified. There is scope for OECD to capitalise on its reputation as a trusted provider of statistical data, proximity to policy decision makers, and engagement with national experts to build a sustainable tool. The bulk of future methodological efforts has to concentrate on developing the value proposition for scientists as ultimate information providers.

References

- Arora, A, Belenzon, S, Patacconi, A. (2017) The decline of science in corporate R&D. Strategic Management Journal; 39: 3– 32. <u>https://doi.org/10.1002/smj.2693</u>
- Bello, M. and F. Galindo-Rueda (2020). "Charting the digital transformation of science: Findings from the 2018 OECD International Survey of Scientific Authors (ISSA2)", OECD Science, Technology and Industry Working Papers, OECD Publishing, Paris. <u>https://doi.org/10.1787/18151965</u>
- Beullens, K., Loosveldt G., Vandenplas C. & Stoop I. (2018). Response Rates in the European Social Survey: Increasing, Decreasing, or a Matter of Fieldwork Efforts? Survey Methods: Insights from the Field. Retrieved from <u>https://surveyinsights.org/?p=9673</u>
- Boselli, B. and F. Galindo-Rueda (2016), "Drivers and Implications of Scientific Open Access Publishing: Findings from a Pilot OECD International Survey of Scientific Authors", OECD Science, Technology and Industry Policy Papers, No. 33, OECD Publishing, Paris
- Digital Science (2016), *The State of Open Data*, Digital Science, https://figshare.com/articles/The_State_of_Open_Data_Report/4036398
- Elsevier (2017). Scopus Content Coverage Guide. Version updated August 2017. Accessed on 26 August 2019 from https://www.elsevier.com/__data/assets/pdf_file/0007/69451/0597-Scopus-Content-Coverage-Guide-US-LETTER-v4-HI-singles-no-ticks.pdf
- Elsevier (2019a). Open Science Monitor. Updated Methodological Note. Access on 26 August from <u>https://ec.europa.eu/info/sites/info/files/research_and_innovation/open_science_monitor_methodologi</u> <u>cal_note_april_2019.pdf</u>
- Elsevier (2019b). Research Futures. Researcher survey results www.elsevier.com/__data/assets/pdf_file/0011/847334/ResearchFutures-Survey-Results-February-2019.pdf
- Franzoni, C., Scellato, G., Stephan, P. (2012) "Foreign-born scientists: mobility patterns for 16 countries", Nature Biotechnology, 30 (12), pp. 1250-1253. https://www.nber.org/globsci/GLOBSCI%20project%20data%20manual.pdf
- Liu, M., & Wronski, L. (2018). Examining Completion Rates in Web Surveys via Over 25,000 Real-World Surveys. Social Science Computer Review, 36(1), 116–124. <u>https://doi.org/10.1177/0894439317695581</u>
- Lax-Martinez, G., Raffo, J. and Saito, K. 2016. "Identifying the Gender of PCT inventors", WIPO Economic Research Working Paper 33
- Market Research, Wiley (2017): Wiley Open Science Researcher Survey 2016 Infographic. figshare. Journal contribution. <u>https://doi.org/10.6084/m9.figshare.4910714.v1</u>

Mei, B. and Brown, G. (2018). "Conducting Online Surveys in China", Social Science Computer *Review*, Vol. 36(6) 721-734. DOI: 10.1177/0894439317729340

- OECD (2019), *Measuring the Digital Transformation: A Roadmap for the Future*, OECD Publishing, Paris, <u>https://doi.org/10.1787/9789264311992-en</u>.
- OECD (2018), "Blue Sky perspectives towards the next generation of data and indicators on science and innovation", in OECD Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption, OECD Publishing, Paris, <u>https://doi.org/10.1787/sti_in_outlook-2018-19-en</u>.
- OECD (2015), Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris, <u>https://doi.org/10.1787/9789264239012-en</u>.
- OECD / SCImago Research Group (CSIC) (2016), Compendium of Bibliometric Science Indicators.
- Thursby, J., Haeussler, C., Thursby, M.C., Jiang, L. (2018). Prepublication disclosure of scientific results: Norms, competition, and commercial orientation. *Science Advances*, Vol. 4, no. 5. DOI: 10.1126/sciadv.aar2133

Annex A. Sampling frame and target sample

Table A A.1. Number of corresponding authors in the sampling frame and target sample, by country/economy of affiliation

Country of affiliation	Number of corresponding authors in the sampling frame	Number of corresponding authors in the target sample
USA	199 867	24 052
CHN	183 621	26 837
Other countries	57 272	5 163
IND	56 894	6 972
DEU	51 087	6 214
GBR	50 654	5 954
JPN	46 630	5 647
FRA	33 139	3 907
ITA	31 196	4 150
BRA	30 511	3 589
RUS	29 073	3 478
ESP	28 467	3 549
CAN	27 386	3 543
AUS	25 746	3 312
KOR	25 222	4 047
IRN	20 022	3 046
TUR	17 101	2 167
NLD	15 558	1 883
POL	15 545	2 086
TWN	12 557	1 812
SWE	11 427	1 422
MYS	10 670	1 584
CHE	10 346	1 385
MEX	8 376	1 197
IDN	8 140	1 347
BEL	7 938	1 118
CZE	7 677	1 255
DNK	7 311	1 077
PRT	7 184	1 157
NOR	6 295	962
AUT	6 260	1 087
ISR	6 148	1 008
ZAF	6 128	965
FIN	6 060	1 006
EGY	5 642	1 025
РАК	5 257	1 020
THA	5 226	1 006
ARG	5 178	964
ROU	5 028	1 105
GRC	4 891	1 008
SGP	4 682	1 058
NZL	4 118	937
SAU	3 945	947

$\mathbf{60} \mid$ the 2018 oecd international survey of scientific authors

UKR	3 755	941
HKG	3 754	923
CHL	3 676	919
IRL	3 618	897
COL	3 610	927
HUN	3 321	935
SVK	2 712	894
HRV	2 059	898
SVN	1 791	855
BGR	1 225	671
LTU	1 210	769
EST	914	677
LVA	661	218
CYP	601	216
LUX	412	213
ISL	304	301
MLT	204	186
Total	1 145 303	158 488

Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Science field	Number of corresponding authors in the sampling frame	Number of corresponding authors in the target sample
Medicine	212 969	29 522
Engineering	137 900	17 192
Computer Science	102 649	12 240
Physics and Astronomy	88 067	11 675
Agricultural and Biological Sciences	76 598	11 944
Social Sciences	72 871	10 177
Mathematics	49 789	6 686
Biochemistry, Genetics and Molecular Biology	48 231	7 684
Materials Science	45 376	6 821
Earth and Planetary Sciences	35 627	4 209
Environmental Science	33 120	4 659
Chemistry	29 691	4 904
Arts and Humanities	29 170	3 362
Pharmacology, Toxicology and Pharmaceutics	23 299	3 774
Business, Management and Accounting	20 993	3 357
Chemical Engineering	20 707	3 315
General	16 767	2 656
Psychology	15 041	2 250
Energy	14 938	1 730
Economics, Econometrics and Finance	12 812	1 940
Immunology and Microbiology	12 356	1 994
Nursing	12 102	1 491
Veterinary	8 201	1 226
Neuroscience	7 847	1 176
Health Professions	7 326	940
Decision Sciences	5 824	868
Dentistry	5 034	696
Total	1 145 303	158 488

Table A A.2. Number of corresponding authors in the sampling frame and target sample, by science field

Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Annex B. Non-response bias analysis

Table A B.1. Comparison of unit survey respondents and non-respondents, selected countries/economies

				·
Economy affiliation	of	Author's average of field-normalised citations per document (1996-2017)	Number of publications (1996-2017)	Share of publications in OA journals (1996-2017)
AUS		-0.0857	-7.390	0.0140
		(0.482)	(0.096)	(0.254)
AUT		-0.0125	11.49*	0.00920
		(0.904)	(0.037)	(0.590)
BEL		0.153	12.91	0.00917
		(0.374)	(0.114)	(0.620)
BRA		0.0719	3.614	0.00172
		(0.206)	(0.095)	(0.927)
CAN		0.0487	6.976	-0.0181
		(0.578)	(0.076)	(0.096)
CHE		-0.104	9.486	0.0121
		(0.592)	(0.112)	(0.485)
CHN		0.00108	10.78	0.0328*
		(0.991)	(0.232)	(0.044)
CZE		0.0324	4.370	0.00377
		(0.688)	(0.410)	(0.860)
DEU		0.0254	9.015**	0.00662
		(0.714)	(0.008)	(0.433)
DNK		-0.111	6.853	0.0256
		(0.398)	(0.271)	(0.203)
EGY		0.311	7.939*	-0.00375
		(0.054)	(0.025)	(0.904)
ESP		0.0975	5.621	-0.0305*
		(0.168)	(0.090)	(0.015)
FIN		0.0321	-0.0870	0.0213
		(0.848)	(0.986)	(0.296)
FRA		0.0831	6.591	-0.00113
		(0.344)	(0.058)	(0.908)
GBR		-0.0261	4.740	0.00268
		(0.784)	(0.160)	(0.766)
GRC		0.0652	14.83**	0.00322
		(0.493)	(0.004)	(0.842)
IDN		0.0508	-1.001	-0.0171
		(0.841)	(0.477)	(0.599)
IND		0.0407	9.425***	-0.00301
		(0.417)	(0.000)	(0.797)
ISR		-0.0910	6.482	-0.0173
		(0.445)	(0.281)	(0.270)
ITA		0.194**	12.60***	-0.00784
		(0.002)	(0.000)	(0.351)
JPN		0.0688	9.962**	0.0187
		(0.178)	(0.004)	(0.059)

T-Tests for null hypothesis of no difference in averages: non respondents - respondents = 0

THE 2018 OECD INTERNATIONAL SURVEY OF SCIENTIFIC AUTHORS 6

Economy affiliation	of	Author's average of field-normalised citations per document (1996-2017)	Number of publications (1996-2017)	Share of publications in OA journals (1996-2017)
KOR		0.0205	15.33**	-0.0142
		(0.886)	(0.008)	(0.419)
MEX		0.123	1.393	0.000645
		(0.108)	(0.630)	(0.976)
MYS		0.0834	-2.751	0.00839
		(0.368)	(0.579)	(0.719)
NLD		0.117	15.49**	0.00828
		(0.379)	(0.007)	(0.590)
PAK		-0.0700	10.00	0.0494
		(0.620)	(0.060)	(0.118)
POL		-0.000512	-1.849	-0.0287
		(0.995)	(0.577)	(0.134)
PRT		0.0233	7.861	-0.0216
		(0.777)	(0.078)	(0.161)
ROU		0.0469	6.997	-0.00419
		(0.507)	(0.080)	(0.854)
RUS		0.00810	5.418	-0.0157
		(0.914)	(0.083)	(0.252)
SGP		-0.0548	3.644	0.00986
		(0.812)	(0.732)	(0.715)
SWE		-0.0772	3.439	0.0445*
		(0.568)	(0.466)	(0.016)
THA		-0.0956	14.01*	0.0272
		(0.393)	(0.025)	(0.286)
TUR		-0.155*	4.377	-0.000117
		(0.037)	(0.084)	(0.994)
USA		0.150*	10.12***	0.00229
		(0.047)	(0.000)	(0.651)

Note: P-values in parentheses. ,* refers to p<0.05; ** refers to p<0.01; *** refers to p<0.001. *Source*: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Table A B.2. Comparison of partial respondents and survey completers, selected economies

Economy of affiliation	Author's average of field-normalised citations per document (1996-2017)	Number of publications (1996-2017)	Share of publications in OA journals (1996-2017)
AUS	0.328	-0.338	-0.0194
	(0.464)	(0.979)	(0.478)
AUT	-0.134	-2.202	0.00556
	(0.592)	(0.784)	(0.888)
BEL	0.163	-15.07	0.0600
	(0.608)	(0.222)	(0.233)
BRA	0.0174	-3.608	-0.0844
	(0.859)	(0.441)	(0.056)
CAN	0.0837	-7.188	0.0747*
	(0.639)	(0.348)	(0.010)
CHE	-0.660	-11.90	0.0636
	(0.402)	(0.285)	(0.108)
CHN	0.359*	-0.132	0.0258
	(0.046)	(0.991)	(0.394)
CZE	-0.240	0.398	-0.000727
	(0.135)	(0.963)	(0.984)
DEU	-0.0830	-1.287	0.00302
	(0.608)	(0.841)	(0.887)
DNK	-0.390	-13.19	-0.0172
2	(0.227)	(0.135)	(0.696)
FGY	-0.261	-4.807	-0.00271
	(0.085)	(0.204)	(0.968)
FSP	0.0978	-7.520	-0.00318
	(0.606)	(0.302)	(0.922)
FIN	0.257	-13 49	0 109*
	(0.405)	(0.262)	(0.013)
FRA	-0.110	-7.949	-0.0461*
	(0.505)	(0.298)	(0.049)
GBR	0.219	-7 859	-0.00915
OBIN	(0.259)	(0 149)	(0.669)
GRC	-0.255	-2 695	0.0172
0110	(0.303)	(0.704)	(0.670)
וחח	0 349*	-1 663	0.0130
	(0.038)	(0.658)	(0.844)
IND	-0.0967	-7 178**	-0.0512*
	(0.328)	(0.008)	(0.027)
ISR	-0 0894	-14 37	-0.0460
	(0.783)	(0 142)	(0.245)
ITΔ	-0.0702	(0.142) _0.730	0.00824
IIA	-0.07 02	-9.750	(0.071)
IDN	0.0617	(0.002)	0.0366
	(0.638)	-0.403	(0.174)
KOR	(0.030)	(0.443) _12.56	0.00225
NUR	(0.112)	-12.00	(0.00220
	(0.113)	(0.439)	0.0404
WEA	0.139	0.020	-0.0491
MVC	(0.273)	(0.257)	(0.364)
WYS	0.128	0.549	-0.0458
	(0.287)	(0.528)	(0.343)

T-Tests for null hypothesis of no difference in averages: partial respondents – completers = 0

The 2018 OECD International survey of scientific authors $\mid 65$

Economy of affiliation	Author's average of field-normalised citations per document (1996-2017)	Number of publications (1996-2017)	Share of publications in OA journals (1996-2017)
NLD	-0.396	-14.44	0.0151
	(0.143)	(0.102)	(0.714)
PAK	0.182	-1.378	0.0902
	(0.488)	(0.746)	(0.099)
POL	-0.145	-6.738	-0.0496
	(0.361)	(0.568)	(0.323)
PRT	-0.202	4.365	-0.00885
	(0.327)	(0.563)	(0.817)
ROU	-0.200	-5.159	0.0916
	(0.103)	(0.308)	(0.099)
RUS	-0.0649	-0.966	-0.00519
	(0.712)	(0.871)	(0.881)
SGP	0.936	-18.35	-0.0000930
	(0.175)	(0.470)	(0.999)
SWE	0.253	3.566	-0.0348
	(0.622)	(0.746)	(0.341)
THA	0.0396	-5.057	-0.0798
	(0.883)	(0.296)	(0.155)
TUR	-0.387	-0.0971	0.0269
	(0.204)	(0.982)	(0.375)
USA	0.0921	-2.432	0.00489
	(0.438)	(0.485)	(0.682)

Note: P-values in parentheses. ,* refers to p<0.05; ** refers to p<0.01; *** refers to p<0.001. *Source*: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Table A B.3. Comparison of early and late respondents, selected countries/economies

Economy of affiliation	Author's average of field-normalised citations per document (1996-2017)	Number of publications (1996-2017)	Share of publications in OA journals (1996-2017)
AUS	-0.222	2.458	0.000274
	(0.402)	(0.871)	(0.992)
AUT	0.387	21.88*	-0.00213
	(0.206)	(0.020)	(0.958)
BEL	-1.129**	-12.38	-0.0399
	(0.005)	(0.453)	(0.477)
BRA	-0.0928	-5.654	-0.0158
	(0.423)	(0.226)	(0.753)
CAN	0.132	0.635	0.00470
	(0.555)	(0.945)	(0.907)
CHE	0.309	-11.22	-0.0267
	(0.769)	(0.220)	(0.386)
CHN	-0.0230	-0.830	0.0121
	(0.950)	(0.939)	(0.838)
CZE	-0.0388	1.576	-0.0370
	(0.836)	(0.851)	(0.350)
DEU	0.182	5.940	0.00244
	(0.227)	(0.407)	(0.918)
DNK	-0.428	10.10	-0.0334
	(0.253)	(0.343)	(0.541)
EGY	0.0481	-1.840	-0.0935
	(0.808)	(0.753)	(0.185)
ESP	-0.279	1.275	0.0837*
	(0.068)	(0.850)	(0.023)
FIN	0.106	9.466	0.0173
	(0.696)	(0.388)	(0.664)
FRA	0.186	-5.707	-0.0311
	(0.295)	(0.389)	(0.316)
GBR	0.00587	15.11*	0.00728
	(0.976)	(0.017)	(0.787)
GRC	0.380	-26.17*	0.0983
	(0.249)	(0.018)	(0.052)
IDN	0.222	0.779	0.0390
	(0.212)	(0.738)	(0.647)
IND	0.0785	7.826*	-0.0593
	(0.620)	(0.021)	(0.081)
ISR	-0.122	-5.327	0.00500
	(0.728)	(0.595)	(0.921)
ITA	0.0423	-0.463	-0.0326
	(0.702)	(0.937)	(0.174)
JPN	0.127	-0.0220	-0.0151
	(0.212)	(0.997)	(0.515)
KOR	0.254	29.35	0.0509
	(0.310)	(0,275)	(0.322)
MEX	-0 175	-10.27	-0.0364
	(0 151)	(0,135)	(0.513)
MYS	0 229	6,690	0.00988
	(0 131)	(0 300)	(0.858)

T-Tests for null hypothesis of no difference in averages: early respondents – late respondents = 0

THE 2018 OECD INTERNATIONAL SURVEY OF SCIENTIFIC AUTHORS $\mid 67$

Economy of affiliation	Author's average of field-normalised citations per document (1996-2017)	Number of publications (1996-2017)	Share of publications in OA journals (1996-2017)
NLD	0.927**	15.19	0.0224
	(0.010)	(0.130)	(0.608)
PAK	-0.0120	-4.298	0.0480
	(0.971)	(0.506)	(0.562)
POL	0.141	3.072	0.00231
	(0.569)	(0.816)	(0.969)
PRT	0.230	-7.166	0.0289
	(0.361)	(0.319)	(0.503)
ROU	-0.0288	-3.202	0.0456
	(0.879)	(0.631)	(0.509)
RUS	0.272	3.833	0.0140
	(0.291)	(0.456)	(0.770)
SGP	1.021	-37.13	-0.0526
	(0.277)	(0.271)	(0.486)
SWE	-0.284	-0.987	-0.0684*
	(0.468)	(0.916)	(0.038)
THA	-0.444	7.567	-0.0104
	(0.155)	(0.458)	(0.853)
TUR	0.0247	4.454	0.0543
	(0.840)	(0.334)	(0.163)
USA	-0.0355	4.935	-0.00818
	(0.822)	(0.257)	(0.569)

Note: P-values in parentheses. ,* refers to p<0.05; ** refers to p<0.01; *** refers to p<0.001. *Source*: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Annex C. Nonresponse adjustment estimation

	Survey participation
Academic age (1996-2017)	0.0217***
	(0.000)
Information availability on author's academic age	0.409
	(0.054)
Al-related reference publication	-0.224**
	(0.002)
Average science journal ranking of the publications (1996-2017)	-0.102***
	(0.000)
Affiliation to a higher education institution	-0.0941***
	(0.000)
Affiliation within OECD	-1.459***
	(0.000)
Reference publication in an OA journal	0.0468
	(0.098)
Information availability on OA status of the reference publication	-0.0460
	(0.268)
Number of publications (1996-2017)	-0.00339***
	(0.000)
Collaboration (1996-2017)	-0.573***
	(0.000)
International collaboration (1996-2017)	0.284***
	(0.000)
Average of field-normalised citations per document (1996-2017)	-0.000623
	(0.932)
Normalised citation ranking of the survey underling document	0.0923*
	(0.027)
Information availability on author gender	-0.193***
	(0.000)
Female	-0.142***
	(0.000)
Long-term mobility (1996-2017)	0.0323
	(0.270)
Constant	-1.531***
	(0.000)
Observations	158,488
Pseudo R-squared	0.091

Logistic regression estimates of an author's probability to participate in the survey

Table A C.1. Survey nonresponse adjustment estimation

Note: P-values in parentheses. Country-fixed and stratum effects included in the regression. * refers to p<0.05; ** refers to p<0.01; *** refers to p<0.001. *Academic age* is calculated as the number of years between the first and last publications in Scopus. Information on *gender* was derived by linking the scientific author's name in Scopus to the worldwide gender-name dictionary compiled by WIPO. *Long-term mobility* is a dummy variable equal to 1 when an author's country of affiliation reported in his/her first publication indexed in Scopus is different from that reported in his/her last publication available in Scopus, and 0 otherwise. *Collaboration* is measured by the proportion of an author's documents involving co-authorship and indexed in Scopus in the period 1996-2017. *International collaboration* is measured by the proportion of an author's documents involving the proportion of an author's documents indexed by the proportion of an author's documents involving the proportion of an author's documents indexed by the proportion of an author's documents involving the proportion of an author's documents involvin

in Scopus in the period 1996-2017 and co-authored among institutions in different countries. *AI-related reference publication* is a dummy variable, which is equal to one if the publication is AI-related, and 0 otherwise. A document's relatedness to AI was derived through a text mining exercise of keywords in scientific publications based on Scopus data (see OECD, 2019).

Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Table A C.2. Item nonresponse adjustment estimation

Logistic regression estimates of an author's probability to complete the survey

	Survey submission
Information availability on author gender	0.314
	(0.108)
Extra information availability	-0.570
	(0.280)
Response to question 8 (data/code development)	1.573***
	(0.000)
Information availability on age	0.542***
	(0.000)
Number of publications (1996-2017)	0.000590
	(0.306)
Average of field-normalised citations per document (1996-2017)	-0.0107
	(0.408)
Affiliation to a higher education institution	-0.0435
	(0.428)
Long-term mobility (1996-2017)	-0.0348
	(0.598)
Al-related reference publication	-0.0291
	(0.859)
	(0.428)
Female	-0.146**
	(0.005)
Age	0.0188***
	(0.000)
Data and code development	-0.226***
	(0.000)
Observations	11 947
pseudo R-squared	0.098

Note: P-values in parentheses. Country-fixed and stratum effects included in the regression. * refers to p<0.05; ** refers to p<0.01; *** refers to p<0.001. Information on *gender* was derived by linking the scientific author's name in Scopus to the worldwide gender-name dictionary compiled by WIPO. *Long-term mobility* is a dummy variable equal to 1 when an author's country of affiliation reported in his/her first publication indexed in Scopus is different from that reported in his/her last publication available in Scopus, and 0 otherwise. *Al-related reference publication* is a dummy variable, which is equal to one if the publication is Al-related, and 0 otherwise. A document's relatedness to AI was derived through a text mining exercise of keywords in scientific publications based on Scopus data (see OECD, 2019).

Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Annex D. Summary descriptive statistics

Table A D.1. Summary descriptive statistics of ISSA2 selected variables

Responses weighted by the author-level final weights

Scopus data/ Question number	Variable	Proportion	Standard error	95% cor inte	nfidence rval	DEFT	Number of responses
				Lower	Upper		
ASJC Scopus	Science subject area						(exclusive)
· ·	Health sciences	23.43	0.32	22.81	24.06	0.82	1 968
	Life sciences	15.02	0.27	14.50	15.56	0.83	1 830
	Physical sciences	48.20	0.39	47.44	48.97	0.86	4 876
	Social sciences	13.35	0.16	13.05	13.66	0.50	3 289
Affiliation Scopus	Name-implied institutional affiliation						(exclusive)
	Higher education sector	74.67	0.67	73.32	75.96	1.69	8 817
	Other	25.33	0.67	24.04	26.68	1.69	3 146
	Basic information section						
Q1	Gender						(exclusive)
	Male	71.92	0.66	70.60	73.20	1.60	8 325
	Female	28.08	0.66	26.80	29.40	1.60	3 495
	Other (suppressed for confidentiality)						
Q2	Age						(exclusive)
	Up to 34	21.29	0.75	19.86	22.79	1.97	1 954
	35 to 44	31.74	0.78	30.23	33.29	1.81	3 674
	45 to 54	21.48	0.69	20.16	22.86	1.81	2 817
	55 to 64	15.83	0.59	14.71	17.02	1.74	1 977
	65 or more	9.67	0.37	8.96	10.42	1.36	1 200
	Access to scientific outputs section						
Q6	Open access to the reference publication						
	Do not know	10.64	0.55	9.61	11.76	1.92	1 131
	Gold open access journal	27.29	0.77	25.80	28.82	1.86	3 290
	Hybrid open access journal	13.01	0.59	11.89	14.21	1.89	1 245
	Green open access	20.66	0.68	19.36	22.02	1.81	2 344
	Not open access	28.41	0.65	27.16	29.70	1.55	3 610
Q7	Author role in the paper						
	Formulating the initial idea	75.3	0.7	73.9	76.7	1.80	8 919
	Having primary responsibility for carrying out the analysis	79.9	0.6	78.6	81.1	1.66	9 144
	Supervising the work of others	46.5	0.8	44.9	48.1	1.74	5 097
	Providing access to key research resources	42.5	0.8	41.0	44.0	1.68	4 732
	Undertaking data analysis	67.1	0.8	65.6	68.7	1.83	7 850
	Writing the manuscript	85.7	0.7	84.3	87.0	2.17	10 153
	Other	2.2	0.2	1.8	2.6	1.53	274
Q8	Reference publication resulting in new data or code						(exclusive)
	Do not know	4.89	0.37	4.21	5.67	1.85	532
	Only data	32.73	0.76	31.26	34.24	1.74	3 905
	Only code	8.67	0.43	7.86	9.54	1.64	940
	Both data and code	22.41	0.68	21.10	23.78	1.77	2 457
	Neither data nor code	31.30	0.81	29.74	32.91	1.87	3 750

Scopus data/ Question number	Variable	Proportion	Standard error	95% co	nfidence	DEFT	Number of responses
Q9	Publication of the new data or code resulting fron publication on a repository	the reference	0.101				(exclusive)
	Only data	25.09	1.10	23.00	27.31	2.07	1 508
	Only code	4.54	0.50	3.66	5.61	1.95	265
	Both data and code	9.68	0.69	8.41	11.13	1.92	589
	Neither data nor code	60.69	1.16	58.39	62.94	1.94	4 313
Q10	Delivery of the new data or code resulting from the	e reference pub	olication to a jo	ournal or pu	ublisher		(exclusive)
	Only data	26.10	1.06	24.08	28.22	1.96	1 605
	Only code	1.47	0.18	1.15	1.87	1.22	108
	Both data and code	7.95	0.87	6.40	9.83	2.62	434
	Neither data nor code	64.49	1.14	62.22	66.69	1.94	4 480
Q11	Sharing of the new data or code resulting from th	e reference pub	lication with o	ther resear	chers		(exclusive)
	Only data	28.99	1.02	27.04	31.02	1.82	1 872
	Only code	5.26	0.42	4.49	6.16	1.54	339
	Both data and code	17.99	0.95	16.21	19.92	2.00	1 100
	Neither data nor code	47.75	1.13	45.55	49.97	1.84	3 271
Q12	Access status and features of the new data or co other researchers	de resulting fror	m the referenc	e publicatio	on with		(multiple: count=yes)
	No fee	90.05	0.82	88.32	91.55	2.24	6 085
	Assignment of digital object identifiers	12.80	0.64	11.59	14.11	1.57	878
	Adoption of a standard access mechanism	23.17	0.87	21.51	24.92	1.68	1 603
	Availability of information online	36.75	1.15	34.52	39.04	1.96	2 325
	Application of an usage licence	14.08	0.99	12.25	16.13	2.32	734
	Use of detailed and comprehensive metadata	24.18	1.00	22.26	26.20	1.91	1 603
	Compliance with standards that facilitate data sources linking	21.38	0.83	19.79	23.06	1.66	1 579
Q13a	Factors influencing open access: formal sharing i	requirements					(exclusive)
	Significantly constrained	26.94	0.86	25.28	28.67	1.92	2 756
	No impact	52.12	1.05	50.07	54.17	2.06	5 020
	Significantly enhanced	20.93	0.86	19.30	22.66	2.07	1 893
Q13b	Factors influencing open access: intellectual prop	erty protection					(exclusive)
	Significantly constrained	23.74	0.95	21.93	25.66	2.18	2 009
	No impact	65.01	1.05	62.93	67.04	2.14	6 471
	Significantly enhanced	11.24	0.79	9.78	12.90	2.45	1 004
Q13c	Factors influencing open access: career objective	es					(exclusive)
	Significantly constrained	8.14	0.55	7.12	9.28	1.95	743
	No impact	71.67	0.97	69.74	73.52	2.07	6 627
	Significantly enhanced	20.19	0.88	18.53	21.97	2.11	1 951
Q13d	Factors influencing open access: norms/peer exp	ectations					(exclusive)
	Significantly constrained	13.06	0.69	11.76	14.47	1.98	1 137
	No impact	59.16	1.06	57.06	61.23	2.09	5 645
	Significantly enhanced	27.78	0.97	25.92	29.71	2.09	2 569
Q13e	Factors influencing open access: resources and o	capabilities					(exclusive)
	Significantly constrained	20.16	1.03	18.22	22.25	2.46	1 780
	No impact	60.67	1.14	58.40	62.89	2.25	5 703
	Significantly enhanced	19.18	0.87	17.54	20.93	2.11	1 764
Q13f	Factors influencing open access: dissemination c	osts					(exclusive)
	Significantly constrained	33.32	1.13	31.14	35.57	2.31	2 970
	No impact	56.08	1.14	53.83	58.31	2.22	5 326
	Significantly enhanced	10.60	0.64	9.40	11.93	2.02	1 005
Q13g	Factors influencing open access: privacy and eth	ics					(exclusive)

THE 2018 OECD INTERNATIONAL SURVEY OF SCIENTIFIC AUTHORS \mid 71

Scopus data/ Question number	Variable Significantly constrained	Proportion	Standard error	95% cor inte	nfidence rval	DEFT	Number of responses
		20.55	1.11	18.46	22.81	2.64	1 517
	No impact	69.44	1.15	67.15	71.65	2.39	6 698
	Significantly enhanced	10.01	0.59	8.91	11.22	1.89	1 017
	Use of digital tools section						
214	Use of digital tools						(multiple: count=yes)
	Literature searches	89.93	0.46	88.99	90.80	1.53	9 066
	Connect with stakeholders	32.07	0.81	30.50	33.69	1.75	3 865
	Search for/Order materials	54.82	1.09	52.67	56.94	2.19	5 753
	Data retrieval	40.86	1.00	38.92	42.82	2.03	4 277
	Connect with other researchers	68.91	1.03	66.86	70.88	2.22	7 388
	Peer reviews	71.24	1.10	69.04	73.34	2.43	7 747
	Search for funding opportunities	45.35	0.93	43.53	47.19	1.88	5 279
	Crowdfunding	2.65	0.38	1.99	3.51	2.39	280
	Conduct/Outsource experiments	26.21	0.90	24.48	28.00	2.05	2 395
	Manage labs	25.68	1.02	23.72	27.74	2.35	2 287
	Share notebooks, protocols or workflows	33.05	0.97	31.18	34.99	2.07	3 285
	Participate in virtual online meetings	44.41	0.89	42.66	46.17	1.80	5 086
	Data collection and processing	58.95	1.07	56.84	61.03	2.18	5 784
	Project management	50.87	1 09	48 74	52 99	2 18	5 139
	Distributed data analysis and visualisation	34.60	1.00	32.67	36.59	2 10	3 467
	Collaborative manuscript writing	65.24	1.00	63.18	67.24	2.10	6 782
	Edit and assist writing	64.09	0.99	62 13	66.00	2.10	6 391
	Manage bibliographic references	55.82	1.08	53.69	57.93	2.00	5 753
	Disseminate/Archive code or protocols	17.48	0.65	16 25	18 78	1 70	1 810
	Archive non digital outputs	12.88	0.00	11 59	14 29	2.05	1 223
	Monitor/Benchmark use of research outputs	27.88	0.00	26.17	29.65	1.98	2 902
	Archive/Share documents other than articles	15.69	1.02	/3 70	17.68	2.04	1 712
	Communicate research findings to the public	38.18	0.08	36.20	40.11	2.04	4 7 12
	Communicate research findings to the public	JU 34	1.05	12 20	40.11	2.01	4 230
	Archive manuscripts and proprints	44.04	1.05	42.23	40.41	2.12	4 000
	Archive manuscripts and preprints	45.20	1.04	41.20	40.00	2.10	4 090
	Develop online personal or team profiles on	19.04	1.03	20.90	00.90 26 E4	2.02	0 291
	research work	34.00	0.90	32.71	30.94	2.00	3 005
	Disseminate/Archive data	34.54	0.96	32.69	36.44	2.01	3 825
	Research methods section						
15	Research methods						(multiple: count=yes
	Gathering, digitising or curating information	38.87	1.06	36.82	40.97	2.13	3 602
	Formulating and testing hypotheses in experimental setting	53.16	1.05	51.10	55.22	2.07	4 793
	Formulating and testing hypotheses in empirical, non-experimental setting	41.54	1.02	39.55	43.56	2.04	4 334
	Formulating theories and studying their properties	46.26	1.08	44.16	48.38	2.12	4 445
	Using computational modelling and simulation methods	39.70	0.99	37.78	41.65	1.98	3 613
16a	Use or development of conventional data						(exclusive
	Use	36.42	1.14	34.23	38.68	2.28	3 302
	Develop	7,20	0.50	6.28	8.23	1.86	647
	Both use and develop	25.98	0.89	24.27	27.76	1.97	2 434
	Neither use and develop	30.40	0.99	28.50	32.37	2.08	2 982

$72 \mid$ the 2018 OECD international survey of scientific authors
Scopus data/ Standard 95% confidence Number of Variable Proportion DEFT Question number error interval responses Q16b Use or development of big data (exclusive) Use 12.47 0.85 10.91 14.23 2.39 830 3.73 0.40 3.02 4.59 1.96 277 Develop Both use and develop 7.36 0.59 6.29 8.60 2.10 537 Neither use and develop 76.43 0.98 74.45 78.31 2.16 7 022 Q16c Use or development of computational methods, processes and systems (exclusive) Use 35.29 1.15 33.07 37.57 2.29 2 986 7.10 8.02 632 Develop 0.44 6.28 1.65 Both use and develop 19.13 0.76 17.68 20.66 1.84 1713 3 7 4 7 38.48 1.04 36.47 40.54 2.04 Neither use and develop Q16d Use or development of connected sensors to collect information (exclusive) Use 16.35 0.87 14.72 18.14 2.19 1 235 Develop 3.38 0.35 2.75 4.15 1.81 260 Both use and develop 6.36 0.51 5.44 7.44 1.93 532 Neither use and develop 73.90 0.98 71.93 75.78 2.08 6 588 Q16e Use or development of programmable machines that execute tasks autonomously (exclusive) 7.64 10.32 2.22 Use 8.89 0.68 621 Develop 2.20 0.25 1.76 2.75 1.59 189 Both use and develop 4.19 0.41 3.46 5.07 1.89 356 84.72 0.80 83.09 86.23 2.06 7 4 1 7 Neither use and develop Q16f Use or development of participative networks for securing data from a range of individual actors (exclusive) Use 10.69 0.70 9.40 12.13 2.09 877 Develop 2.17 0.42 1.49 3.15 2.64 167 Both use and develop 308 3.92 0.47 3.10 4.96 2.24 Neither use and develop 83.22 0.88 81.41 84.88 2.19 7 216 Q17 (multiple: Type of data count=yes) 29.55 Collected from statistical offices 27.66 0.95 25.85 2.02 2,838 Collected and owned by private companies 15.42 0.71 14.07 16.87 1.88 1 501 80.84 0.92 78.98 82.57 2.22 7 4 1 3 Collected by yourself or your team 4 688 Shared by colleagues or peers 51.32 1.11 49.14 53.49 2.12 Personal data 31.91 1.13 29.72 34.17 2.32 2 958 From social networks 11.19 0.67 9.95 12.56 2.02 1074 0.91 Collected and owned by governments 26.55 24.80 28.38 1.97 2 672 Administrative data 17.39 0.78 15.91 18.97 1 648 1.96 Q18 Most important challenge (exclusive) Access to the right set of skills 25.28 0.84 23.67 26.95 1.83 2 387 Access to the right set of data 33.07 1.07 31.01 35.20 2.16 2 995 Access to the right infrastructure and tools 34.07 32.14 36.05 2.00 2 996 1.00 Other 7.58 0.94 5.92 9.65 3.39 636 Q19 Most important skills (exclusive) 38.76 1.00 36.83 40.74 1.97 3 776 Project management Advanced programming 32.40 0.99 30.48 34.38 2.05 2 823 25.69 24.11 27.34 2711 Project definition 0.82 1.82 Knowledge of legal aspects 11.49 1.03 9.62 13.66 3.11 857 50.98 4 7 4 7 Data collection, curation and preservation 53.06 1.06 55.14 2.05 Q20 Most important infrastructure (max 2 terms) Cloud services 9.13 0.60 8.03 10.37 2.00 996 Computer hardware 20.63 0.95 18.82 22.56 2.28 1 454

8.09

0.61

6.97

9.37

2.17

558

THE 2018 OECD INTERNATIONAL SURVEY OF SCIENTIFIC AUTHORS | 73

OECD SCIENCE, TECHNOLOGY AND INDUSTRY WORKING PAPERS

Security services

Scopus data/	Variable	Proportion	Standard	95% confidence		DEFT	Number of
Question number	Infrastructure for high performance computing	20 14	0.88	18 47	21.92	2 12	1 820
	Internet or other computer networks connectivity	56.69	1.08	54.55	58.80	2.11	5 723
	Software	51.35	1.06	49.27	53.43	2.06	4 972
	Storage capacity	10.92	0.56	9.88	12.06	1.72	1 007
	Digital identify and outputs section						
Q21	Identifiers						(multiple
-							possible)
	Author IDs or personal profiles associated to citation indices	49.37	1.10	47.22	51.53	2.11	4 890
	IDs maintained by research funding organisations or authorities in the country	18.01	0.78	16.54	19.58	1.94	1 826
	ORCID	68.65	1.05	66.55	70.68	2.18	6 476
	None	15.01	0.78	13.54	16.61	2.10	1 308
	Other	1.56	0.15	1.29	1.90	1.20	203
Q22	Channels used to provide information on research	work online					(multiple)
	Generic social network accounts	34.59	0.99	32.67	36.57	1.96	3 358
	Individual webpage or site within the organisation	47.95	1.06	45.87	50.03	2.00	4 589
	Specialised academic profile or networking site	81.98	0.90	80.15	83.67	2.19	7 401
	Individual webpage created and maintained by you	20.41	0.81	18.87	22.06	1.90	1 818
Q23a	Information on academic qualifications, work expe	rience					(exclusive)
	Do not know if available online	3.80	0.44	3.03	4.76	2.18	317
	Not relevant or applicable	5.02	0.54	4.07	6.19	2.34	357
	Relevant but information not available online	9.96	0.53	8.97	11.04	1.67	856
	Information available online	81.22	0.79	79.62	82.72	1.92	7 481
Q23b	Information on working papers and peer-reviewed	publications					(exclusive)
	Do not know if available online	2.18	0.25	1.73	2.74	1.65	188
	Not relevant or applicable	3.07	0.41	2.36	3.98	2.25	216
	Relevant but information not available online	10.41	0.66	9.18	11.78	2.06	944
	Information available online	84.34	0.78	82.75	85.81	2.04	7 653
Q23c	Information on books or book chapters						(exclusive)
	Do not know if available online	4.87	0.39	4.15	5.69	1.70	371
	Not relevant or applicable	26.11	0.85	24.47	27.81	1.81	1 944
	Relevant but information not available online	17.28	0.73	15.90	18.75	1.80	1 619
	Information available online	51.75	0.99	49.80	53.69	1.86	4 830
Q23d	Information on conference participation or organis	ation					(exclusive)
	Do not know if available online	7.02	0.99	5.31	9.22	3.62	520
	Not relevant or applicable	8.06	0.66	6.86	9.44	2.26	613
	Relevant but information not available online	23.73	0.84	22.13	25.42	1.85	2 085
	Information available online	61.19	1.15	58.91	63.43	2.22	5 578
Q23e	Information on editorial work						(exclusive)
	Do not know if available online	10.33					773
	Not relevant or applicable	33.08					2 631
	Relevant but information not available online	27.82					2 192
0.001	Information available online	28.78					2 909
Q23t	Information on databases and computer code	<u> </u>	o	0.0=	40.55	/ =-	(exclusive)
	Do not know it available online	9.34	0.57	8.27	10.53	1.78	726
	Not relevant or applicable	42.73	1.08	40.62	44.87	1.98	3 /31
	Relevant but information not available online	19.20	1.06	26.70 17.04	30.84 21.56	2.11	2 250

$74 \mid$ the 2018 oecd international survey of scientific authors

THE 2018 OECD INTERNATIONAL SURVEY OF SCIENTIFIC AUTHORS	75

Scopus data/ Question number	Variable	Proportion	Standard error	95% confidence interval		DEFT	Number of responses
Q23q	Information on teaching, training or mentoring a	ctivity					(exclusive)
	Do not know if available online	6.00	0.44	5.19	6.93	1.73	480
	Not relevant or applicable	16.89	0.81	15.37	18.54	2.00	1 367
	Relevant but information not available online	33.69	1.15	31.47	35.98	2.26	2 768
	Information available online	43.42	1.11	41.24	45.61	2.09	3 992
Q23h	Information on patented inventions applied for o	r granted					(exclusive)
	Do not know if available online	7.01					559
	Not relevant or applicable	58.04					5 447
	Relevant but information not available online	10.51					871
	Information available online	24.44					1 432
Q23i	Information on contract research or other consu	Itancy work					(exclusive)
	Do not know if available online	8.26	0.68	7.02	9.68	2.25	617
	Not relevant or applicable	49.03	1.14	46.79	51.28	2.09	4 079
	Relevant but information not available online	24.61	0.85	22.98	26.31	1.80	2 125
	Information available online	18.11	0.90	16.41	19.94	2.14	1 540
Q23j	Information on entrepreneurship activity						(exclusive)
	Do not know if available online	7.47					563
	Not relevant or applicable	70.13					5 890
	Relevant but information not available online	10.09					723
	Information available online	12.31					1 042
Q23k	Information on management or advisory board r	nembership					(exclusive)
	Do not know if available online	6.33					479
	Not relevant or applicable	50.11					4 040
	Relevant but information not available online	17.24					1 443
	Information available online	26.32					2 441
Q23I	Information on mainstream and social media co	verage					(exclusive)
	Do not know if available online	9.22					809
	Not relevant or applicable	46.81					3 889
	Relevant but information not available online	11.86					952
	Information available online	32.10					2 616
Q23m	Information on public service or political work						(exclusive)
	Do not know if available online	9.59					700
	Not relevant or applicable	62.24					5 278
	Relevant but information not available online	13.97					1 024
	Information available online	14.19					1 194
Q23n	Information on societal engagement						(exclusive)
	Do not know if available online	10.17					730
	Not relevant or applicable	67.90					5 778
	Relevant but information not available online	12.68					942
	Information available online	9.25					731
Q24	Indicators of research outputs						(maximum 3 items)
	Prestige or citations of the journal	61.54	1.16	59.24	63.80	2.24	5 655
	Citations by peers	62.66	1.15	60.38	64.90	2.24	5 661
	Counts of outputs	57.42	1.14	55.17	59.63	2.16	5 067
	Online use indicators (e.g. views)	20.32	0.85	18.72	22.03	1.97	1 795
	Qualitative methods	26.26	0.83	24.67	27.92	1.77	2 453
	Context-specific indicators	9.76	0.69	8.49	11.21	2.19	858
Q25	Decisions informed by the use of quantitative in	dicators of resea	rch outputs				(multiple)
	Payment for services provided	16.10	0.90	14.41	17.94	2.25	1 171

Scopus data/ Question number	Variable	Proportion	Standard error	95% confidence interval		DEFT	Number of responses
	Project funding	63.13	1.05	61.05	65.17	2.01	5 356
	Hiring or job promotion	56.57	1.06	54.49	58.62	1.96	4 788
	Possibility to join research teams	68.11	0.99	66.14	70.01	1.94	5 928
	Selection for managerial or advisory roles	33.80	0.93	32.00	35.64	1.80	3 171
	Career information section						
Q27	Education level						(exclusive)
	Doctoral degree or equivalent	84.13	0.81	82.47	85.66	2.09	7 645
	Master's degree or equivalent	15.18	0.80	13.67	16.83	2.10	1 106
	Other	0.69	0.14	0.47	1.02	1.54	42
Derived from Q28	Academic age (years since first indexed publica	tion)					(exclusive)
	Up to 5	33.54	1.00	31.61	35.52	2.00	2 697
	6 to 10	19.58	0.88	17.91	21.35	2.09	1 739
	11 to 20	24.40	1.11	22.29	26.65	2.45	2 297
	21 to 40	19.20	0.74	17.80	20.69	1.77	1 923
	41 or more	3.28	0.23	2.86	3.77	1.24	324
Q31 and Q32	Employment status						
	Employed	90.19	0.69	88.74	91.47	2.21	8,268
	Not employed						(exclusive)
	Of which, inactive	47.53					382
	Of which, unemployed	52.47					279
Q33	Main reasons for not working						(exclusive)
	Retired	51.03					234
	Other	48.97					124
Q34	Time of the last job						(exclusive)
	Never worked	23.96					86
	Less than 1 year ago	27.33					199
	Between 1 and 5 years ago	30.91					251
	More than 5 years ago	17.80					162
Q35	Research work as part of the job core responsibilities/remuneration	82.46	0.83	80.77	84.04	2.08	7 637
Q36	Need to compute for resources within the organisation	64.58	0.95	62.68	66.42	1.89	5 758
Q37	Need to compete for and secure funding from external organisation	71.47	0.85	69.77	73.12	1.79	6 559
Q38	Average working hours per week						(exclusive)
	Less than 0	1.39	0.16	1.11	1.75	1.30	120
	10-14	2.03	0.22	1.65	2.50	1.46	179
	15-19	1.07	0.15	0.81	1.42	1.41	118
	20-24	2.11	0.22	1.71	2.59	1.47	189
	25-29	1.94	0.27	1.47	2.56	1.87	152
	30-34	3.64	0.44	2.87	4.60	2.21	303
	35-39	9.49	0.48	8.59	10.47	1.56	971
	40-44	28.43	0.99	26.53	30.41	2.08	2 656
	45-49	16.12	0.69	14.80	17.53	1.79	1 600
	50-54	14.63	1.11	12.58	16.94	2.98	1 200
0.00	More than 55	19.15	0.89	17.47	20.95	2.14	1 514
Q39	Percentage of weekly working time spent on res	earch					(exclusive)
	0-9%	4.91	0.39	4.20	5.74	1.71	418
	10-19%	6.86	0.39	6.13	7.66	1.46	664

$76 \mid$ the 2018 OECD international survey of scientific authors

95% confidence Scopus data/ Standard Number of DEFT Variable Proportion Question number error interval responses 10.40 20-29% 0.67 9.17 11.78 2.06 996 30-39% 10.99 0.65 12.34 1 0 3 1 9.78 1.97 40-49% 9.31 0.51 8.35 10.36 1.67 946 10.69 1 138 50-59% 11.87 0.63 13.17 1.84 60-69% 10.67 9.36 12.13 2.16 973 0.71 1 140 70-79% 14.43 1.07 12.46 16.65 2.86 80-89% 10.76 0.66 9.53 12.12 2.01 837 90-100% 9.80 0.66 8.59 11.17 2.08 760 Q40 Holding an indefinite contract 0.55 6.91 441 Not relevant 5.73 4.73 2.25 41.85 39.79 43.95 3 367 No 1.06 2.04 Yes 52.42 1.12 50.23 54.60 2.11 5 1 4 7 Q41 Possibility of being dismissed for reasons other than gross misconduct (exclusive) Not relevant 6.00 0.55 5.01 7.17 2.17 561 No 61.82 0.98 59.87 63.73 1.91 5 398 Yes 32.18 0.87 30.51 33.90 1.75 2 955 Q43 Holding a secondary occupation 15.17 0.61 14.01 16.41 1.61 1 697 Q45 and Q46 Sector of employment (exclusive) 74.54 1.06 72.40 76.56 2.30 6849 Public sector 18.36 0.64 17.13 19.65 1.56 1,757 For-profit organisation Q47 Main area of the employer's activity (exclusive) Agriculture, forestry, fishing, mining and 2.86 0.28 2.36 3.46 1.58 280 quarrying Manufacturing 1.56 0.23 1.18 2.08 1.72 117 Electricity, gas, steam, A/C, water supply, waste 1.10 0.26 0.69 1.73 2.31 68 management, construction Wholesale and retail trade, rapid, transportation, 0.24 0.06 0.15 0.40 1.19 21 storage, accommodation, food services 1.63 0.34 2.97 2.21 2.16 167 Information and communication Financial and insurance activities, real estate 0.35 0.11 0.19 0.63 1.69 34 Professional, scientific and technical activities 16.33 0.70 15.00 17.75 1.78 1 421 Administrative and support activities 0.30 0.08 0.18 0.52 1.43 28 Public administration and defence 1.39 0.14 1.13 1.70 1.15 144 Education (other) 5.75 0.47 4.89 6.73 1.89 492 5 4 3 0 Higher education 57.61 1.09 55.45 59.74 2.08 Human health and social work activities 10.07 0.72 8.74 11.57 2.25 613 Arts, entertainment and recreation, other 0.24 0.07 1.33 27 0.14 0.42 services Q48 Annual income (exclusive) 8.36 0.69 9.80 I prefer not to answer 7.11 2.27 590 Under \$15 000 13.27 0.63 12.08 14.55 1.70 1 1 3 9 \$15 000-\$29 999 15.54 1.21 13.32 18.06 3.05 1 288 \$30 000-\$44 999 12.81 0.87 11.19 14.61 2.39 1 1 1 2 \$45 000-\$59 999 11.90 0.67 10.65 13.28 1.89 1 0 4 3 \$60 000-\$74 999 8.76 0.45 7.91 9.68 1.46 829 7.69 694 \$75 000-\$89 999 0.44 6.87 8.59 1.51 6.64 539 \$90 000-\$104 999 0.41 5.88 7.50 1.51 \$105 000-\$119 999 4.45 0.46 3.63 5.45 2.06 369 \$120 000-\$149 999 4.27 0.30 3.72 4.90 1.36 369 \$150 000-\$199 999 3.59 0.38 2.91 4.43 1.90 247 0.25 2.27 1.42 \$200 000 or more 2.72 3.26 187

THE 2018 OECD INTERNATIONAL SURVEY OF SCIENTIFIC AUTHORS 77

$78 \mid$ The 2018 OECD International survey of scientific authors

Scopus data/ Question number	Variable	Proportion	Standard error	95% confidence interval		DEFT	Number of responses
	Mobility						
Derived from Q3 and Q4	Proportion of authors that are born in a country different from their country of residence	29.93	0.58	28.82	31.08	1.37	3 966
Derived from Q29 and Q4	Proportion of authors that studied in a country different from their country of residence	22.36	1.08	20.31	24.56	2.45	2 284
Q49	Plan to move to another country	9.14	0.57	8.09	10.31	1.85	768

Note: Author-level sampling weights adjusted for nonresponses are used. Figures referring to question 9 onwards are also adjusted for item nonresponses. *Source:* OECD International Survey of Scientific Authors (ISSA), 2018. <u>http://oe.cd/issa</u>.