Note to Delegations:

This document is also available on O.N.E under the reference code:

DSTI/STP/NESTI(2019)2/FINAL

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), participants at the OECD Applied Economics Work-in-Progress Seminar, CSTP, Global Science Forum (GSF) and NESTI delegates for comments and advice. Within 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 3

Abbreviations 7

Charting the digital transformation of science: Findings from the 2018 OECD International Survey of Scientific Authors (ISSA2) 8

Executive summary 9

The study 9

Key findings 9

Implications for further work 12

1. Introduction 13

2. Measuring the digitalisation of science: the ISSA2 study 15

3. A multidimensional approach to measuring the digitalisation of science 17

4. Factors influencing digitalisation of science 45

5. Drivers and impacts of digitalisation in science 49

6. Concluding remarks 61

References 63

Annex A. Benchmark bibliometric indicators on digitalisation 65

Annex B. Factor analysis of digitalisation patterns 67

Annex C. Economy and country-level indicators 70

Annex D. Detailed regression analysis results 77

Tables

Table 2.1. Distribution of ISSA2 survey participants 15

Table B.1. Description of the questionnaire-based variables used in the factor analysis 67
Table B.2. Factor variance 69
Table B.3. Factor score evaluation tests 69
Table A D.1. Links between digitalisation patterns and author profiles 77
Table D.2. Links between digitalisation patterns and author research methods 79
Table D.3. Drivers of authors' perceived challenges in their research work 81
Table D.4. Factors explaining respondent's reporting of most important research skills 82
Table D.5. Factors explaining reporting of most important research infrastructure 84
Table D.6. Drivers of probability of engagement in activities beyond core research 86
Table A D.7. Link between research quality and author profile 88
Table D.8. Link between scientific authors' income and their profiles 89
Table D.9. Authors' attitudes towards digitalisation and its potential impacts 91
Table D.10. Authors' attitudes towards digitalisation and their profiles, continued 93

Figures

Figure 3.1. Research methods by science field 18
Figure 3.2. Survey-based measures of open access to reference publication, by science field 19
Figure 3.3. Survey-based measures of open access to reference publication 20

Figure 3.4. Scientific production resulting in new data or code, by science field ... 22
Figure 3.5. Dissemination channels for data and code resulting from reference publication ... 23
Figure 3.6. Dissemination channels for data and code, by country of residence ... 23
Figure 3.7. "FAIR" quality features of the data or code resulting from reference publication... 24
Figure 3.8. Incidence of fee payment requirements to access data or code, by science field ... 24
Figure 3.9. Factors affecting the level of access granted to research outputs ... 25
Figure 3.10. Use of digital tools by research activity ... 26
Figure 3.11. Use and development of advanced digital tools or methods, by science field ... 27
Figure 3.12. Use and development of advanced digital tools, by country of residence... 28
Figure 3.13. Type of data used by scientific authors ... 29
Figure 3.14. Use of identifiers (IDs) by scientific authors to track their research work... 30
Figure 3.15. Use of identifiers (IDs) by scientific authors, by sector ... 30
Figure 3.16. Use of identifiers (IDs) by scientific authors, by country of residence ... 31
Figure 3.17. Channels used for online dissemination of information about research... 32
Figure 3.18. Channels used for online dissemination of research information, by country ... 32
Figure 3.19. Engagement in research-related activities, by science area... 33
Figure 3.20. Engagement in research-related activities, by country of residence ... 34
Figure 3.21. Engagement in research-related activities, by sector of employment ... 35
Figure 3.22. Online availability of information on research-related activities, by science area ... 36
Figure 3.23. Use of research metrics, by type and science field... 37
Figure 3.24. Decisions informed by quantitative indicators of research, by science field ... 37
Figure 3.25. Decisions informed by quantitative research indicators, by country of residence ... 38
Figure 3.26. Decisions informed by quantitative indicators of research, by sector ... 38
Figure 3.27. Correlation of digitalisation factors with question items ... 40
Figure 3.28. Patterns of digitalisation in science, by field... 43
Figure 3.29. Patterns of digitalisation in science, by country of residence... 44
Figure 4.1. Links between digitalisation patterns and author profile ... 46
Figure 4.2. Links between measures of digital intensity and research methods ... 48
Figure 5.1. Most important challenge faced by scientific authors in their research work ... 50
Figure 5.2. Most important challenge faced by scientific authors, by country of residence ... 50
Figure 5.3. Most important skills for scientific authors' research work... 51
Figure 5.4. Most important skills for scientific authors' research work, by country... 51
Figure 5.5. Most important infrastructure for scientific authors' research work, by field... 52
Figure 5.6. Most important infrastructure for scientific authors' research work, by country... 52
Figure 5.7. Authors' perceived challenges in their research work and digitalisation patterns ... 53
Figure 5.8. Most important skills and digitalisation patterns ... 54
Figure 5.9. Most important infrastructure and digitalisation patterns ... 55
Figure 5.10. Drivers of lifetime citation impact and journal prestige... 56
Figure 5.11. Engagement in activities beyond core research and digitalisation patterns ... 57
Figure 5.12. Authors' earnings and their profiles... 58
Figure 5.13. Scientific authors' views on the digitalisation of science and its potential impacts... 59
Figure 5.14. Authors' attitudes towards digitalisation and their profiles ... 60

Figure A.1. Open access journal publishing by science fields, 2017, ISSA2 sampling frame ... 65
Figure A.2. Open access journal publishing, selected economies, 2017, ISSA2 sampling frame... 65
Figure A.3. AI-related scientific document content, by fields, 2017, ISSA2 sampling frame ... 66
Figure A.4. AI-related scientific document content, selected economies, 2017, ISSA2 sampling frame ... 66
Figure C.1. Scientific production resulting in new data or code, by country or economy of residence ... 70
Figure C.2. Absence of fee payment requirements to access to data or code... 71
Figure C.3. Open access to scientific publications, by an author's country or economy of residence ... 71
Figure C.4. Use and/or development of advanced digital practices, by country or economy of residence ... 72
Figure C.5. Use of identifiers (IDs) by scientific authors to track his/her research work, by country or economy of residence ... 72
Figure C.6. Channels used for disseminating information about research, by country or economy of residence ... 73

Figure C.7. Engagement in research-related activities by country or economy of residence ... 73
Figure C.8. Online availability of information on research-related activities by country or economy of residence ... 74
Figure C.9. Use of research metrics, by type and country or economy of residence ... 74
Figure C.10. Decisions affected by the use of quantitative research indicators, by country or economy of affiliation ... 75
Figure C.11. Challenges faced by scientific authors in their research work ... 75
Figure C.12. Most important skills for scientific authors' research work ... 76
Figure C.13. Most important infrastructure for scientific authors' research work ... 76

Boxes

Box 3.1. Evaluating the degree of determinacy/indeterminacy of the factor scores ... 41

Abbreviations

Science fields code description

Code	Field
AGRBIO	Agricultural and Biological Sciences
ALL	**All fields combined**
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

The abbreviated field codes presented above are used throughout this publication.

Charting the digital transformation of science: Findings from the 2018 OECD International Survey of Scientific Authors (ISSA2)

Michela Bello and Fernando Galindo-Rueda

This paper presents the results of the 2018 OECD International Survey of Scientific Authors (ISSA2), a global online survey designed and implemented to measure the key features of the digital transformation of science. The paper explores the potential impacts of digitalisation based on a combination of different indicators on research impact and responses from nearly 12 000 authors across the world. The evidence shows that although digital activity is pervasive, the transformation is uneven across fields and sectors, and is influenced by factors such as norms, experience, skills and data availability. Overall, scientists appear to be optimistic about the potential of digitalisation, especially in relation to the efficiency of research and collaboration across national borders. This paper is also the first analysis to leverage a new OECD approach to data collection in priority science policy topics for which evidence might be scarce or insufficiently timely.

Executive summary

About the study

This paper presents the results of the 2018 OECD *International Survey of Scientific Authors* (ISSA2), a global online survey that examined key features of the digital transformation of science. ISSA2 was designed and implemented with the intent to support statistical analysis on the nature and effects of digitalisation in scientific research. The study was set out to contribute 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. ISSA2 was designed to address the priorities set out in the Programme of work of the OECD Committee for Scientific and Technological Policy (CSTP), as well as in the OECD Blue Sky Agenda (OECD, 2018) adopted by the CSTP's Working Party of National Experts in Science and Technology (NESTI).

The study targeted the corresponding authors of scientific publications whose contact information is available in a large global bibliographic database. A sample of scientific authors listed as corresponding authors were invited by email to participate in an online survey conducted directly by the OECD and were asked to report on their use of a broad range of digital tools and related practices, in addition to other key demographic and career information. Responses were collected for a total of approximately 12 000 scientific authors from all over the world and across all disciplinary areas, representing to a varying extent the subset of the research population engaged in scholarly publication work, including those in the business sector. Although suffering from a significantly high non-response rate, this is comparable to other online surveys and the study's quality checks suggest that the results, through reweighting adjustments, can be considered representative of the target population for the majority of countries and economies covered.

This paper's output tables, the underlying anonymised data and software code required to replicate the analysis can be obtained by following the indications provided on the OECD ISSA website at http://oe.cd/issa.

Key findings

The survey results provide a rich snapshot picture of the multi-faceted nature of science and digitalisation, providing a baseline for charting its digital transformation and the mechanisms through which it influences scientific research and its impacts on society. Rather than pointing to a single composite index of digitalisation, the analysis of responses across a broad set of question items identifies four distinct facets of digitalisation in scientific research:

- the adoption of digital scientific collaboration and productivity tools throughout all stages of the scientific process;
- the digitally-enabled diffusion and access to data and code;
- the use of advanced and data-intensive digital tools to gain insights and develop predications;
- the development of digital identity and online communication of scientific work.

The characterisation of these distinctive facets proves effective at mapping how different fields, sectors and even countries score in relation to major digitalisation paradigms. The results suggest considerable room for greater utilisation of digital opportunities across all of them, while taking into account the evidence provided in this and other studies on the incentives faced by scientists.

Digitalisation and the open science paradigm

In nearly 70% of cases, data and code are by-products of research that resulted in scientific publications. Although digital technology facilitates the sharing of scientific outputs, informal person-to-person sharing mechanisms are still dominant, while less than 50 percent of scientific authors deposit their data or code on a repository or share it with the publisher as supporting information. Data and code sharing practices may often be ineffective as adequate information to reuse them is rarely available. Less than 30% of authors carry out activities that contribute towards assuring findability, accessibility, interoperability and reusability of their data (FAIR principles of Open Science). This stands in contrast with the access status of scientific publications, which are estimated to be openly and freely available in nearly 70% of cases according to authors' responses.

Data sharing practices are field dependent and influenced positively by social norms and peer expectations, while dissemination costs, IPR protection and privacy and ethics considerations are often identified as *the* major obstacles. Data sharing practices are most widely diffused in the combined area of biochemistry, genetics, molecular biology, immunology and microbiology, and less common in the arts and humanities, social sciences and mathematics.

Across many fields, it is striking to note the rather limited uptake of relevant digital practices amongst scientists whose work entails the development and curation of data. It is therefore surprising that scientists across all fields coincide in pointing most frequently to data collection and curation skills as most important skills for their scientific work, above programming and project management skills.

Digitalisation and the new paradigm of data-intensive research

The survey results characterise scientific activities relating to big data, robotics and the Internet of things in addition to the use and development of advanced computational tools. The use of data intensive advanced digital tools is more widespread in computer sciences, materials science, engineering, and earth and planetary sciences. The research activities most strongly associated to this mode of digitalisation are those that involve computational modelling as well as those that entail the testing of hypotheses in experimental settings.

Scientific authors in People's Republic of China (hereafter 'China') and India exhibit profiles that are highly oriented towards this mode of research, consistently with a separate analysis of AI intensity of the content of scientific publication abstracts. Scientists in the business sector display higher average scores on this mode of digitalisation compared to scientists in other sectors, in particular higher education. Scientists with experience in the use of advanced digital tools are less likely to have attained a doctorate degree and are more likely to report access to data as a major challenge to research.

Digitalisation transforming research identity and collaboration

Digitalisation is also transforming research in ways that are common to other domains of social interaction but are very pronounced since the scientific community is highly interconnected globally. The results show that the use of digital productivity and collaboration tools is more intense within Higher education institutions and small economies, and it is associated with higher rates of personal scientific collaboration and productivity.

A significant number of scientific authors are actively engaged in defining their online identity to assert links to their work and communicate their research beyond conventional channels. This type of digitalisation mode is particularly prevalent in the Higher education and Non-profit sectors as well as within the fields of arts and humanities, the social science and agricultural and biological sciences.

Online identity is important for researchers because a significant but incomplete part of their science-related activities leave a digital trace. In turn, this digital footprint supports the production of metrics that researchers use to inform decisions that impact on their research careers, such as hiring and promotion decisions in universities. They do note, for example, that journal prestige indicators are widely used, more so than indicators of actual citation impact or even qualitative indicators. A significant number of authors point to the growing use of online usage indicators (26% of authors in the Pharma and related fields).

The ORCID identifier appears to be the most widely used identification mechanism by scientists to assert online identity, especially in the Higher Education and Government sector, where it has become a de facto standard. Adoption within the Business enterprise sector is much more limited even among those who publish in scholarly journals and conference proceedings. National identifiers are also reported in some instances but they appear to fail to attain the scale of global-born approaches such as ORCID or even those used by publishers.

Digitalisation divides in science and implications

In general, challenges faced by authors in the digital era concern principally access to data and infrastructure, including basic Internet connectivity. The perception of this challenge increases with the relevance of data for research.

This paper finds evidence of a marked digital divide across gender and age groups. Female authors are less involved than their male counterparts in the use of advanced tools and in data/code sharing practices. However, they are more likely to report engagement in activities contributing to their digital online identity and communication. Younger scientists also exhibit higher digitalisation scores across the various dimensions identified.

There is no evidence of an earnings premium from using advanced digital tools or engaging in data sharing, whereas a more intensive use of digital tools seems to be correlated to a greater involvement in activities beyond core research. For instance, scientists with a more intense use of advanced digital tools are more likely to be involved in business / managerial activity (e.g. start-ups) or to apply or register for IPR protection. However, a major driver of reported incomes is the lifetime average prestige of the journals where scientific authors have published, more so than the average normalised citation record. This is a poignant reminder of the value that the scientific market assigns to the decisions made by editorial boards in journals over and above their actual citation influence, and it points to entrenched patterns explaining the incentives facing by researchers at different points in their careers when considering potential policy reforms, for example, around open science practices.

Overall, scientists seem to have a positive view of the impact of digitalisation in science, and especially on the efficiency of research work and collaboration across national borders. Younger researchers are more sceptical about its effects on research valuation and rewarding systems.

Implications for further work

This paper does not only provide important insights into the digitalisation of science, but also showcases the potential of a brand new approach to data collection from an OECD perspective in priority science policy areas when there is scarce empirical evidence to substantiate the debate.

The ISSA instrument has the potential to become an important tool to measure the impacts of scientific work and investigate the ways in which scientific process takes place. While this approach is not expected to replace existing and more established science and research indicators, it can be particularly useful to study topics and issues for which standard approaches do not offer satisfying solutions. The ability to collect granular behavioural data on scientists, combined with other sources such as bibliometric data, coupled to the flexibility to easily adapt the tool based on emerging research and policy needs are among its major advantages.

Taking into account the room for improvements, the ISSA approach can provide a basis for distributed data collection within countries, which, if necessary, can also target a different population (such as that of researchers), or the development of a more consolidated data infrastructure within the OECD to use for the statistical analysis of major science and research policy questions. Finally, the process leading to the identification and prioritisation of survey topics can also provide an important opportunity to strengthen communication within the OECD, between policy and statistical communities working on science issues, as well as between the OECD and the global scientific community. The dissemination of the anonymised microdata for research purposes seeks to support this vision.

1. Introduction

The development and adoption of digital technologies is transforming science through all stages of the scientific process, from agenda setting and experimentation to research output communication and evaluation. Several studies and anecdotal evidence suggest that scientific activity has not only been transformed – like all other areas of economic activity – by the uptake of productivity and collaborative ICT tools but has also become increasingly data-intensive. Indeed, IT infrastructures and software tools are allowing the collection and use of large amounts of data in scientific research. The use of ICTs is also enabling a deeper and more efficient analysis of the data permitting the investigation of new and more complex research questions as well as the adoption of new, data-driven research methods. This has led some observers to argue that the techniques and technologies for such data-intensive science are so fundamentally different that a "fourth paradigm" for scientific exploration, distinct from empirical, theoretical and computational approaches to sciences, has emerged (Hey et al., 2009). This new paradigm raises both opportunities and challenges for the established processes by which hypotheses are developed and scientific facts are established.

The use of digital tools in science has co-evolved with the emergence of digitally native journals, repositories, pre-print servers as well as new tools for metadata and publications of intermediate research outputs and pieces of information with important implications on different aspects of science. In many instances, scientific publications and underlying data are becoming freely and openly available, as part of a drive towards "open science" paradigm to make the entire scientific process more open and inclusive to all relevant actors (OECD, 2015b). Digital tools facilitate data reuse, generating an expectation of increased research efficiency and reproducibility, but also represent a challenge for traditional appropriation mechanisms and may call for the design of new institutions that incentivise research and its translation into workable solutions.

The digital traceability of research outputs and other related dimensions is another salient feature of scientific research, particularly noteworthy at a time of growing demands for accountability in the use of public funds and interest in developing performance-based incentives for research. What gets recorded and measured has a potentially deep impact on the ways in which research and related activities can be assessed and rewarded. New metrics are being developed, based on information captured not only from administrative systems but also from the Internet, for example, through generic and specialised social media applications. Many of such efforts pursue the objective of capturing additional areas of research activity and impact beyond those traditionally measured (Priem et al., 2010).

Digitalisation has also dramatically changed scholarly and non-scholarly communication. Science blogs and tweets, for instance, are becoming an important source of information and appear to play an increasingly important role in the specialist and public debate. New forms of interactions between researchers and the public are also arising. Digital tools allow in principle for citizens to be not only more informed about science but can also more actively contribute to it by, for instance, participating in the agenda setting, data gathering activities, or providing intellectual efforts. Conversely, heightened communication raises the risks of scientific information being misused or even manipulated.

Policy makers have a strong interest in understanding how digitalisation is transforming the world of science. The tools being used by scientists are affecting research activity and 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 broadening of access to data and knowledge are, for instance, raising concerns about 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 large amounts of data that is becoming more popular in most scientific domains requires scientists and researchers to acquire new skills. This drives policy interest in identifying emerging skill requirements and pathways to promote their development, for example within doctoral training programmes but also in terms of retraining established researchers.

This paper presents the results of a recent OECD survey targeting the scientific community – the OECD International Survey of Scientific Authors (ISSA). This survey was conducted in the last quarter of 2018 and gathered statistical evidence on the use of digital tools and practices in science with the aim to document the nature and effects of digitalisation within the scientific community.

This document is structured in three main parts covering the ISSA2 data analysis and its main findings. Firstly, it provides evidence on the adoption of new digital practices in science based on a descriptive analysis of the survey data. Differences in digitalisation patterns across science fields are illustrated showcasing areas where adoption of digital practices are at an advanced stage and where they may be lagging behind current possibilities. Four new synthetic indicators capturing different levels of adoption of digital tools in science have been derived through factor analysis from the rich set of survey data collected in order to support this characterisation and further analysis. The paper continues by investigating whether different author profiles are associated with different levels and patterns of digitalisation. While a causal relationship cannot be identified from the data collected, this analysis provides valuable insights into the nature of the digital transformation of science and illustrates the factors that are more highly correlated to a greater adoption of digital practices. Finally, the analysis proceeds to explore the implications of the digitalisation of science in terms of scientists' new needs (e.g., infrastructure and skills), performance and research impact and is complemented with survey findings on the ways scientists think that digitalisation will impact on science.

The rest of the document is organised as follows. Section 2 introduces the ISSA2 project and related data. Section 3 presents several descriptive statistics on the use of digital tools in science. Section 4 provides evidence on the link between an author's profile and the adoption of digital practices. Section 5 discusses the impacts on science from digitalisation. Section 6 concludes.

2. Measuring the digitalisation of science: the ISSA2 study

The OECD developed the ISSA2 survey with the intention to contribute 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. This work was carried out in line with the priorities set out in the Programme of work of the OECD Committee for Scientific and Technological Policy (CSTP), as well as the OECD Blue Sky Agenda (OECD, 2018) adopted by the CSTP's Working Party of National Experts in Science and Technology (NESTI).

Interested readers can find a full technical description of the study in a separate technical paper (Bello and Galindo-Rueda, 2020). The ISSA2 study targeted the corresponding authors of scientific publications whose contact information is available in a bibliographic database. This approach builds on the model of a first pilot carried out in 2016 (ISSA1), which focused on the relationship between scientific publishing and open access. In the 2018 ISSA edition, a sample of scientific authors listed as the document's corresponding authors were invited to participate in an online survey conducted directly by the OECD in the last quarter of 2018 and were asked to report on their use of a broad range of digital tools and related practices, in addition to other key demographic and career information. The ISSA2 final dataset contains data on approximately twelve thousand scientific authors worldwide. Table 2.1 illustrates the number of responses across science fields.

Table 2.1. Distribution of ISSA2 survey participants

Field	Number of responses	Share in the total dataset (%)
Multidisciplinary	185	1.55
Agricultural and Biological Sciences	1,069	8.94
Arts and Humanities	451	3.77
Biochemistry, Genetics, Molecular Biochemistry, Genetics, Molecular Biology, Immunology and Microbiology	469	3.92
Business, Management and Accounting	543	4.54
Chemical engineering and Chemistry	379	3.17
Computer Science	928	7.76
Earth and Planetary Sciences	328	2.74
Economics, Finance and Decision sci.	464	3.88
Energy and Environmental Science	554	4.63
Engineering	1,013	8.47
Materials Science	334	2.79
Mathematics	518	4.33
Medicine and Health Professions	1,783	14.9
Neuroscience, Pharmacology, Toxicology	292	2.44
Physics and Astronomy	822	6.87
Social Sciences and Psychology	1,831	15.31
Total	**11,963**	**100**

Note: The indicated fields characterise the journal in which the selected document and corresponding author has been published and correspond to the All Fields Journal Classification assigned by Elsevier.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Detailed information on the survey design and implementation and calculation of the survey weights used in this paper's analysis is provided in the survey's technical annex document. This document also assesses the representativity of the achieved sample and the indicators derived from it, comparing the study's target population to that of researchers, showing that the population of corresponding scientific authors diverges from that of researchers (as defined in the OECD Frascati Manual) in relation to a number of dimensions. When considering the results of this survey, it is important for readers to bear in mind the following key points on the representativeness and interpretability of the results:

- The results are based on data collected from individuals who have produced scientific outputs considered suitable for publication in an extensive list of peer-reviewed journals or conference proceedings. Coverage of a country's scholarly production in global bibliographic indexes is dependent on publishing norms and indexing criteria that relate to disciplinary, language and cultural differences.
- While there is a significant number of scientific authors based in the business sector, the results are *not* representative of the business research community as only a fraction of the latter are allowed to publicly disclose their results in scholarly journals. The results do however show differences within the scientific community that relate to the tension between the disclosure and appropriability paradigms of research. While it would have been technically possible to extrapolate to the general research population, it was ultimately decided against this because of likely unobserved systematic differences between the publishing set of business authors compared to other business researchers.
- The use of an online and by-invitation-only survey approach restricts the study population to scientific authors who are designated as corresponding authors, as contact email information is otherwise not systematically available for sampling purposes. This implies that results are *slightly biased* towards representing the personal experience of scientists acting as principal investigators. However, this approach enables a more comprehensive assessment of the research activity under study. The seniority of individuals in such position can vary across fields.
- The use of an online survey results in differences in response patterns by country and other personal characteristics. In particular, Internet access restrictions and anti-phishing protections have prevented a significant number of targeted respondents from becoming aware of the survey requests or discouraged response. Results have been adjusted to take into account such response patterns as well as possible.
- Results for countries with particularly low effective response rates should be treated with extreme caution. For example, where strong language or Internet access restrictions apply, respondents are more likely to be characterised by having international connections and are, as a result, potentially less representative even when the non-response adjustments are applied.
- While the main focus of this paper is on presenting results that characterise the global scientific community, a number of aggregate indicators at the country/economy level are presented throughout the paper's main section (for countries with more precise estimates) and in the annex. Readers should recall that simple differences in averages across countries can be driven by compositional effects such as the fields a country specialises on or the demographic characteristics of its author population. Regression analysis presented in this document controls for the combined influence of these different characteristics.

3. A multidimensional approach to measuring the digitalisation of science

This section describes the multifaceted transformation of science by examining the adoption of digital tools and related practices by scientists across the multiple stages of the scientific research process. It does so by covering the following aspects: (a) types of research methods adopted by scientific authors; (b) the production of data and code and related sharing and curation practices; (c) use of a broad range of digital tools; (d) online presence of scientific authors and communication of scientific work; and (e) main digitalisation dynamics in science.

3.1. Research paradigms - methods adopted by scientific authors

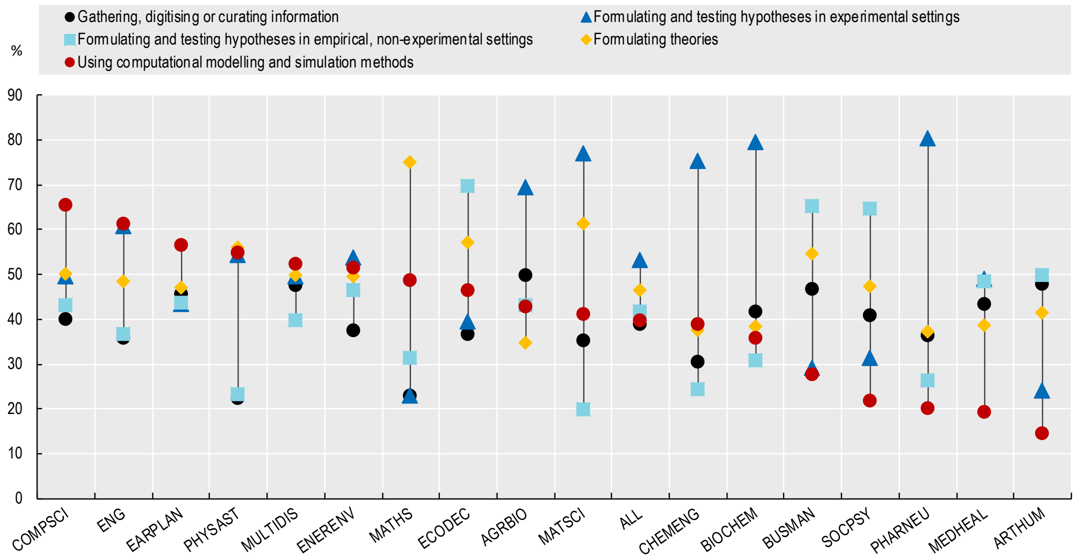
The ISSA2 survey collected information on the involvement of scientists in a range of substantive functional types of research activities associated to different science paradigms. The survey questionnaire is available online[1]. This includes gathering or curating information, formulating theories and studying their properties and predictions; using computation modelling and simulation methods; formulating and testing hypotheses in experimental settings and formulating and testing hypotheses in empirical, non-experimental settings. This information can help identify the main types of research methodologies used across different scientific domains and facilitates interpreting the role and nature of digitalisation.

As Figure 3.1 shows, computer sciences, engineering, and earth and planetary sciences stand out in the use of computational modelling and simulation methods, whereas theory formulation seems to be more common in mathematics. The formulation and testing of hypotheses in experimental settings is a key feature of Pharmacology, toxicology, pharmaceutics and neuroscience, as well as Materials science, whereas the formulation and testing hypotheses in empirical, non-experimental settings is widespread in Economics, finance and decision sciences. Data gathering and curation was reported by less than 50% of authors in all fields, with the exception of Agricultural and biological sciences, where the percentage is slightly above 50%.

[1] www.oecd.org/sti/ISSA2questionnaire.pdf

Figure 3.1. Research methods by science field

Percentage of authors in each field involved in a given research method



Note: Weighted estimates based on sampling weights adjusted for nonresponse. In their answers to the relevant question, respondents can select as many research methods or paradigms as they wish to.
How to read: Nearly 70% of authors in Computer sciences used computational modelling and simulation methods in their research work, compared to 15% in the Arts and humanities.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

3.2. Production and dissemination of publications, data and code

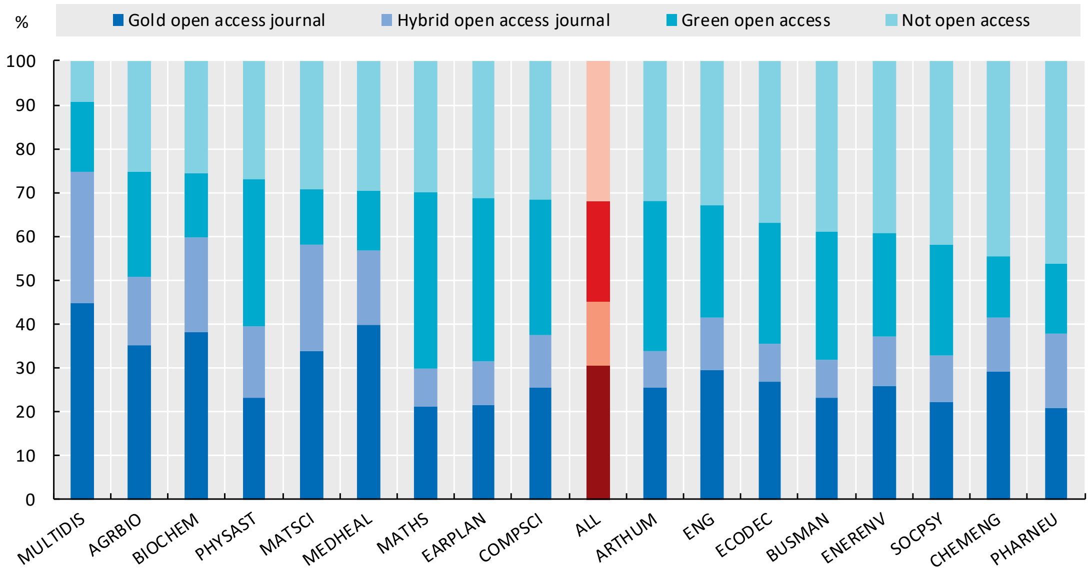
Open access to publications

An important feature of digitalisation is the considerable potential for facilitating access to scientific publications and implementing dissemination models that are more in tune with actual review and publication costs. The subject of open access was covered at length in this survey's previous edition (Boselli and Galindo-Rueda, 2016), investigating not only the extent and nature of open access but also the complex interaction between incentives and scientists' preferences that allowed to generate estimates of the economic value of open access.

ISSA2 results indicate that nearly 70 percent of scientific publications published in 2017 were available in late 2018 on an open access (OA) basis (Figure 3.2). Journal-based open access (gold-OA) seems to be more prevalent in Medicine and health professions as well as in Biochemistry, genetics, molecular biology, immunology and microbiology, whereas repository-based open access (green-OA) is more common in Mathematics and Earth and planetary sciences. Hybrid open-access publishing (subscription journals where the articles is on open access basis) is most often found for documents featured in Multidisciplinary and Materials science journals.

Figure 3.2. Survey-based measures of open access to reference publication, by science field

Percentage of authors

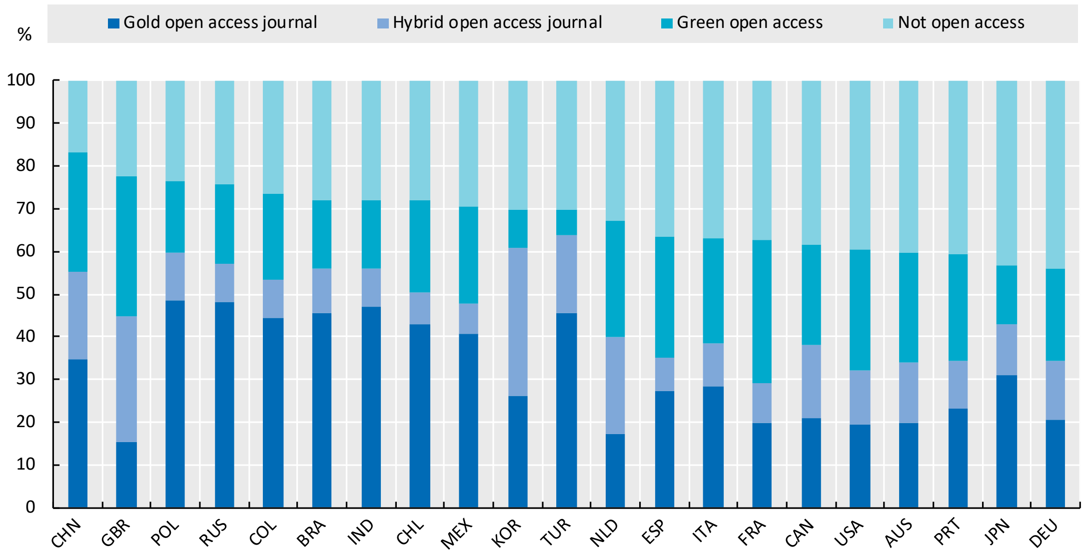


Note: Weighted estimates based on sampling weights adjusted for nonresponse. These estimates exclude an average of 10% of respondents who felt unsure about the access status of their documents.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

An indicative comparison of the ISSA2 results with those obtained in this survey's previous edition (ISSA1) for the fields that have been covered on both occasions suggests an increase in the incidence of open access to scientific publications between 2015 and 2017. This appears to be most pronounced in physics and astronomy as well as in materials science by almost 20 percentage points, compared to 10 percentage points in the arts and humanities. It is also worth comparing the survey responses to the available digital information on the access status of scientific publications. Journal-based OA (gold OA) can be gauged from the detail about the OA status of the journals based on the information kept in the online, community curated, Directory of Open Access Journals (DOAJ) records. Based on the latter, survey responses might appear to overstate gold-OA compared to DOAJ records, and it is apparent that some respondents fail to fully appreciate their document's legal access status from a third party perspective, in particular when journals apply hybrid OA approaches. However, this gap is also partly accounted for by conference proceedings and several other open journals that do not have a DOAJ OA rating. Furthermore, a significant number of journals have evolved in their access policies over time.

Figure 3.3. Survey-based measures of open access to reference publication

Percentage of authors, by economy of residence



Note: Weighted estimates based on sampling weights adjusted for nonresponse. These estimates exclude an average of 10% of respondents who felt unsure about the access status of their documents.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Concerning overall OA status, comparisons with indicators such as those published by OECD (2017) and the EU Open Science Monitor, which use information about the legal access status of Scopus documents in the Unpaywall[2] database,[3] also point to significant differences of appreciation. Authors tend to report considerably higher rates of OA by third parties (from 60% to 70% depending on the treatment of uncertain respondents) compared to those implied by linking Unpaywall data for the same year (36%) using DOIs when available. It might appear that authors are aware of opportunities for finding exact replica of their documents that are not captured by Unpaywall. In conclusion, depending on the OA question asked, survey and algorithmic approaches such as those supported by Unpaywall may need to be combined to develop a more accurate picture about access to scientific documents, particularly in light of widespread practices that cannot be considered to be fully compliant with or actually opposed to defined formal property rights. Comparable results on a country of residence basis are available in Figure 3.3 for large countries, while more detailed results are available in the Annex.

[2] The Unpaywall database has one record for every article with a Crossref DOI (about 95 million). After harvesting different sources in search of Open Access content, the results are matched to DOIs using content fingerprints. For any given DOI, it is possible to learn about OA versions that exist elsewhere. For more information, see https://unpaywall.org/

[3] See https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/trends-open-access-publications_en

Access to data and code

A more novel dimension of ISSA2 is its examination of open science issues beyond OA to publications. The survey gathers information on the generation and use of data and code with regard to the work undertaken in the context of the survey's reference publication[4]. As reported in Figure 3.4, 65% of scientific production results in new data or code. The development of data is more common than that of code. Agricultural and biological sciences, and Biochemistry, genetics, molecular biology, immunology and microbiology are the fields most oriented towards data development, whereas the opposite holds for Mathematics as well as in Arts and humanities. Code generation is more widespread in Computer sciences, followed by Mathematics and Earth and planetary sciences.

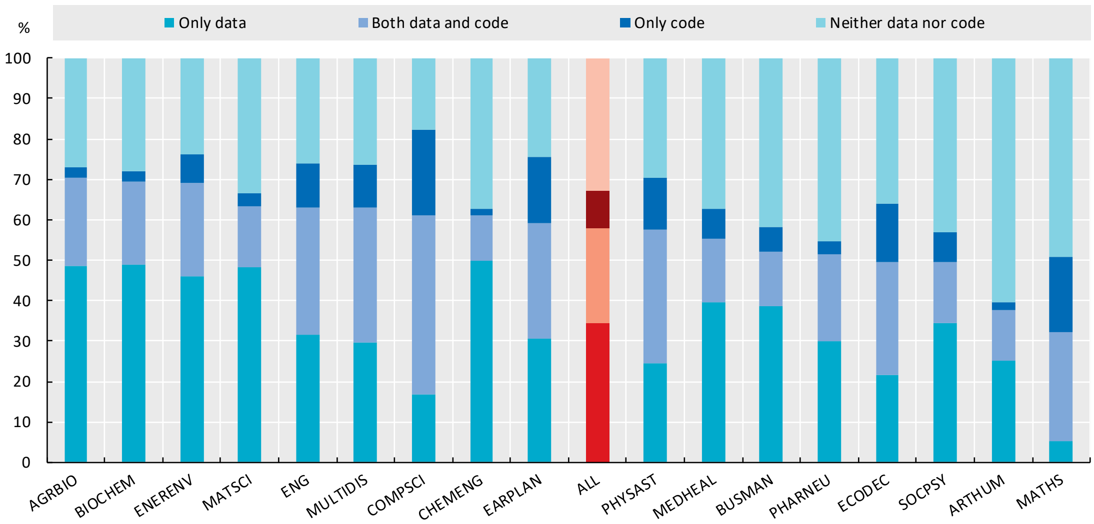
ISSA2 examines the connection between generation of and access to data and code in science. A number of publishers, funding agencies, research institutions, and policy makers encourage researchers to share their data and code to allow for the verification of the research results and, ultimately, spur greater efficiency and quality in research. The survey provides a means to assess whether and how researchers share their data, including possibilities for reuse. Survey participants were asked to indicate whether they have delivered data or code supporting the survey reference publication to the journal, made it available on a repository, or shared it with other researchers. Furthermore, they were also requested to provide details on specific features of their data, through which compliance with the FAIR (findability, accessibility, interoperability, reusability) principles[5] for data sharing can be assessed.

[4] Authors were included into the ISSA2 sample if one of their documents had been randomly selected from a bibliographic database, and only one retained per author in the case of multiple draws per author. See the technical annex for a detailed description of the sampling procedures.

[5] The intent behind the "FAIR" data principles is to provide guidelines for those wishing to enhance the reusability of their data holdings. The principles have since received worldwide recognition by various organisations. To be findable (F), the principles state that data should be assigned globally unique and eternally persistent identifiers, be described with rich metadata, be registered or indexed in a searchable resource, and the metadata should specify the data identifier. Accessibility (A) implies making the data retrievable by their identifier using a standardized communication protocol, which should be open, free and universally implementable and should allow for an authentication and authorisation procedure. To be interoperable (I) the data should use agreed formats, language and vocabularies as well as the metadata should use a community agreed standards and vocabularies. Re-usability (R) means that data and metadata are richly described by a plurality of attributes. It should have a clear and accessible data usage license and be associated with detailed provenance.

Figure 3.4. Scientific production resulting in new data or code, by science field

Percentage of authors with outputs in each category



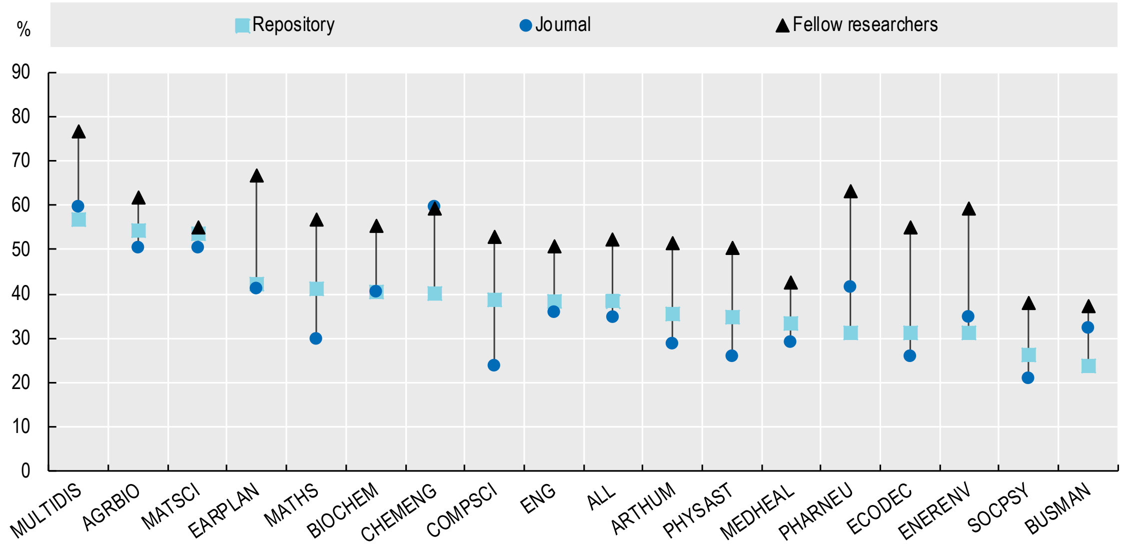
Note: Weighted estimates based on sampling weights adjusted for nonresponse. Research data include numerical scores, textual records, images and sounds used as primary sources for scientific research. Codes include custom developed software and code, laboratory notebooks, and other computer enabled documents describing every step of the research work and protocols followed. It excludes drafts of scientific papers, plans for future research, peer reviews, or personal communications with colleagues as well as physical and non-digitised objects.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

As Figure 3.5 shows, data access and sharing practices are highly dependent on the field of research, including data and code sharing through interpersonal exchanges, the most common type of access mechanism across all fields. Repositories are used on average in close to 40% of relevant cases, including 5% of instances that apply solely to code and 10% combining data and code. Repository archiving practices seem to be more common in Agricultural and biological sciences and in Materials sciences, whereas they are much less diffused in business, management, and accounting. Despite the fact that pre-publication data sharing has highly been advocated in some biomedical sciences such as genomics (Birney et al., 2009), only 40 percent of authors in the combined field of Biochemistry, genetics, molecular biology, immunology and microbiology (BIOCHEM in the figures) provide access to their data from a repository. Less than 40% of authors report delivering their data or code to a journal as supplementary information. This practice seems to be slightly more common in agricultural and biological sciences, material sciences as well as in chemical engineering and chemistry.

Among countries for which sufficiently precise estimates can be derived, India-based scientists have the largest use of repositories whereas the opposite is the case in Japan (Figure 3.6).

Figure 3.5. Dissemination channels for data and code resulting from reference publication

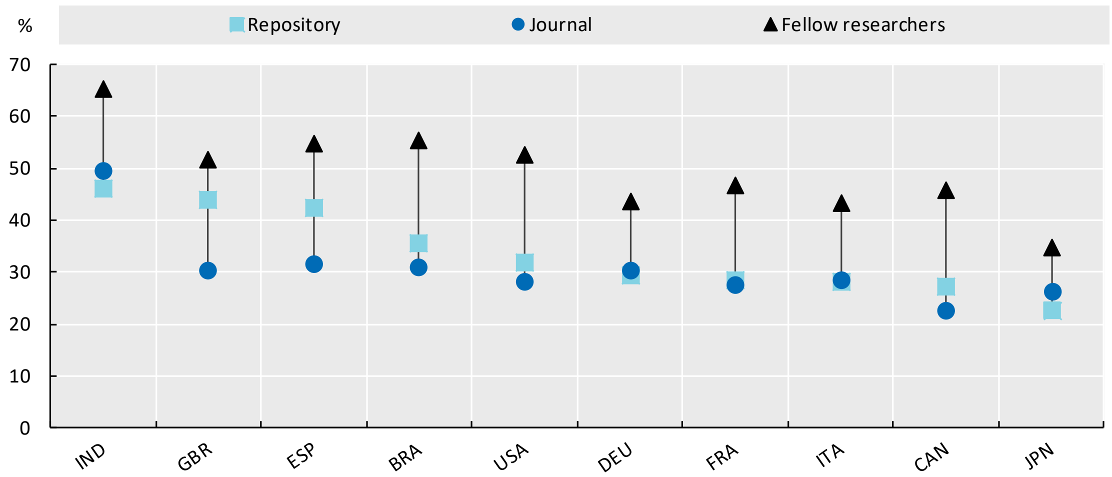
As percentage of authors with data or code outputs, by field



Note: Weighted estimates based on sampling weights adjusted for nonresponse.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure 3.6. Dissemination channels for data and code, by country of residence

As percentage of authors with data or code outputs within each country



Note: Weighted estimates based on sampling weights adjusted for nonresponse.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

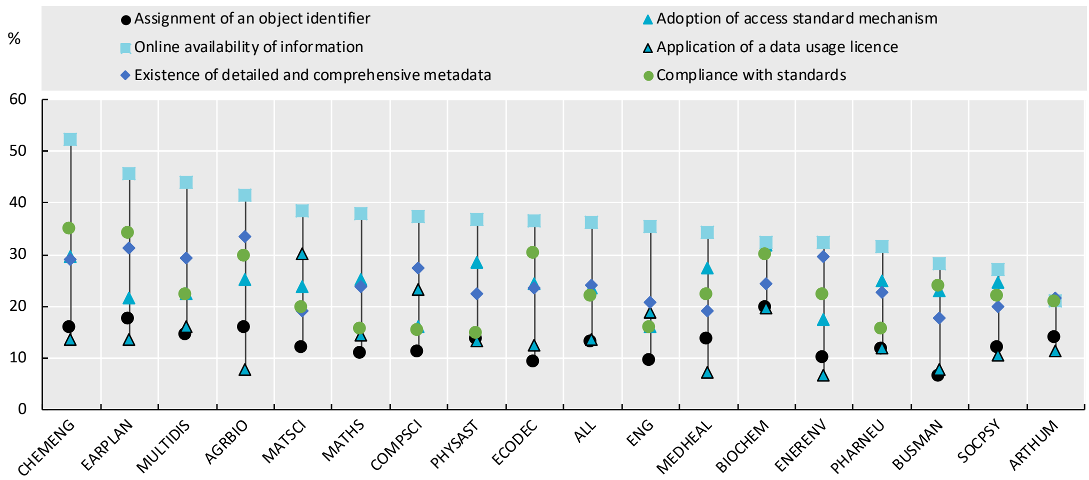
Even when data are shared, this does not mean that they can be re-used and reusability is an important feature of open data[6]. Survey results also point to the possibility that data or code sharing may be ineffective because of inadequate information for reuse (Figure 3.7). Less than 30 percent of authors generating data or code reported facilitating practices such as providing detailed and comprehensive metadata, ensuring compliance with standards

[6] Although an official definition of open data does not exist, it is generally recognised that it refers to data that can be: i) accessed, ii) reused and distributed; and iii) used by anyone (OECD, 2015b).

that allow its integration with data from other sources, or allowing for standardised access request mechanisms. The assignment of a unique and permanent object identifier is more diffused in biochemistry, genetics, molecular biology, immunology and microbiology than in other fields, whereas the practice of applying a data usage licence is more common in materials sciences.

Figure 3.7. "FAIR" quality features of the data or code resulting from reference publication

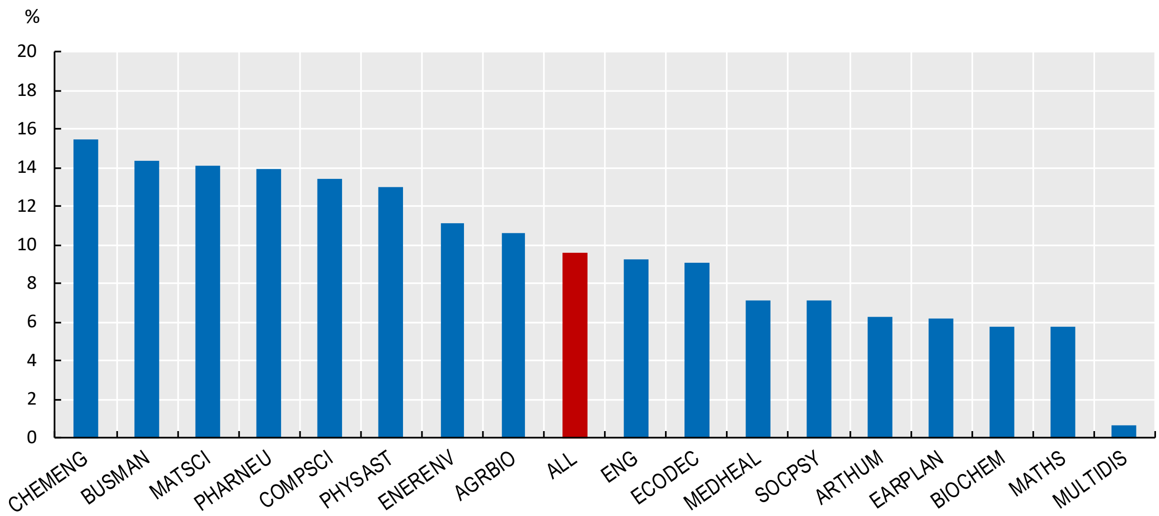
As percentage of authors with data or code outputs, by field



Note: Weighted estimates based on sampling weights adjusted for nonresponse.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure 3.8. Incidence of fee payment requirements to access data or code, by science field

As a percentage of authors with data or code outputs



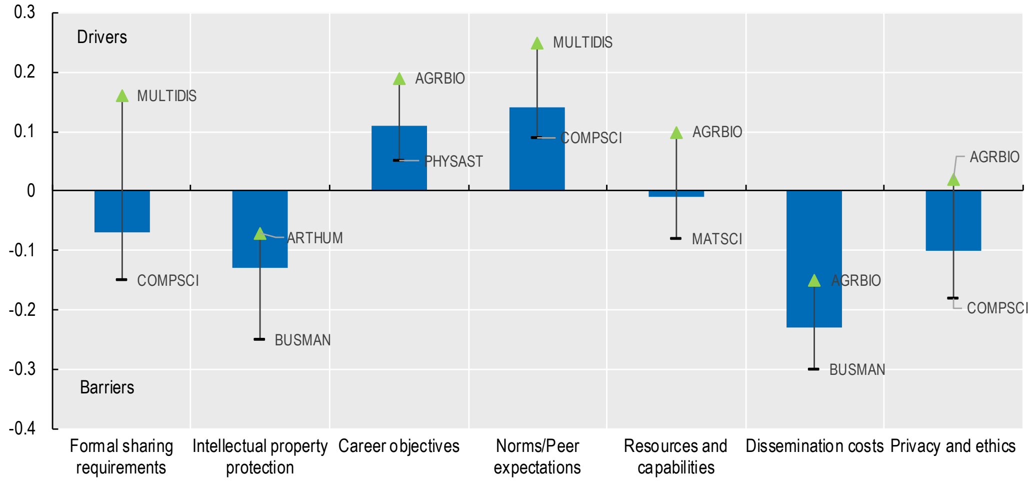
Note: Weighted estimates based on sampling weights adjusted for nonresponse. "Data or code" refer to the data or code developed as part of the work undertaken for the survey reference publication.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

According to survey respondents, data access requires the payment of a fee or a subscription in only 10% of cases (Figure 3.8). Rather than being indicative of free availability of research outputs, this appears to show a lack of systematic access channels upon which pricing mechanisms can be implemented.

A number of studies have shown that trust and intellectual property or confidentiality issues are key reasons why researchers hesitate to share their data (Meijer et al., 2017; Wiley, 2017). The results in ISSA2 coincide in highlighting the inhibiting role of privacy requirements, but point to dissemination costs playing a more prominent role, in terms of money and time, in limiting access to scientific outputs (Figure 3.9), as suggested by Wallis et al (2013) and DAMVAD (2014). Wallis et al. (2013) argue that researchers seem to be concerned about the efforts required to make their data useful to others, so that they are willing to do so either only for the people they know and trust or if they expect to receive credit for it. The ISSA2 findings point in a similar direction as career objectives and peer expectations have been reported by scientific authors as main factors leading to granting enhanced level of access.

It is worth noting that formal publisher, funder or institutional sharing requirements do not seem to significantly influence authors' behaviour as reported by the latter. In light of the fact that an increasing number of journals are changing their policies to make data available[7], this result suggests that such measures may need to be accompanied by other practices and incentives that recognise the need for a shift in social norms around the use of scientific commons and their fit into career objectives.

Figure 3.9. Factors affecting the level of access granted to research outputs



Note: Weighted averages based on response variable adopting values: -1 = Significantly constrained, 0= No significant impact, 1= Significantly enhanced. Sampling weights adjusted by nonresponses have been used. The bar represents the average score across all fields. The figure also displays the values for the fields with highest (diamond) and lowest (dash) average scores.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

[7] One example is the Public Library of Science (PLoS) policy introduced in 2014, which requires authors to state where the data associated with the research can be accessed (Bloom et al., 2014).

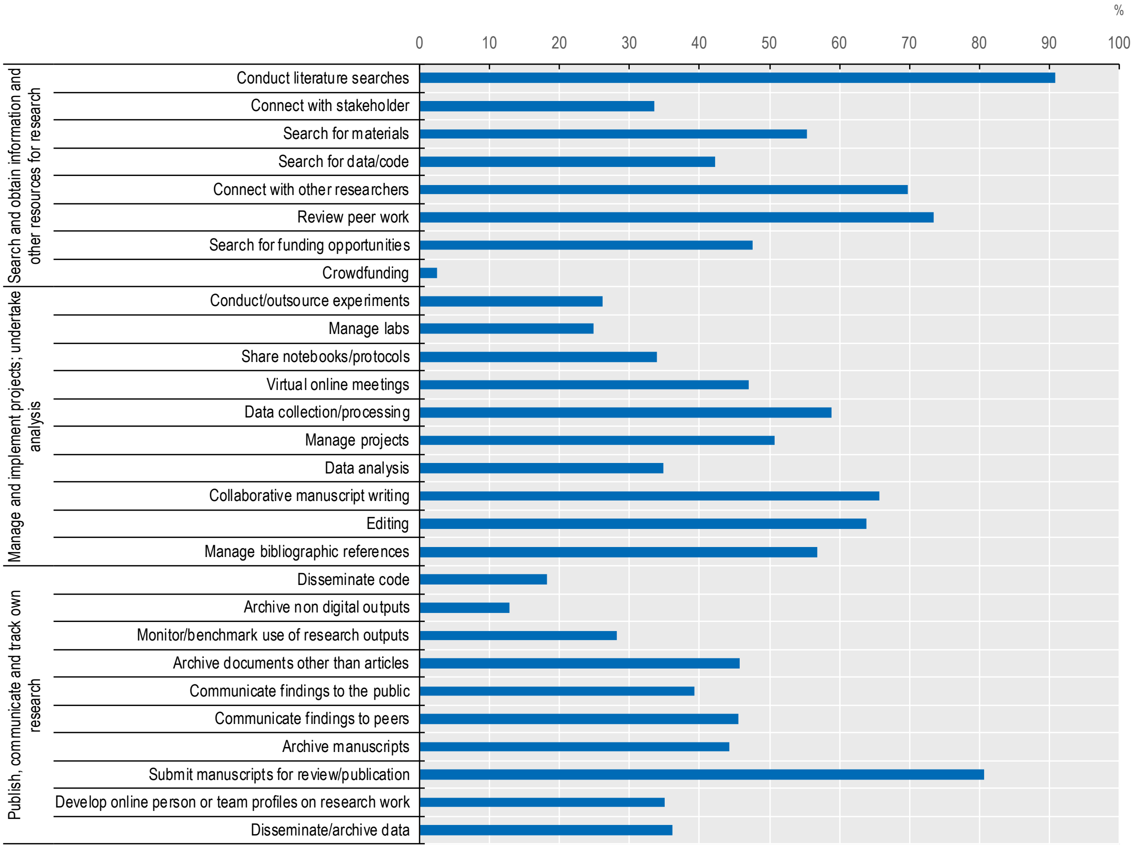
3.3. Researchers' use of digital tools and data

Research lifecycle online tools

ISSA2 collects information on the use of a wide range of digital tools for purposes covering the entire scientific research cycle. Respondents were asked to indicate their use of digital tools for a battery of 28 research-related activities, which can be grouped under the following main dimensions: search of information and resources, project management and data analysis, writing, and dissemination of scientific outputs. The survey results present a multifaceted view of digitalisation of scientific activity (Figure 3.10).

Figure 3.10. Use of digital tools by research activity

Percentage of authors using digital tools for each activity



Note: Weighted estimates are based on sampling weights adjusted for nonresponse. To interpret this figure, it is worth noting that responses reflect a combination of two combined effects, the respondents' propensity to engage in the listed activities and to do so using online digital tools. Given the large list of items, respondents were not asked to separate between the two dimensions in order to limit response burden. For a large number of activities that are largely implicit to the status of corresponding author, the indicator is fully indicative of the digitalisation effect. Other activities such as building online profiles or crowdsourcing, when incurred, are by definition mediated through online tools.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa .

Digital tools have attained full adoption in activities that are essentially common to all publishing scientists such as literature search and manuscript submission, as well as very high adoption in other equally common tasks such as collaborative manuscript writing, editing, managing bibliographic references and communication with other researchers.

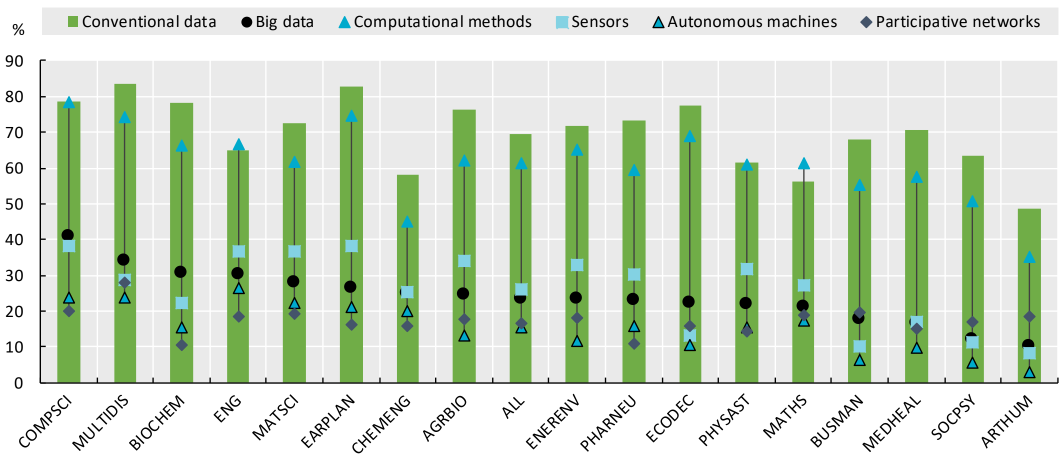
Survey results also provide evidence that activities such as crowdfunding are still uncommon among scientific authors, whereas more than 40% of them disseminate and archive scientific outputs other than articles on online platforms.

Use and development of big data and advanced digital tools

In addition to the adoption of a broad range of productivity and collaborative tools, another key facet of science digitalisation concerns the development and use of advanced digital tools by scientists - not only within the computer sciences and ICT engineering domains - to gain knowledge about phenomena, evaluate models as well as formulate and test hypotheses. This includes the use of big data analytics, artificial intelligence, sensor-nets and the Internet of Things (Hey et al., 2019), whose high potential as general purpose technologies contribute to generate expectations of step changes in research quality and efficiency at a time in which there are concerns about a decline in research productivity (Bloom et al., 2017). The use of cloud computing in science, for instance, is allowing groups to host, process, and analyse large volumes of multidisciplinary data creating economies of scale, facilitating data sharing and collaboration, and enabling long-term data preservation (Gannon et al, 2009), whereas AI is enabling to tackle complex computation problems, collect large-scale data, optimise experimental design, among others (OECD, 2019).

Figure 3.11. Use and development of advanced digital tools or methods, by science field

Percentage of authors using or developing a given tool or method

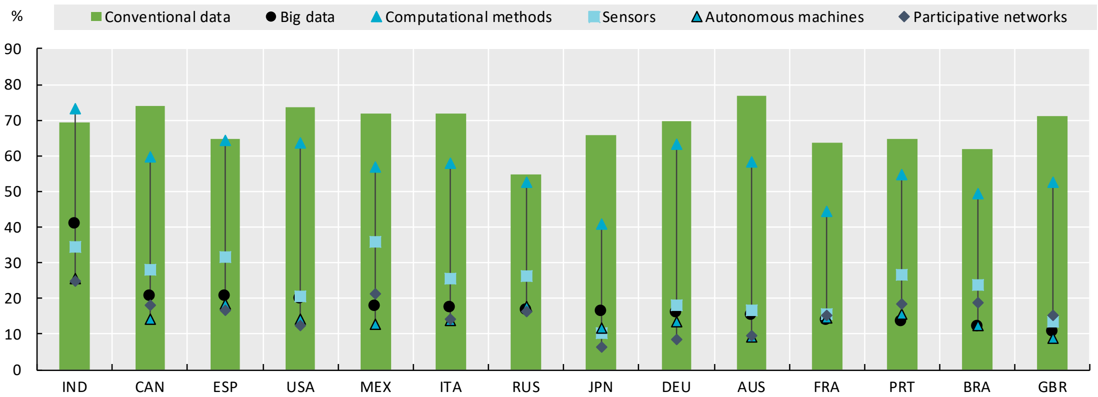


Note: Weighted estimates based on sampling weights adjusted for nonresponse. Based on the wording adopted in the survey questionnaire, *big data* refer to data with size, complexity and heterogeneity features that can only be handled with unconventional tools and approaches; *computational methods* include all computational methods, processes and systems to extract knowledge or insights from structured or unstructured data; *sensors* include all connected sensors to collect information from environment and systems in an automated fashion; *autonomous machines* refer to programmable machines that execute tasks autonomously, such as robots.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure 3.11 depicts ISSA2 results reflecting respondents' answers to questions on the use and development of advanced digital tools. This includes big data (as defined in the figure's note), compared to that of more conventional data, as well as computational methods, processes and systems, connected sensors for data collection, programmable machines that execute tasks autonomously (e.g. robots) and participative networks for gathering data. The results suggest that the use of more advanced digital tools highly varies by field. Big data is more diffused in Computer sciences, Biochemistry, genetics, molecular biology, immunology and microbiology, and Engineering, whereas the use of computational methods is more significant in Computer science and Earth and planetary sciences. Autonomous machines and sensors are more widespread in engineering and computer sciences, respectively. Participative networks exhibit higher use rates in Computer sciences and Business, management and accounting. Overall, the results point to considerable, unexploited digitalisation potential in the arts and humanities, psychology and the social sciences. The share of authors reporting to work with big data is largest in the case of India (40% of authors) (Figure 3.12).

Figure 3.12. Use and development of advanced digital tools, by country of residence

Percentage of authors using or developing a given tool or method within each country



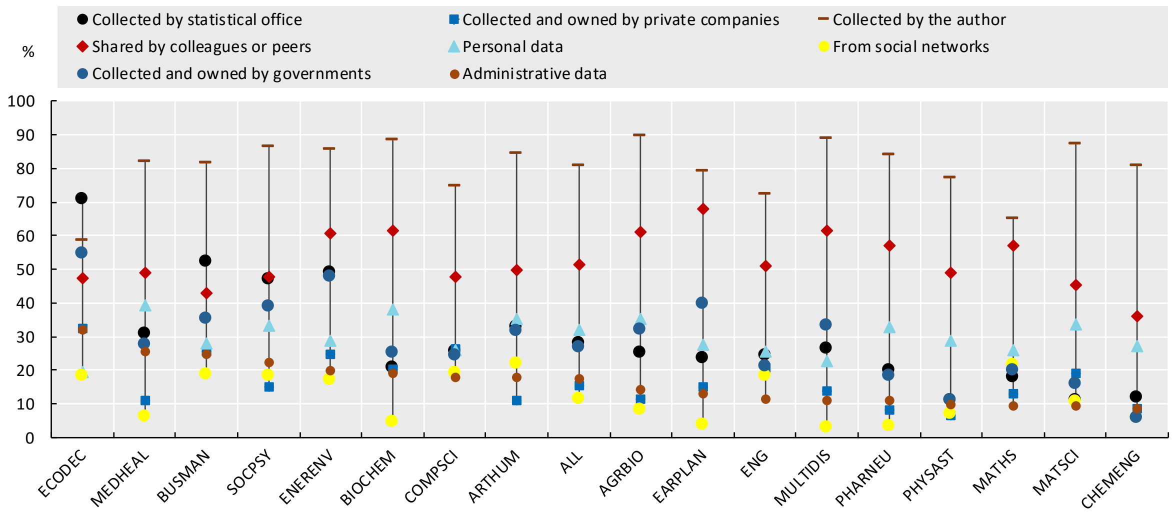
Note: See notes to Figure 3.11.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

The use of digital tools is not only expected to increase the volume but also the complexity and nature of data used and processed by scientists. Figure 3.13 provides evidence on the breadth and origin of data sources, corresponding to overlapping categories, used by scientific authors. Across all fields, a majority of researchers tend to collect their own data, especially within Agricultural and biological sciences and Biochemistry, genetics, molecular biology, immunology and microbiology. Scientists in are the most likely to recognise their reliance on data shared by colleagues. Statistical offices are key data sources for scientists in Economics, finance and decision sciences at more than 70%. The use of personal data is most common in Medicine and health professions, and Biochemistry, genetics, molecular biology, immunology and microbiology.

Nearly 20 percent of scientific authors work administrative data – and nearly 30% in Economics, finance, and decision sciences. The rising importance of social networks is reflected in the 10 percent of authors using data originating from such sources, a significant step for a completely new field.

Figure 3.13. Type of data used by scientific authors

Percentage of authors using a given type of data



Note: Weighted estimates based on sampling weights adjusted for nonresponse.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

3.4. Digital identity of scientific authors

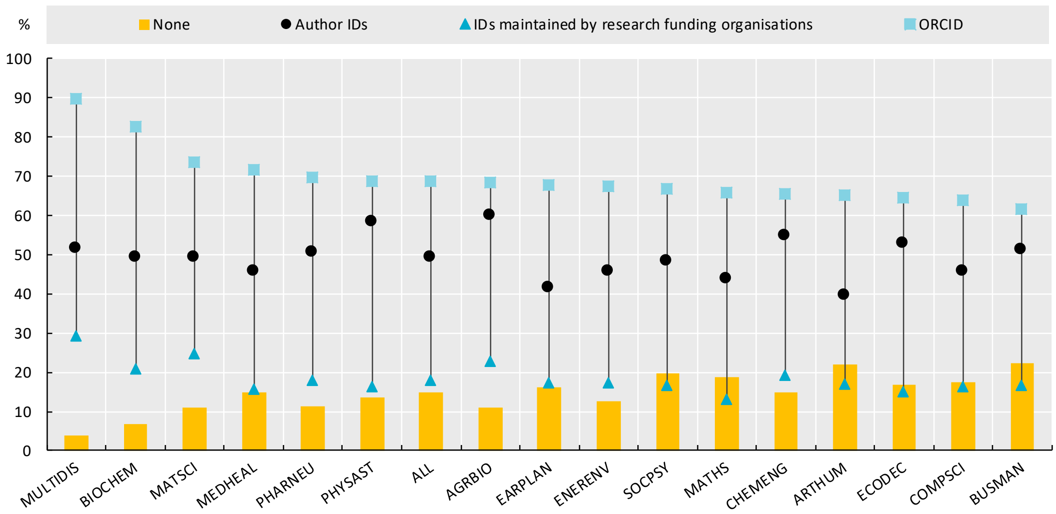
ISSA2 also examines various aspects relating to digital identity, understood as the body of information about an individual or organisation that exists online. For scientists, digital identity can be a crucial reputation asset and vehicle of communication with their peers and society more broadly. Digital identity in science is important for the organisations for which scientists work as well as for third parties that need or may need to interact with them, for example publishers or funding agencies in search of reviewers, or firms in need of specialised scientific or technical services. The increasing availability of information about scientists and their work that exists online relative to other media can in turn change the way in which scientists communicate scientific results and information to peers or the public as a whole, influencing in turn how research and researchers are reviewed and evaluated. ISSA2 explores the extent to which information on an author's research activities is available online and provides information on the link between such information and the way researchers and their output are currently evaluated.

Use of identifiers and communication mechanisms

Use of IDs to assert identity in digital environments is widespread across all fields. Only 15% of authors report not using them to claim authorship over their work and distinguish themselves from other researchers or individuals. The "Open Researcher and Contributor Identifier" (ORCID), promoted by the namesake international non-profit organisation, appears to have become the prevailing global standard as it is the most diffused type of identifier used by scientific authors worldwide (more than 60 percent), followed by other author IDs or personal profiles associated to citation indices (Figure 3.14)[8]. Organisational or country-specific IDs are widely reported but their lack of global reach restricts their overall level of uptake with localised exceptions.

Figure 3.14. Use of identifiers (IDs) by scientific authors to track their research work

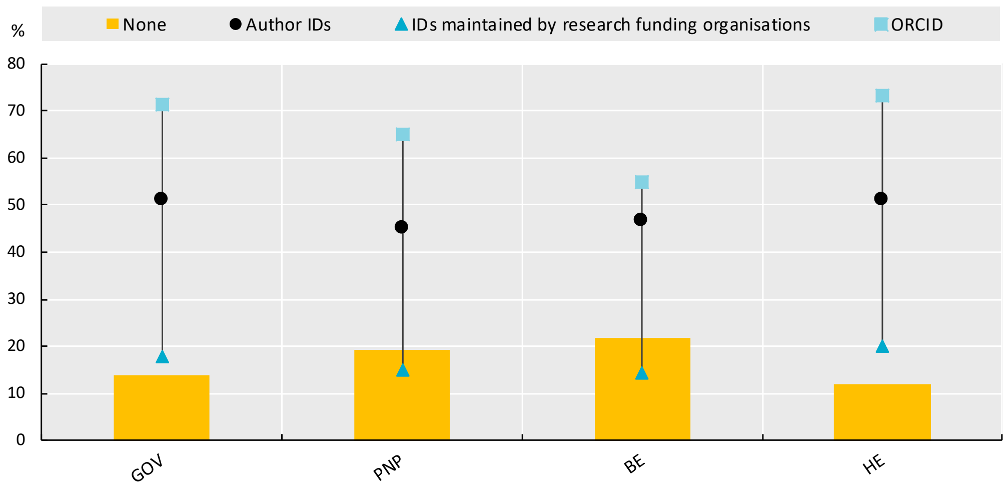
Percentage of authors that use selected types of IDs



Note: Weighted estimates based on sampling weights adjusted for nonresponse. Less than 2.5 percent of authors in each field reported using other types of identifiers.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure 3.15. Use of identifiers (IDs) by scientific authors, by sector

Percentage of authors within sector that use selected types of IDs



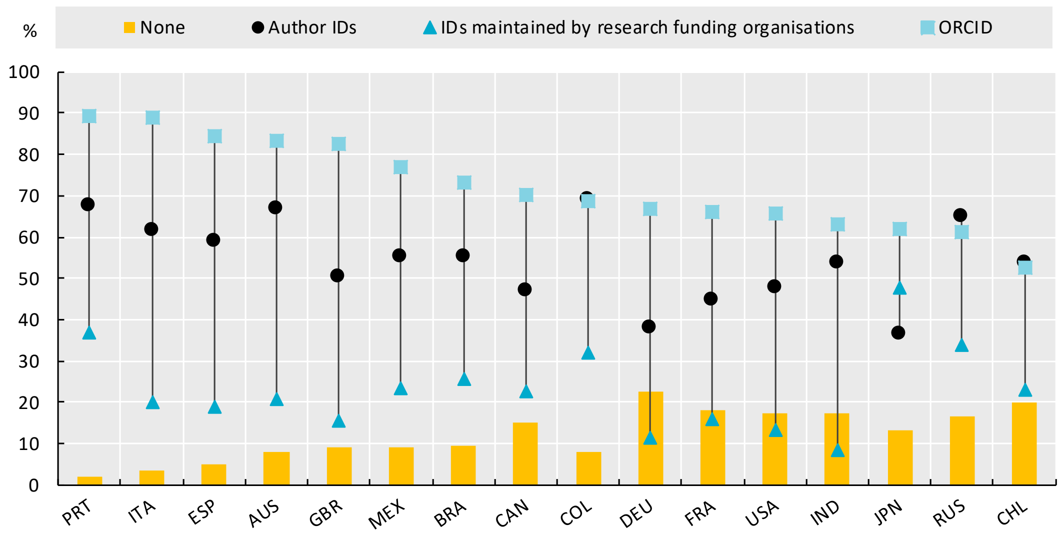
Note: Weighted estimates based on sampling weights adjusted for nonresponse.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

[8] All authors included in the ISSA2 study have a SCOPUS Author ID, which is assigned automatically to every author after one of their documents is indexed in Scopus. However, some authors may not only fail to make use of it but might also be unaware of it.

Results provided on a sector of employment basis (Figure 3.15) highlight the very marked use of ORCID within Higher Education and within Government (about 75%), while adoption is significantly lower among scientific authors within business (55%). Use of ORCID is most extended in Portugal and Italy at over 90%, while adoption is at or below 60% in the cases of Japan, Russia and Chile (Figure 3.16). National identifiers are most prevalent in Japan, Portugal, Colombia and Russia.

Figure 3.16. Use of identifiers (IDs) by scientific authors, by country of residence

Percentage of authors within sector that use selected types of ID



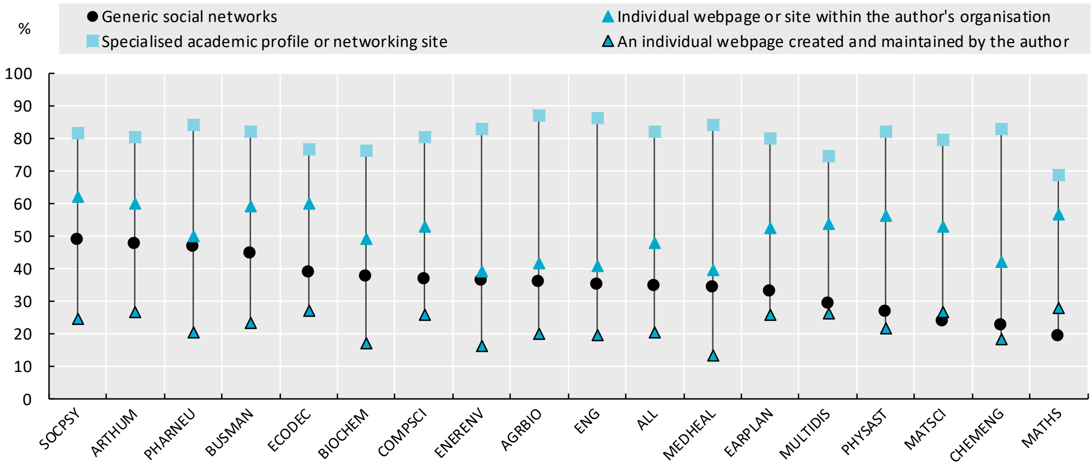
Note: Weighted estimates based on sampling weights adjusted for nonresponse.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

While IDs are a key component of digital identity in science, scientific authors can develop their public profiles by several active ways. As Figure 3.17 shows, responses to the ISSA questionnaire suggest that they prefer to provide information about their work through specialised academic profile or networking site, at close to 80% of cases. This stands in contrast with the traditional, organisational web page containing an academic profile. Nearly 50% of authors, mostly individuals working in higher education institutions, use individual pages within sites maintained by their organisations in order to showcase their work and career trajectories. There is also evidence of the widespread impact of digital platforms stemming from the significant uptake of generic, non-specialist, social networks for communicating research, with nearly 35% of authors now using them for such purpose. This use is particularly pronounced in the social sciences and humanities science area. In contrast, individually maintained webpages are only reported in 20% of cases.

At over 50%, authors based in the United Kingdom and in Australia are the most likely to use generic social networks to communicate and disseminate information. This is about twice as much the incidence of reported use among scientists in Germany, Japan and Russia (Figure 3.18).

Figure 3.17. Channels used for online dissemination of information about research

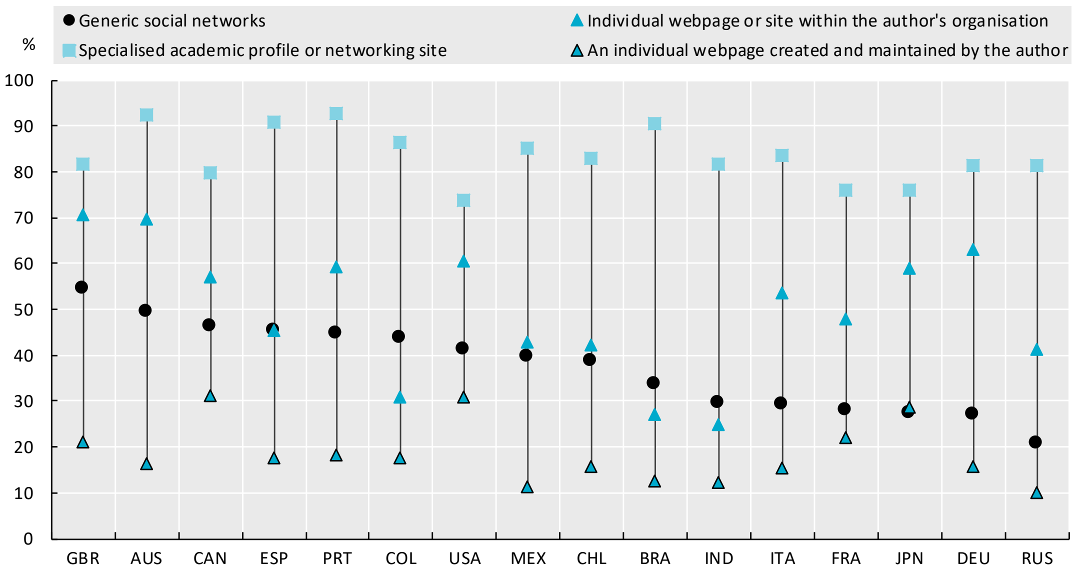
Percentage of authors using a given channel within each field



Note: Weighted estimates based on sampling weights adjusted for nonresponse. *Specialised academic profile or networking site* encompasses the use of ORCID in addition to other specialised academic websites. Less than 5% of authors do not use any said vehicle, with the exception of Engineering (7%) and Physics (10%).
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure 3.18. Channels used for online dissemination of research information, by country

Percentage of authors using a given channel within each country



Note: Weighted estimates based on sampling weights adjusted for nonresponse. *Specialised academic profile or networking site* encompasses the use of ORCID in addition to other specialised academic websites.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

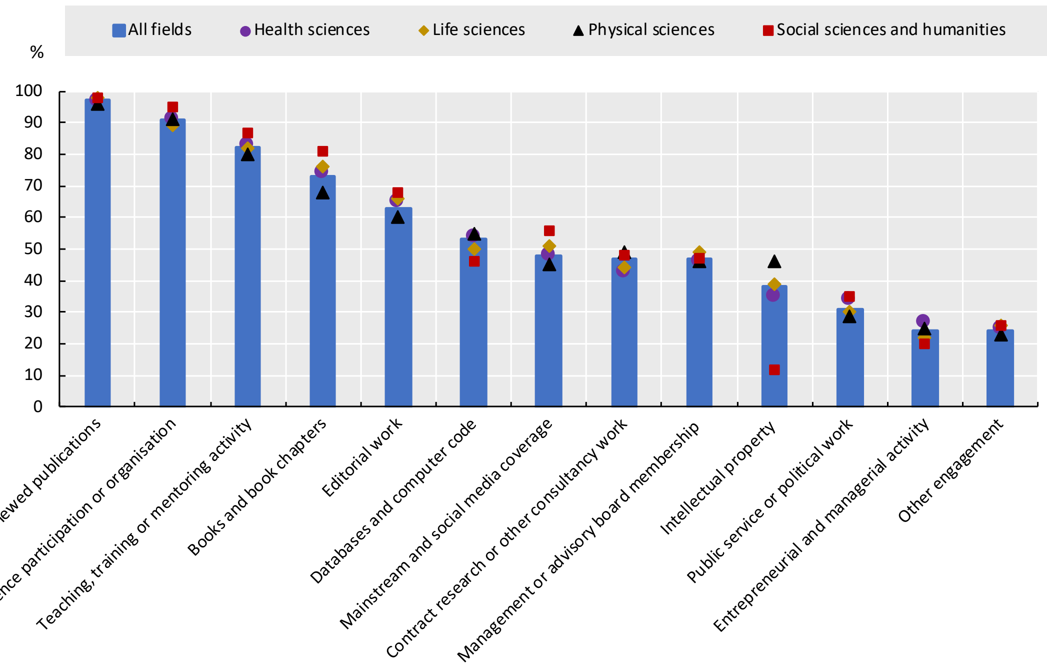
Digital footprint of scientific activity

The ISSA2 study follows up on the online identity section by enquiring on the type of information scientists provide publicly online, compared to the science-related activities they actually engage in. This is relevant for two main reasons. Firstly, this information can provide a better understanding of the entire breadth of researchers' activity beyond what traditional bibliometric indicators regularly capture. Secondly, this can help assess the extent to which new metrics based on the digital footprint of scientific activity can be used to provide a comprehensive view of mechanisms of research impacts.

ISSA2 respondents were provided with an extended list of possible activities and science-related outputs, being asked in first instance to indicate if they engaged in the relevant activity or produced the indicated output and, if so, whether information was available about it. Starting with the former, as Figure 3.19 shows, only activities traditionally considered as core academic were reported by more than 50% of scientific authors. This includes activities and outputs directly connected with the production, review and promotion of scientific papers among the scientific community, as well as teaching or mentoring activity and book or book chapters. Approximately 50% of authors conduct research-related consultancy work or are members of a management or advisory board. There are no large differences in authors' "third mission" engagement across science areas, with the exception of generation of registered IPRs which is reported in less than 40% of cases. IPRs registration (which excludes non-registered copyrights) is highest in Physical sciences (concerning approximately 50% of authors) and much lower (around 10%) in Social sciences and humanities. 30% of scientists carry out public service or political work, slightly above the 25% of authors who have business executive roles (including start-up entrepreneurial activity).

Figure 3.19. Engagement in research-related activities, by science area

Percentage of authors engaged in the specific activity

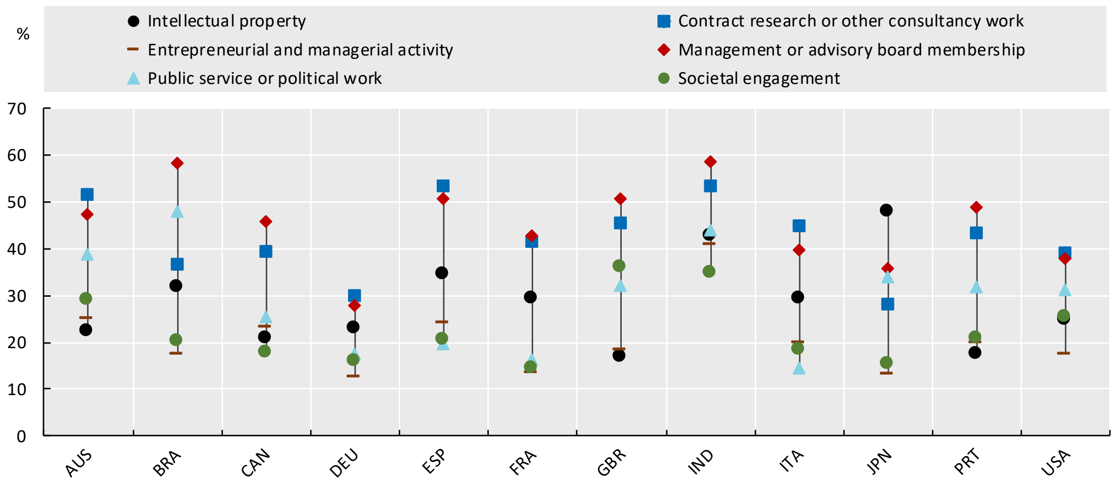


Note: Weighted estimates based on sampling weights adjusted for nonresponse.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Overall, 25% of scientists did not engage in any of so-called third mission activities. It should also be borne in mind that the population includes non-academic authors whose main activities may correspond to those considered to be as "third mission" for academics. The scientific communities in different countries are characterised by different propensities to engage in such activities. For example, scientists in Japan are the most likely to report having registered or applied for intellectual property protection (Figure 3.20), while this is only reported in 20% of the cases in the United Kingdom.

Figure 3.20. Engagement in research-related activities, by country of residence

Percentage of authors within each country engaged in the specific activity

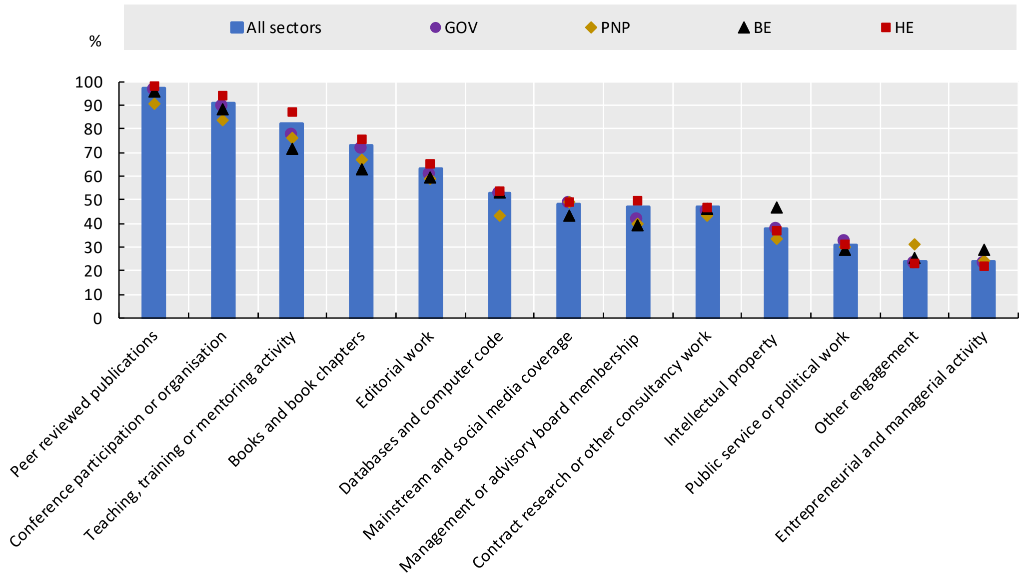


Note: Weighted estimates based on sampling weights adjusted for nonresponse.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

As shown in Figure 3.21, there are some significant differences in engagement among scientific authors by sector of affiliation that reveal differences in priorities and motivations in those sectors. However, these differences are probably smaller than might be expected for the general researcher population.

Figure 3.21. Engagement in research-related activities, by sector of employment

Percentage of authors engaged in the specific activity



Note: Weighted estimates based on sampling weights adjusted for nonresponse.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

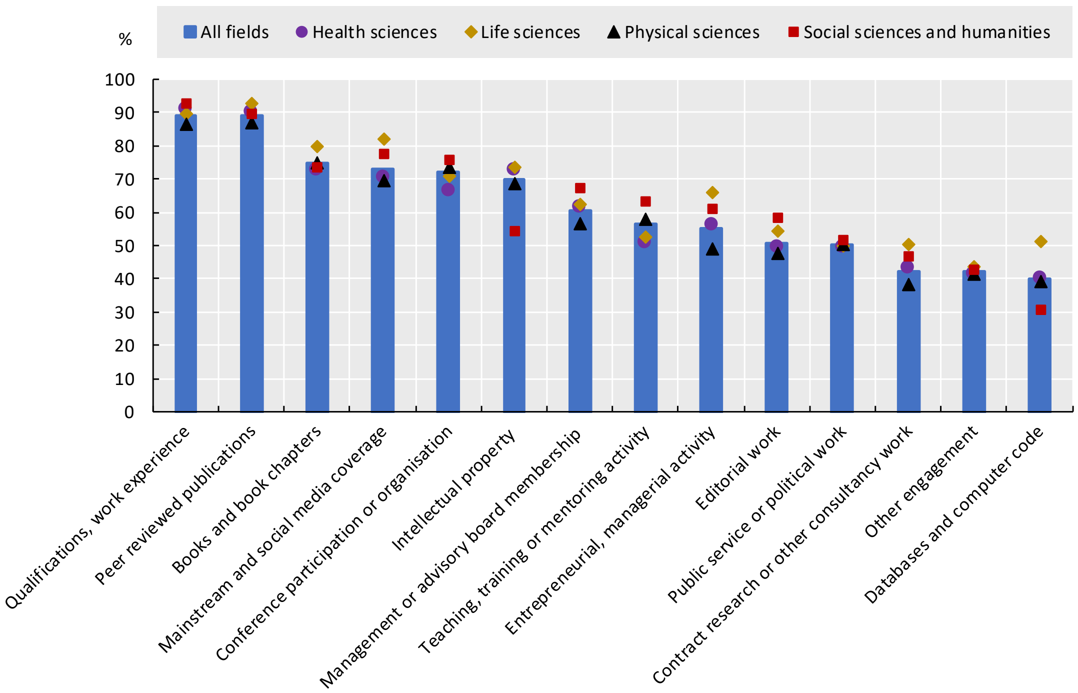
Turning to the digital footprint for the reported activities, responses indicate considerable variation in online availability of information about them. (Figure 3.22). Digital coverage appears to be rather comprehensive for basic information such as lists of peer-reviewed publications, qualifications and experience. A second layer of activities attain consistently high digital footprints, namely in the areas of book contributions, conference participation IP and media. These are areas with well-established or growing capabilities for compiling comprehensive databases and indicators.

Some activities attain rather low rates of digital traceability. For example, consistent with results reported in Section 3.2, only 40% of authors indicate that information about their data and code outputs, or about contract research and consultancy work, are available online.

These results have implications for the future of measurement and assessment of scientific activity based on its digital trace. Several areas are still not leaving an accessible digital trace, so digital tools are imperfect substitutes for a broad range of possible assessments. Furthermore, as noted earlier, data availability is only a pre-condition of usability. Digital information is in many instances unstructured and scattered across different sites, often under control of companies providing associated intelligence services. This highlights some of the major limitations for the use of altmetrics to characterise and monitor the nature and impact of scientific activity.

Figure 3.22. Online availability of information on research-related activities, by science area

As a percentage of authors engaged in the relevant activity



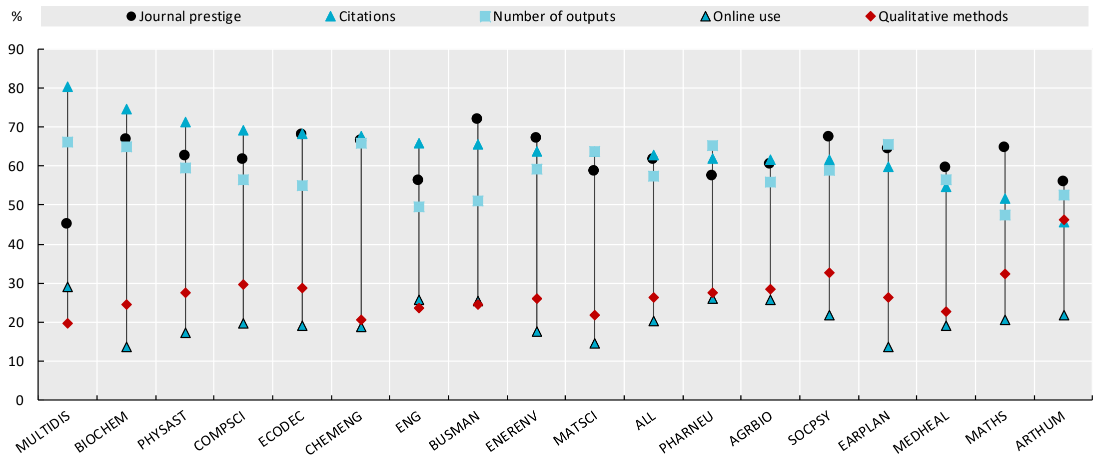
Note: Weighted estimates based on sampling weights adjusted for nonresponse.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Measurement and assessment of science in the digital age

Given the importance of metrics supported by online digital sources and tools for the assessment of impacts, ISSA2 enquires about how researchers perceive that they are used in their work environments. The results demonstrate the pervasive use of traditional indicators based on measures of journal prestige, counts of outputs such as number of peer-reviewed papers and number of citations (Figure 3.23). Although rather novel, online usage metrics, which measure the numbers of views, downloads, etc., are already reported to be used in 20% of cases. These metrics, which in some cases provide insights into the usage of scientific outputs outside the population of scientific authors (for example, medical practitioners or policy makers who download scientific documents to keep to up to date with the literature and support their decisions) are beginning to be systematically provided by publishers. A relative minority of respondents in most fields appreciate that qualitative methods are systematically used in their areas, only approaching 50% of cases for the Arts and the humanities. If respondents have correctly interpreted the question, this suggests that quantitative approaches have displaced qualitative evaluation approaches.

Figure 3.23. Use of research metrics, by type and science field

Percentage of authors reporting the use of a given type of research metrics

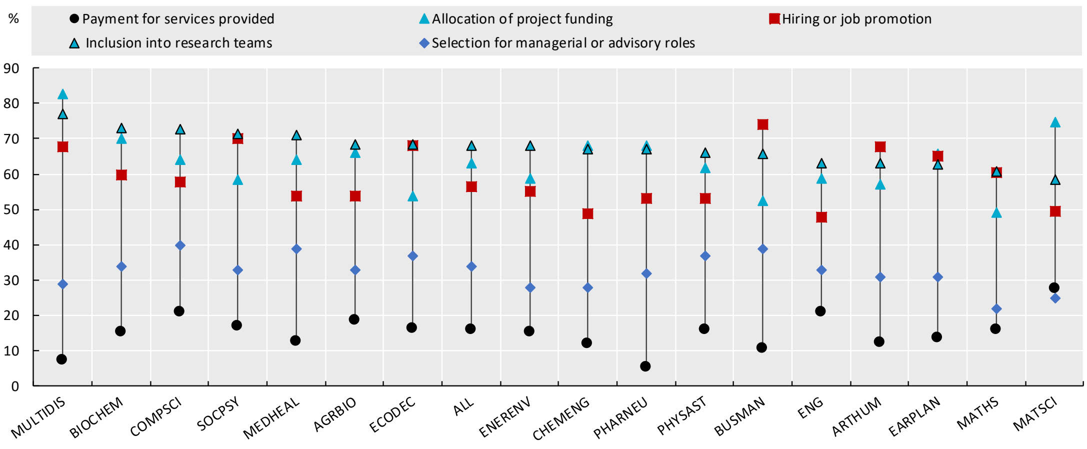


Note: Weighted estimates based on sampling weights adjusted for nonresponse.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

The previous findings raise the question of how research metrics are used. ISSA2 respondents appear to point to a number of key applications of research indicators, primarily informing decisions to include individuals in research teams (70% of cases), closely followed by decisions to allocate project funding (Figure 3.24). Over 55% of scientific authors report that quantitative indicators are being used to inform hiring and job promotion decisions in their work environments. This type of application is particularly marked among scientists in the social sciences and humanities, while significantly less so in engineering and materials sciences.

Figure 3.24. Decisions informed by quantitative indicators of research, by science field

Percentage of authors indicating different channels of indicator-based influence on their careers

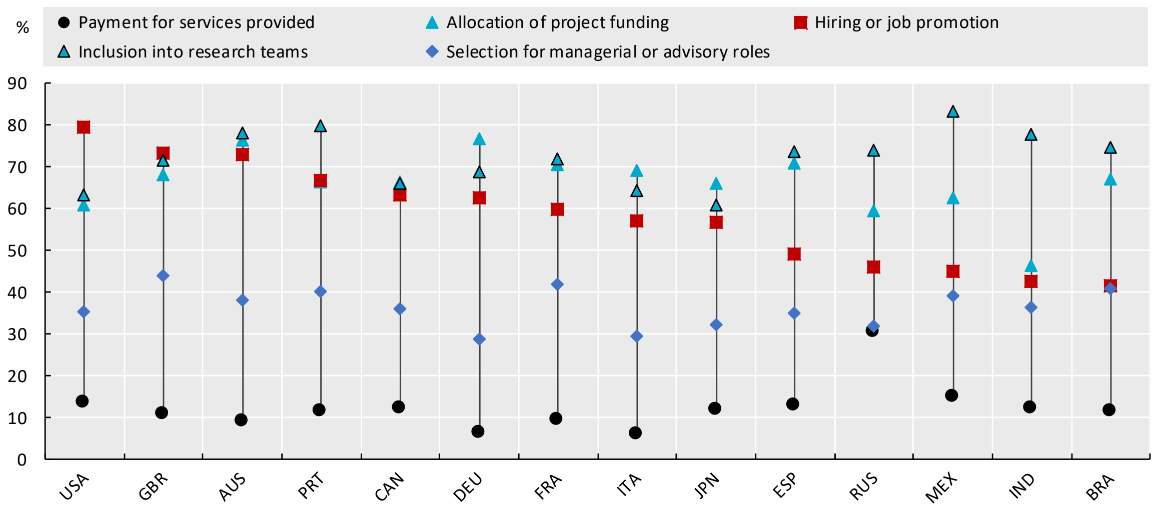


Note: Weighted estimates based on sampling weights adjusted for nonresponse.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Scientists in the United States, the United Kingdom and Australia are among those most likely to report that quantitative indicators are used in their areas of work to inform hiring and job promotion decisions (Figure 3.25).

Figure 3.25. Decisions informed by quantitative research indicators, by country of residence

Percentage of authors indicating different channels of indicator-based influence on their careers

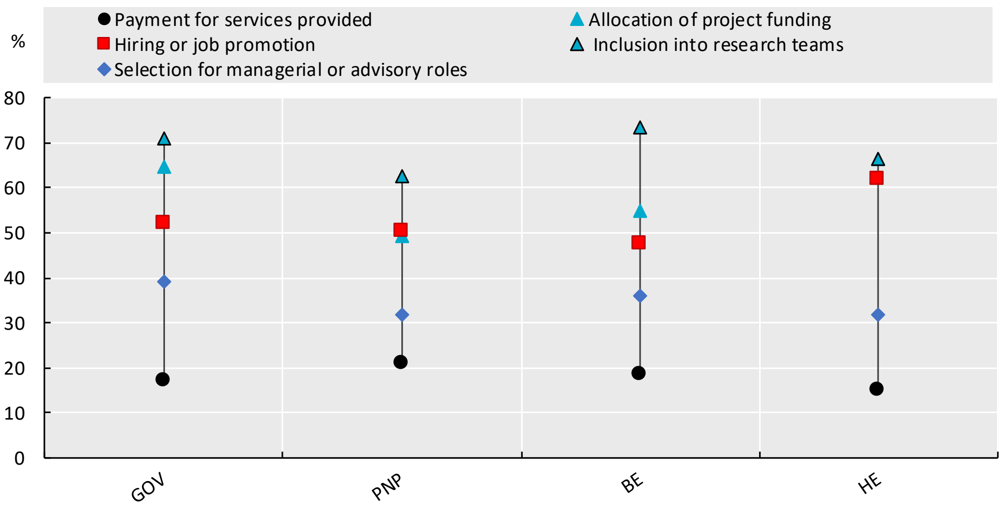


Note: Weighted estimates based on sampling weights adjusted for nonresponse.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

As shown in Figure 3.26, the Higher Education sector appears to make the most intensive use of indicators to inform hiring and job promotion decisions, especially compared to the Business sector. The latter sector does however make intensive use of indicators to guide project funding decisions.

Figure 3.26. Decisions informed by quantitative indicators of research, by sector

Percentage of authors indicating different channels of indicator-based influence



Note: Weighted estimates based on sampling weights adjusted for nonresponse.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

3.5. Patterns of digitalisation of scientific activity

The presentation of descriptive statistics corresponding to multiple and interrelated survey items helps introduce a more exhaustive exploration of general patterns of digitalisation in scientific activity. The wealth of data collected, corresponding to a large number of variables, needs to be synthesised to identify core features and patterns. This further serves the purpose of facilitating the subsequent analysis of the drivers and impacts of digitalisation in its various facets. This paper uses factor analysis as a dimensionality reducing technique to identify a reduced number of synthetic indicators capturing distinct and relevant patterns of science digitalisation.[9]

The full list of variables contributing as inputs in the factor analysis can be found in Table B.1. The list has been selected from the entire set of survey items on the use of digital tools or adoption of digitally-enabled practices of potential relevance to all respondents regardless of their individual contexts. These include responses to questions described in the previous section relating to the use of online platforms or related applications or tools in research work, use or development of data, code or more advanced digital tools or methods (e.g., big data or sensors), quality features of the data/code used or developed in research work, use of identifiers to track own research work, and use of networking sites or other specialised websites to provide information on own research activities or outputs. Given the binary (dichotomous) nature of these variables, tetrachoric correlations for each pair of variables were calculated and the factor estimation was then applied to the resulting pairwise correlation matrix, using the principal-component factor model.

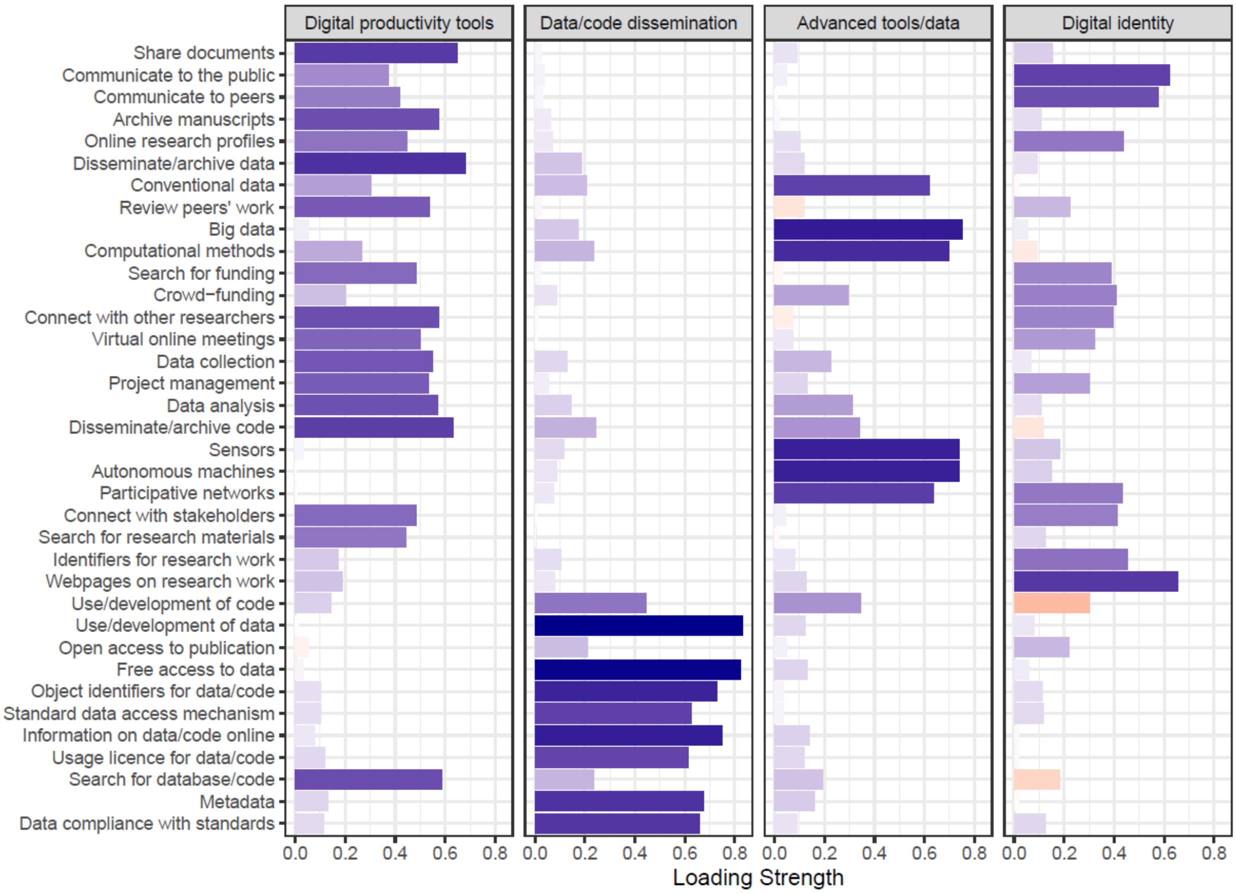
The principal component factor method enables the extraction of a minimum number of factors that accounts for a maximum proportion of the variables' total variance. The number of factors selected was restricted to four based on the analysis of the eigenvalues, screen plot and the variance accounted by each factor. To facilitate the factors' interpretability, an orthogonal rotation generating mutually uncorrelated factors was applied to the factor loadings, i.e. the correlations between the factors and the underlying variables. The resulting rotated factor loadings are provided in Figure 3.27. The factors are interpreted and labelled based on their loadings (associations) with the observed variables. These turned out to refer to four major facets of the digital transformation of science:

- use of digital scientific collaboration and productivity tools;
- development and management of digital access to data and code;
- use of advanced, computing oriented, digital tools (e.g. big data analytics);
- digital identity in online environments and communication of scientific work.

[9] Factor analysis is widely used, in particular in the in social sciences, to explain correlations among variables in terms of a lower number of unobserved, latent factors, and to help assess the robustness and internal validity of survey items.

Figure 3.27. Correlation of digitalisation factors with question items

Factor analysis loadings (associations) in absolute values (blue=positive; red=negative)



Note: The factor analysis is based on the responses by scientists to questions relating to the use of digital tools or adoption of digitally-enabled practices. The resulting four factors have been interpreted and labelled based on how strongly they correlate with the survey-based underlying variables. Factor loadings represent the correlations between the factors and the underlying variables and can take values ranging from -1 to +1. See Box 3.1 for more information on the factor analysis procedures and results.
How to read: The variable “Use/development of data” is strongly and positively correlated with the factor named “data/code dissemination”, whereas the variable “Use/development of code” is negatively correlated with the fourth factor on digital identity.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

For instance, the data/code dissemination-related indicator correlates more strongly with data sharing practices, whereas the indicator relating to authors’ digital identity is more strongly related to the use of online platforms to communicate scientific results to peers or the public, and the use of specialised webs or networking sites to provide information on own research work. Overall, the factors account for nearly 50 percent of the underlying variance (see Table B.2).

Box 3.1. Evaluating the degree of determinacy/indeterminacy of the factor scores

Factor analysis does not produce determinate factor scores because, as the common factor model attempts to define a number of variables that is larger than that of the model equations, a unique solution for the common factor model does not exist (Grice, 2001). Under some conditions, this may result in overly ambiguous results that impact on the interpretability of the analysis. It is suggested to assess the degree of indeterminacy in a given analysis and evaluate the factor scores using a number of indicators referring to a set of criteria (Grice, 2001):

1. *Validity,* i.e. is there sufficient correlation between the factor score estimates and their respective factors?
2. *Univocality*, i.e. are the estimated factor scores excessively or insufficiently correlated with other factors in the same analysis?; and
3. *Correlation accuracy*, do the correlations between the estimated factor scores match the correlations among the actual factors?

Diagnostic indicators relevant for the assessment of these three criteria include two indeterminacy indices, namely the multiple correlation between each factor and the original variables and the minimum possible correlation between two sets of competing factor scores, the validity coefficients, as well as univocality and correlational accuracy matrices. Annex Table B.3 reports the results of the tests applied to the digitalisation-related factor scores. In summary, these tests seem to reveal good levels of validity, univocality, and correlational accuracy for the four factors derived. The multi correlations between each factor and the original variables are higher than 0.8 for all factors, whereas the minimum possible correlation between two sets of competing factor scores is never too low, i.e. never below 0. Minimum possible correlation values that are equal to or below 0 would indicate that two sets of competing factor scores could be constructed for the same factor that are orthogonal or negatively correlated. The validity coefficients are also high and above 0.8. The univocality criterion is assessed by comparing the univocality matrix with the matrix with the factor correlations as the two should ideally match. The highest difference is found between factor scores for *Digital productivity tools* and the *Digital identity and communication* factor and is equal to only 0.117. Finally, the differences between the matrix with the correlations among the estimated factor scores and that the correlations among the factors themselves are also very low (correlation accuracy criterion). The largest difference is found for the *Digital collaboration and productivity* and *Digital identity and communication* factors and is equal to 0.125.

These factors lend themselves to being interpreted as indicators of the strength of different digitalisation paradigms:

- Factor 1 (digital collaboration and productivity) does to some extent capture a form of generic digitalisation of the kind that transforms common types of research activities.
- Factor 2 (data and code access) instead relates directly to the digital basis for the open science paradigm.
- Factor 3 (advanced digital methods) alludes to the new data driven research methods paradigm.
- Factor 4 (digital identity and communication) reflects yet another key dimension of digitalisation that has significant elements in common with wider societal trends as it concerns the digital trace of research activity and its implications for careers and reputation.

These are, however, general interpretations solely based on correlation patterns across a battery of survey items. Their fundamental validity will be assessed in the ensuing analysis, using individual level indicators or scores for these four factors.

To conclude the description of the factor analysis, a least-squares-regression approach was employed to estimate factor scores. The latter illustrate the positioning of each individual on each factor and can be used in subsequent analysis. Since, factor scores are not always well defined, a range of additional tests have been carried out (see Box 3.1) to assess the robustness of the factor analysis approach to the digitalisation of science. The results of these tests confirm the determinacy and stability of the scores as synthetic measures of digitalisation patterns.

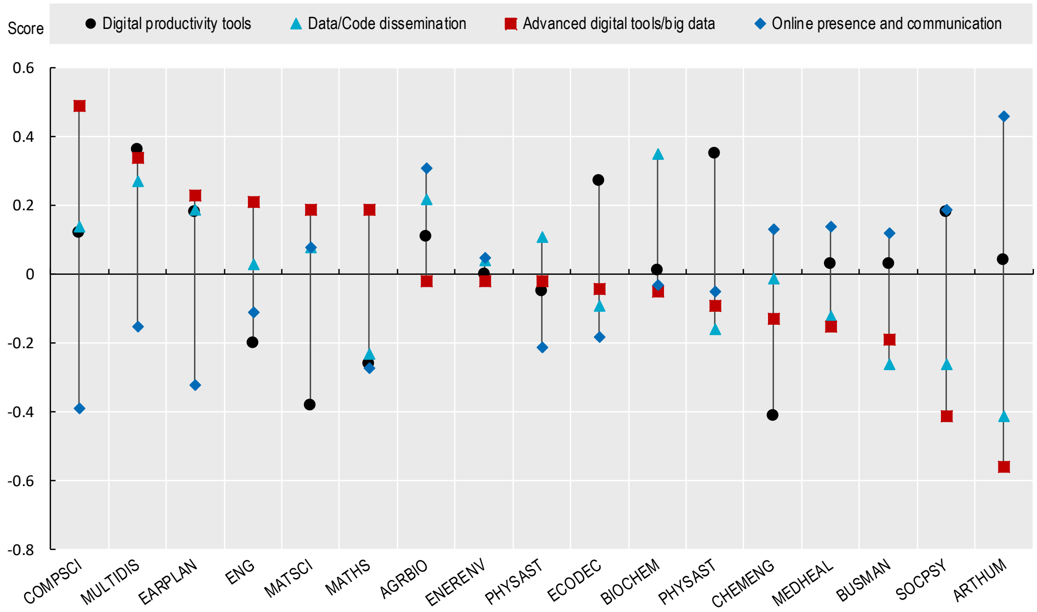
Average factor scores by field of science and country of residence are provided in Figure 3.28 and Figure 3.29, respectively. These figures capture the extent to which scientific authors in different groups are involved in digital practices across the four digitalisation dimensions. Scientific authors publishing in different domains have distinct digitalisation patterns. Consistent with the findings presented in the previous section, the intensity of adoption of advanced digital practices is highest in Computer sciences, Earth and planetary sciences, Engineering, and Materials science. Data/code sharing practices appear to be more diffused in biochemistry, genetics, molecular biology, immunology and microbiology than in other fields. This confirms previous studies documenting the strong uptake in the development of infrastructure, resources and policies that promote data sharing in biomedical sciences, especially in genetics (Kaye et at., 2009; Choudhury et al., 2014).

Authors in Mathematics combine high scores in advanced digital methods with low scores in terms of data/code dissemination practices. Authors in the Arts and humanities as well as in Agricultural and biological sciences present the highest scores for the latent factor that reflects intensity of digital presence and communication of research work. The fields of Engineering and Materials science are characterised by a lower intensity in the use of digital productivity and collaboration tools compared to other fields. This appears to be explained by the fact that these tools, as presented in the ISSA2 questionnaire, are more associated to academic usage. Statistics by sector of affiliation appear to confirm this result.

Contrasting patterns of digitalisation are also present across different economies (Figure 3.29). As noted at the outset of this paper, country-level indicators should be interpreted with caution taking into account response challenges and limited precision of estimates for some economies. Furthermore, indicators at the country level involve compositional effects owing to differences in the national distribution of scientific activity and the national specificity of authors publishing in journals and proceedings covered in Scopus. The results suggest that authors based in China, India, and Turkey stand out in terms of average use of advanced digital tools.

Figure 3.28. Patterns of digitalisation in science, by field

Averages of standardised latent factor scores



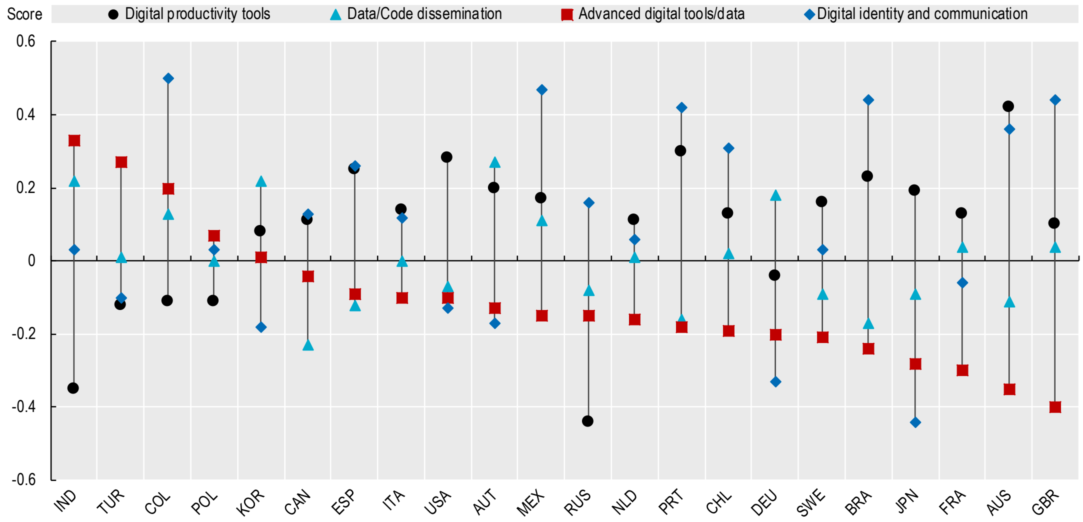
Note: Figures refer to the weighted average of the four standardised factor scores representing latent digitalisation indicators within each science field. Sampling weights adjusted by nonresponses are used in the weighting procedure. The factor analysis is based on the responses by scientists to questions relating to the use of digital tools or adoption of digitally enabled practices. The resulting four factors have been interpreted and labelled based on how strongly they correlate with the survey-based underlying variables. Factor scores are estimated in units of standard deviations from their means and represent a person's relative position on a latent factor compared to the rest of the individuals.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

This survey-based result on advanced computational intensity is consistent with the analysis of indicators derived from the content of the authors' scientific publications. Documents identified as having an AI-related content, as reported in OECD (2019), have been used to estimate the AI-intensity at a country and field level. The Spearman's rank correlation between the average of this measure and the latent factor scores relating to the adoption of advanced digital practices is particularly high at the level of fields (0.69) and still positive and rather significant (0.43) at the level of countries.[10]

[10] This is also the case at the micro respondent level. The AI intensity indicators at field and country/economy levels are reported in Figure A A.3 and Figure A A.4, respectively. Computer sciences, Engineering, Mathematics, and Earth and planetary sciences rank among the top-7 fields in terms of both adoption of digital practices and share of domestically based corresponding authors with AI-related publications. Likewise, China, India, Malaysia and Turkey are among the top-7 countries in relation to both AI intensity indicators.

Figure 3.29. Patterns of digitalisation in science, by country of residence

Averages of standardised latent factor scores



Note: See notes to Figure 3.28. Only countries with more than 100 observations are shown.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

This lends a considerable degree of external validity to the survey-based measures which, in addition, provide indicative profiles of the areas in which different countries have stronger science digitalisation strengths. For example, scientific authors in the United States exhibit *on average* a fairly balanced profile, with higher overall results in terms of use of digital productivity and collaboration tools relative to the three other dimensions which are slightly below the world average. In contrast, in countries[11] like China and India the "digital profile" is significantly more unbalanced and tilted towards the use of advanced digital tools, the opposite of what is found for countries like Australia, United Kingdom and Ireland, which score particularly low on that dimension but display very large scores in terms of digital identity and communication. Small economies like Hungary, Estonia, Slovenia and New Zealand exhibit high intensity in the use of digital productivity and collaboration tools, while these are particularly low in the case of Russia, China and India.

It should be noted that many of these differences across countries disappear or become insignificant as one controls for compositional differences such as those related to field specialisation. A country's degree of specialisation in some areas accounts for a great part of the observed variation, but some of these differences still remain and should be in the future the focus of additional analysis, which may require in some instances to boost the number of survey participants.

[11] Precise estimates of the average factor scores for a number of countries including China, Hungary, and Estonia, among others, cannot be calculated because the number of responses available in these cases is not considered to be large enough for producing accurate and representative estimates. Therefore, figures referring to these countries are not shown in Figure 3.29.

4. Factors influencing digitalisation of science

This section aims to explore the link between an author's characteristics such as gender, age, education, research methods, and his/her digital practices. Although a causal relationship between these indicators cannot be assessed given the "snapshot" nature of the data collected, the results presented in this section can still provide important insights into the patterns of use of digital tools in science. The multivariate analyses presented seek to account for variation in digitalisation intensity across different groups of scientific authors keeping other characteristics constant. This section's results are based on a regression analysis which explains the variation across authors in the 4 digitalisation-related measures derived in Section 3.5 according to authors' characteristics. All regression results are provided in the Annex. The results of this type of work can facilitate a better understanding of factors that appear to facilitate or inhibit the adoption of digital tools in science and ultimately help explain their impacts, as considered in the following section.

4.1. Links between digitalisation patterns and the profile of scientific authors

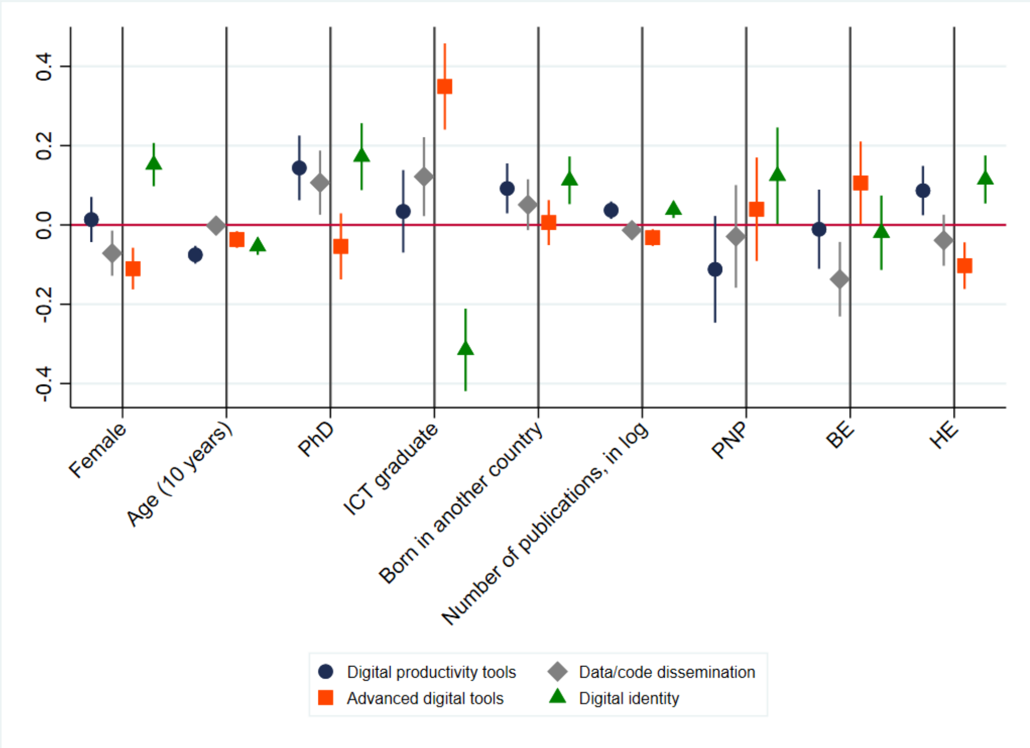
Regression analysis examines the correlation between the use of digital tools and a number of scientists' characteristics, including gender, age, education, employment, place of birth, and research performance, while controlling for research field and country of residence (Figure 4.1).

The digital gender divide in science

Regression results point to significant digital divides between men and women even for a highly professional group such as that of scientific corresponding authors. Female authors score significantly lower than their male counterparts in the use of advanced tools. They are also significantly less involved on average in data and code sharing practices. However, women score higher than men in the factor representing digital identity and communication of research work. For this reason, it is more appropriate to speak about gender divides than a single modality of gender differences around digitalisation. Concerning the use of digital productivity and collaboration tools, no significant differences are found.

Figure 4.1. Links between digitalisation patterns and author profile

Least square regression coefficients (standard deviation units) and confidence intervals



Note: Full regression results are provided in Table A D.1. The coefficient on *Age* corresponds to 10 years increase in age. *ICT graduate* is a dummy variable, which is equal to 1 if an author holds a degree in computer and information sciences, electrical engineering, electronic engineering, or information engineering, and 0 otherwise. GOV is the baseline of PNP, BE and HE coefficients.
How to read: One unit increase in the author profile-related variables (or 1% increase in the number of publications) leads to a change in the indicators on digitalisation patterns in terms of points of standard deviation from their means and as captured by the regression coefficients. For instance, 1% increase in an author's number of publications corresponds to an increase of 0.04 points of standard deviation in the use of digital productivity tools, whereas the adoption of data/code dissemination practices among women is 0.07 points of standard deviation lower than that of men.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Age and other digital divides in science

Age

An author's age appears to be strongly correlated with a lower score in all measures of digital intensity with the exception of digital productivity and collaboration tools, where the estimated effect of age is also negative but not statistically significant. A difference of 10 years of age drives on average a significant gap in the factor capturing the take up of advanced digital tools of 7% standard deviations. The existence of a generational effect is common to measures of digitalisation in other domains (OECD, 2019). In national contexts where the scientific population is ageing particularly fast, this can be a serious concern towards ensuring the pursuit of high quality and impactful research when that needs to be supported by the effective adoption of digital tools.

Qualifications

These age effects are indeed robust to the inclusion of information on whether the authors graduated in an ICT subject.[12] ICT graduates do not only score higher than their non ICT specialist counterparts in the use of advanced digital tools, but they do so also in relation to data and code dissemination activities. Interestingly, these authors with an ICT-related qualification exhibit low scores on the digital identity and communication factor.

Scientific authors with a doctorate degree tend to exhibit high digitalisation scores with the exception of those corresponding to advanced digital tools, for which authors without a doctorate tend to exhibit a higher score.

Institutional sector of employment

Authors in the Higher education sector score higher in digital productivity and collaboration as well as identity and communication factors, but significantly lower in relation to the adoption of more advanced digital practices, compared to authors in the Government sector, who represent the estimation baseline. Scientists in the Private non-profit sector achieve comparable results in terms of digital identity and communication scores to those in the Higher education sector. In contrast, authors in the Business enterprise sector exhibit the highest scores in terms of advanced digital tools, while featuring the lowest scores in data and code sharing.

These results point towards a significant advantage in the business sector towards the use advanced computational practices over scientists in other sectors. This might call for an exploration of career incentives and market opportunities for highly qualified researchers in AI and related areas. As alluded to earlier in this report, the way in which scientists engage in digital activity seems to be very closely related to their organisation's approach to disclosure and appropriability of research outcomes. Academically oriented authors appear to be more inclined to engage in digital activities that enhance academic productivity and collaboration and communication of their results and their profiles. Even within academia, the results obtained thus far point to significant differences in approaches to digital practices with the aim of facilitating access to data and code. Scientists in Government institutions appear to play a key role with regard to this dimension of digitalisation.

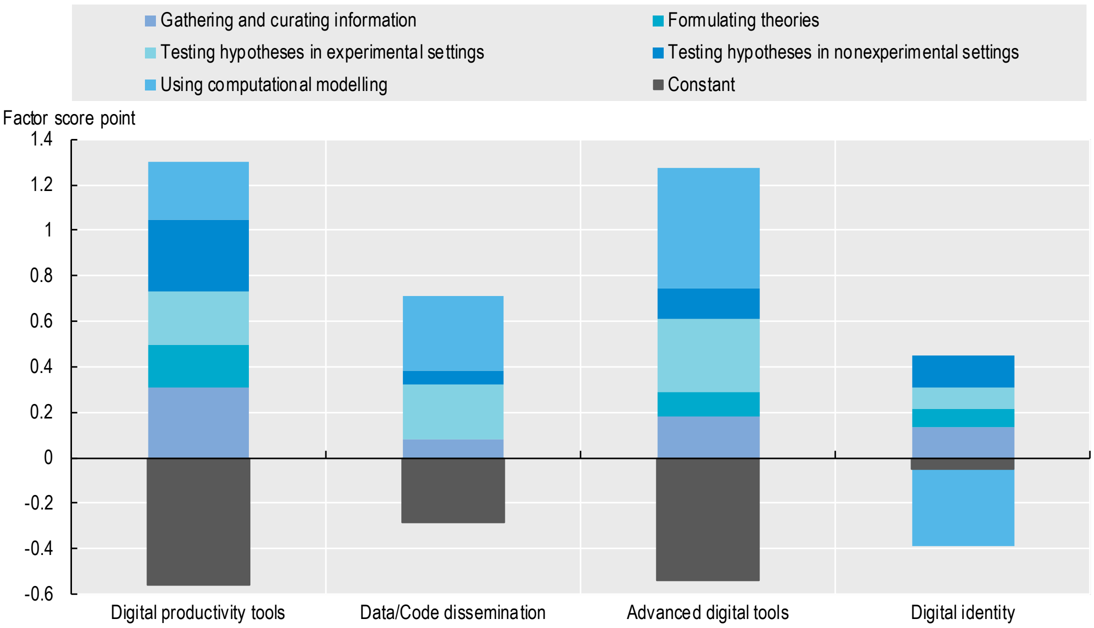
[12] This accounts for the differences within research fields driven by students with and without advanced degrees in ICT subjects.

4.2. Links between digitalisation patterns and research methods

To complete the analysis of digitalisation patterns this subsection explores how the digitalisation factors relate to the fundamental research modes adopted by scientific authors and their teams. Figure 4.2 presents the analysis of each of the four digitalisation indicators, which are regressed on five variables reporting on an author's involvement in the research modes or activities listed in Section 3.1: data gathering and curation, theory formulation, use of computational modelling, hypothesis testing in experimental settings, and hypothesis testing in empirical, non-experimental settings.

Figure 4.2. Links between measures of digital intensity and research methods

Cumulative estimated marginal effects of research methods on digitalisation factors



How to read this chart: each coefficient represents the average difference in factor scores between the respondents that carry out a given research activity and those that do not, whereas the average factor score for each group of respondents is given by the sum of each coefficient with the constant. Since the (z- standardised) factor scores can be interpreted in terms of units of standard deviation from the sample mean, when this sum is below 0 it follows that the respondents involved in the corresponding research method are below the average with regard to the specific factor. For instance, authors that are involved in gathering and curating information, formulating theories and testing hypotheses in experimental settings at the same time, or those that are involved in both testing hypotheses in non-experimental or experimental settings and use of computation modelling score above the average with regard to the use of more advanced digital tools.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

As authors typically engage in one or more research modes and there are no other covariates in the model, the blocks within each column (one per factor) can be combined (summed up and compared to the baseline) to indicate which research profiles result in digitalisation scores above or below the global average (normalised to zero). The main purpose of the chart is to highlight which are the research modes that boost each digitalisation scores, which can be gauged by their relative size. Thus, for example, the adoption of advanced digital tools (third bar) is primarily, but not only, boosted by researchers that carry out research using computational modelling tools, followed by those who test hypothesis in experimental settings. This suggests that it may be appropriate to speak of a fourth paradigm that is distinct from the traditional computational modelling approach and that is relevant to all data-rich research areas.

In the case of the adoption of digital productivity tools, contributions from different modes are more equally distributed, with significant contributions from authors engaged in gathering and curating information as well as those involved in testing hypotheses in non-experimental settings. The mode of information gathering and curating has apparently very limited impact on the digital mode of data and code dissemination, suggesting that there are significant opportunities for repository-based activity and digitalisation of archives provided that the right incentives and tools can be provided. The theory formulation mode is the one with the smallest independent contribution to digitalisation scores. Theorists make use of digital tools mostly when they engage in complementary modes of research, for example, computational modelling.

5. Drivers and impacts of digitalisation in science

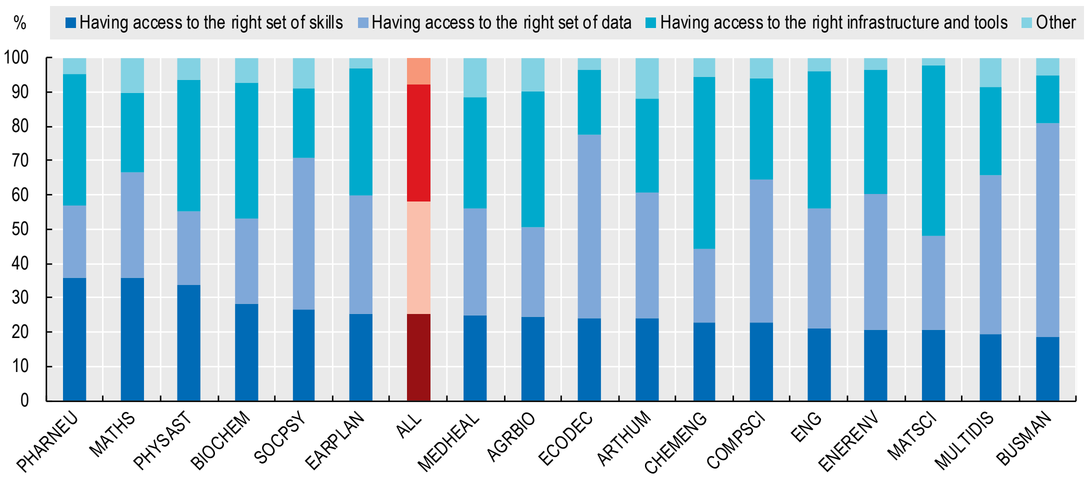
Having mapped the multifaceted nature of digitalisation in science as well as its potential explanatory factors linked to the modes and features of scientific research, the main questions remaining to be addressed in this paper relate to the identification of the key drivers and impacts of digitalisation. Such questions are unpacked in three steps: identifying key barriers faced by scientists in their work; examining the link between digitalisation and measures of scientific impact and exploring authors' personal views on how they themselves view the impact of digitalisation in science.

5.1. Challenges faced by scientific authors in digital environments

Directly following questions on their use of data and code, scientific authors were asked in ISSA2 to report on the greatest challenge they face in their scientific work in digital environments. As illustrated in Figure 5.1, authors appear almost equally split across the three main options provided to them, namely skills, data, and infrastructure and tools. Infrastructure and tools are more important for scientists working in chemical engineering, material science and a number of life sciences. Access to data was considered most challenging in the fields of Economics and business, as well as in other social sciences. The fields to put the greater emphasis on skills, the least reported challenge on average, were those of Pharma and neurology, Mathematics, and Physics and astronomy.

Figure 5.1. Most important challenge faced by scientific authors in their research work

Percentage of authors within each field selecting the relevant option

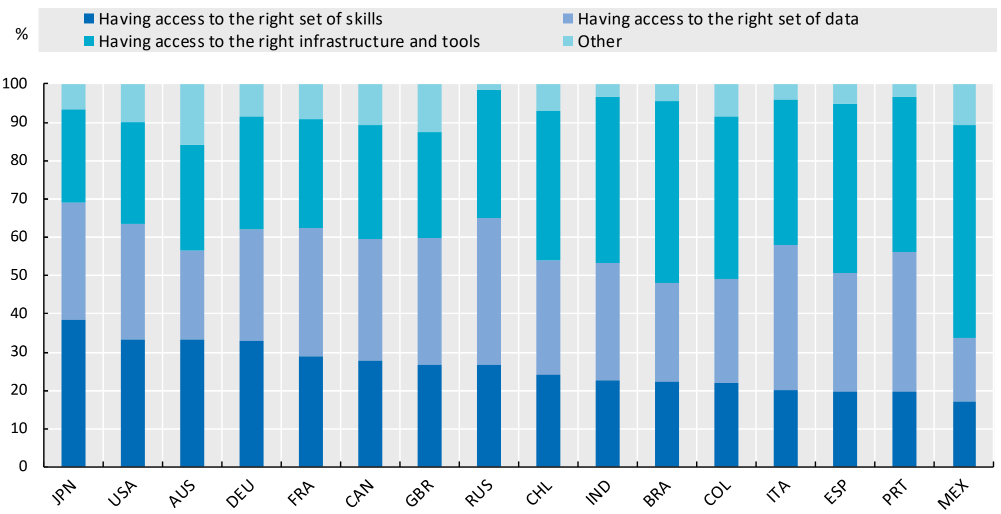


Note: Weighted estimates based on sampling weights adjusted for nonresponse.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Scientists in Japan and the United States appear to be assigning greater importance to challenges posed by skills. In contrast, those in Mexico and Brazil are more likely to point to infrastructure-related challenges (Figure 5.2).

Figure 5.2. Most important challenge faced by scientific authors, by country of residence

Percentage of authors within each field selecting the relevant option



Note: Weighted estimates based on sampling weights adjusted for nonresponse.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Diving into the types of skills and infrastructure that scientific authors highlight as most important for their work provides evidence on the factors, which, if not available, may eventually constrain their work. Starting with skills, scientists across different areas (from the natural sciences to the arts and humanities) coincide in highlighting data collection and curation skills as important with the greatest frequency (Figure 5.3). The appreciation of the importance of advanced programming skills varies significantly across areas, being more critical in Computer sciences, Earth and planetary sciences, and in Mathematics. The knowledge of legal aspects related to intellectual property, privacy and confidentiality appears to be more essential for authors in medicine and health professions. Authors in Business management and accountings, and those in Agricultural and biological science report data collection and curation skills to be more important than authors in other field.

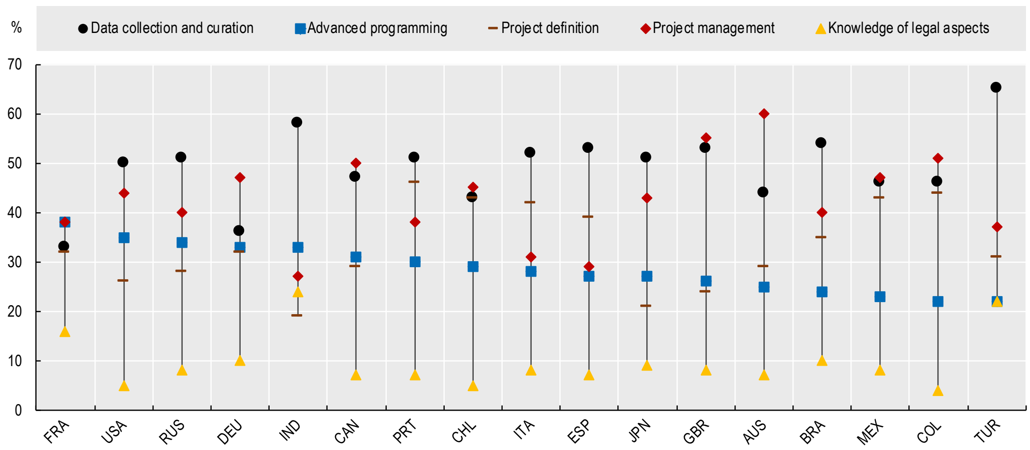
Figure 5.3. Most important skills for scientific authors' research work

Percentage of authors who deem each type of skill as important

Note: Weighted estimates based on sampling weights adjusted for nonresponse. Respondents can select a maximum of two options.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Advanced programming skills appear to be more important for authors in France and the USA, whereas data collection and curation skills are more crucial in India and Turkey (Figure 5.4)

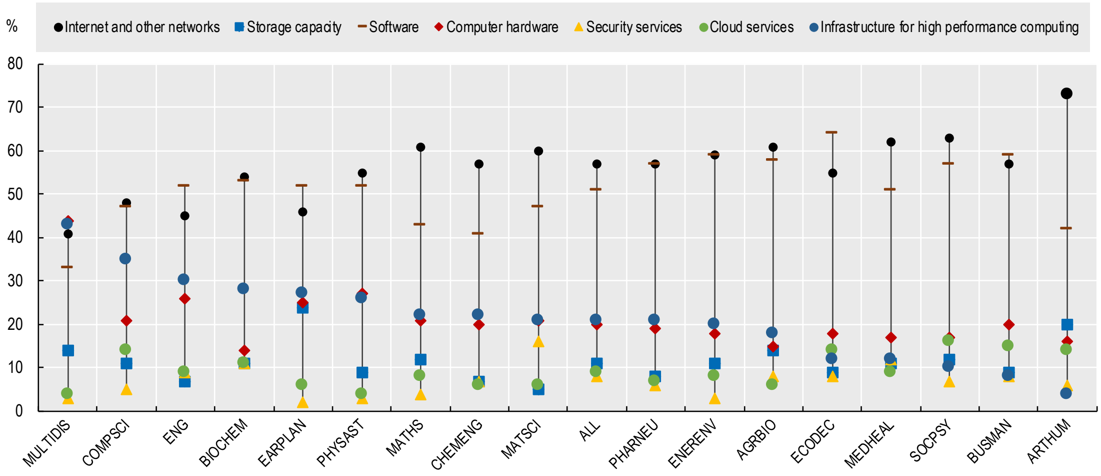
Figure 5.4. Most important skills for scientific authors' research work, by country



Note: Weighted estimates based on sampling weights adjusted for nonresponse. Respondents can select a maximum of two options.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure 5.5. Most important infrastructure for scientific authors' research work, by field

Percentage of authors deeming a given type of infrastructure as important

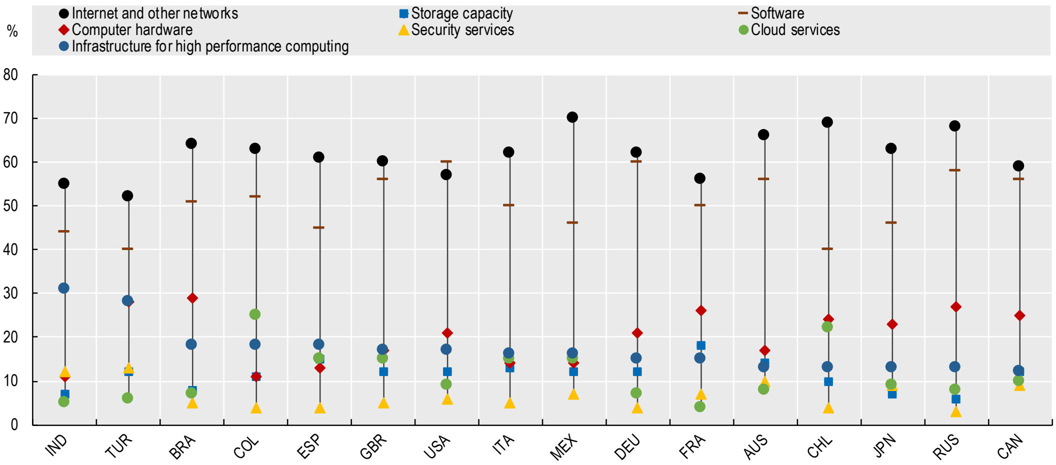


Note: Weighted estimates based on sampling weights adjusted for nonresponse. Respondents can select a maximum of two options.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Considering information and communication infrastructures (Figure 5.5), authors across different fields regard access to Internet and networks, as well as software as of major importance. Access to high performance computing infrastructure is particularly important, in addition to authors with publications in Multidisciplinary journals, to those in Computer sciences, Engineering, and Biochemistry, genetics, molecular biology, immunology and microbiology at close to 30% of cases. Storage capacity is of particular important in Earth and planetary sciences.

Figure 5.6. Most important infrastructure for scientific authors' research work, by country

Percentage of authors deeming a given type of infrastructure as important



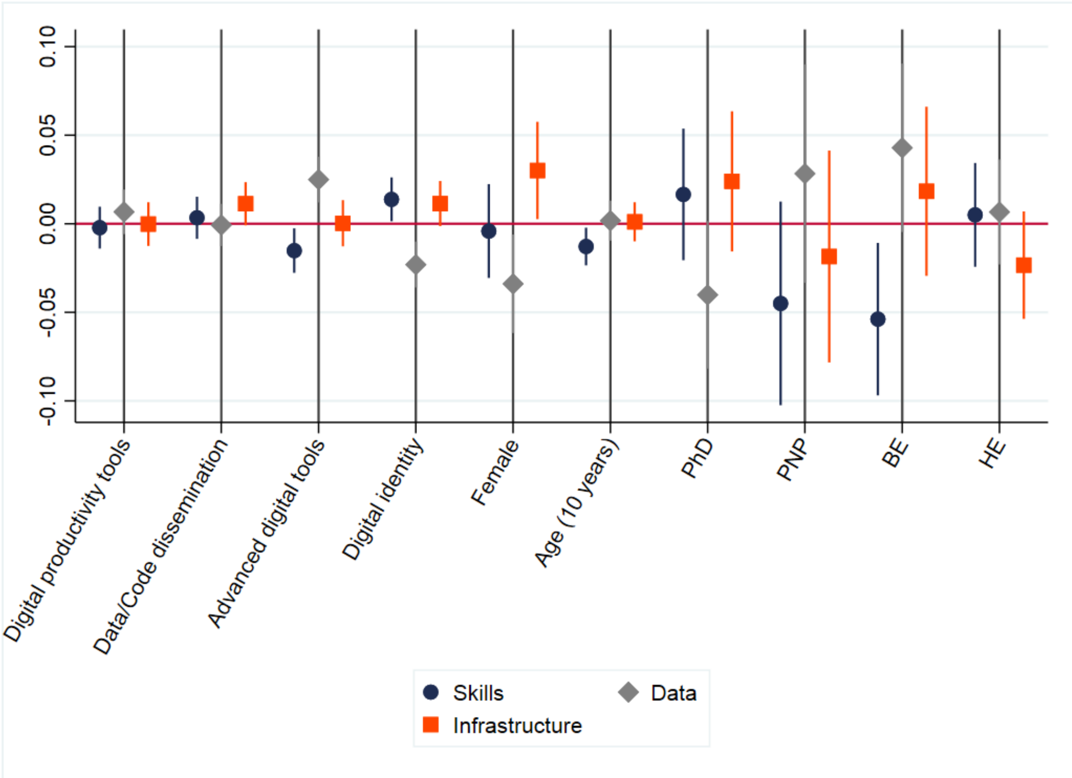
Note: Weighted estimates based on sampling weights adjusted for nonresponse. Respondents can select a maximum of two options.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Infrastructure for high computing performance is particularly important in India and Turkey, whereas storage capacity and security services are slightly more critical in France and Turkey, respectively, than are in other countries (Figure 5.6).

Multivariate analysis of the probability of reporting one particularly issue as the key challenge, controlling for digitalisation profiles, reveals some additional patterns of interest. Whereas the probability of encountering problems in accessing the right set of skills does not seem to be correlated with an author's gender, female authors appear more likely to report access to infrastructure as a challenge (Figure 5.7). In contrast, age and an author's sector of employment seem to be relevant only in relation to skills: younger authors and those in the business enterprise sector seem to be less likely to encounter difficulties in accessing the right set of skills. Authors with higher scores in the factor capturing the use of advanced digital tools are more likely to encounter problems in accessing data, but they are less likely to lack the right set of skills.

Figure 5.7. Authors' perceived challenges in their research work and digitalisation patterns

Least square regression coefficients and confidence intervals

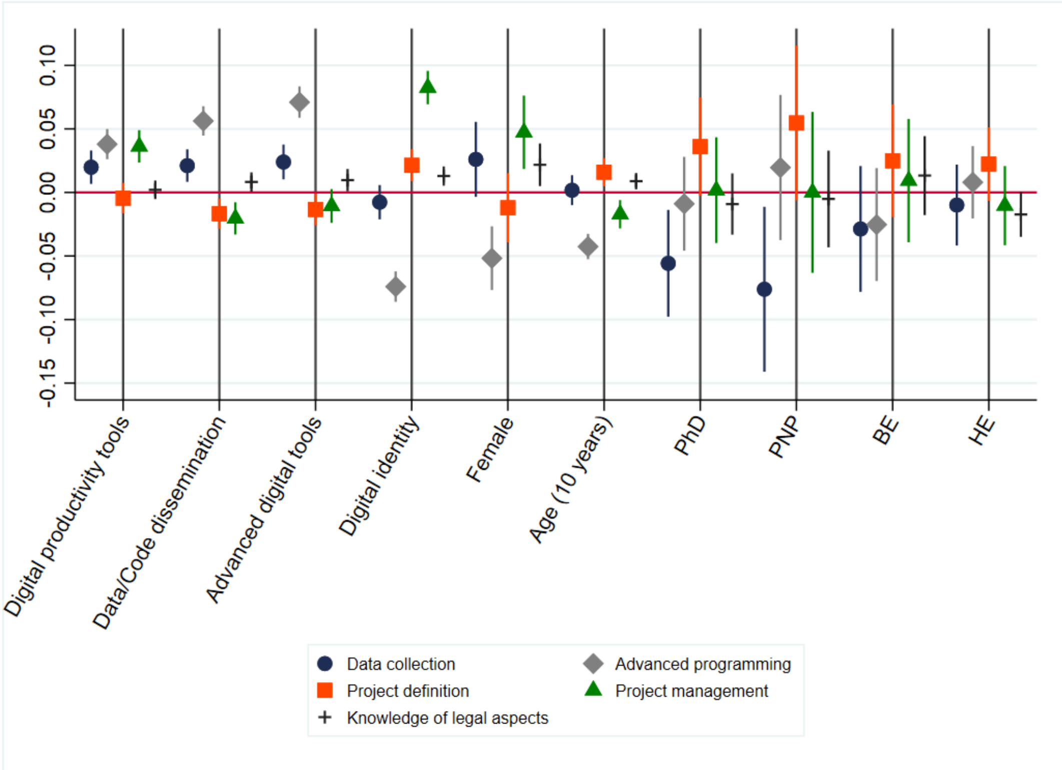


Note: Regression full results are provided in Table D.3 PNP stands for private, non-profit sector. BE stands for business enterprise sector. HE stands for higher education sector. Government (GOV) is the baseline of PNP, BE and HE coefficients.
How to read this chart: Authors with one standard deviation higher degree of intensity in advanced digital tools are close to 2% less likely to perceive skills as the most important challenge. Female authors are nearly 3% more likely than men to report infrastructure as a challenge.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Analysis of the probability of reporting specific skills as important indicates that data collection and curation skills are more likely to be reported by female authors and those with higher digital factor scores in all dimensions except for those with high digital identity scores and with a PhD (Figure 5.8). In contrast, programming skills are more relevant to authors involved in advanced digital practices than they are to the others. Project management skills are more likely to be reported as important by individuals with higher digital identity and communication scores and by women.

Figure 5.8. Most important skills and digitalisation patterns

Least square regression coefficients and confidence intervals

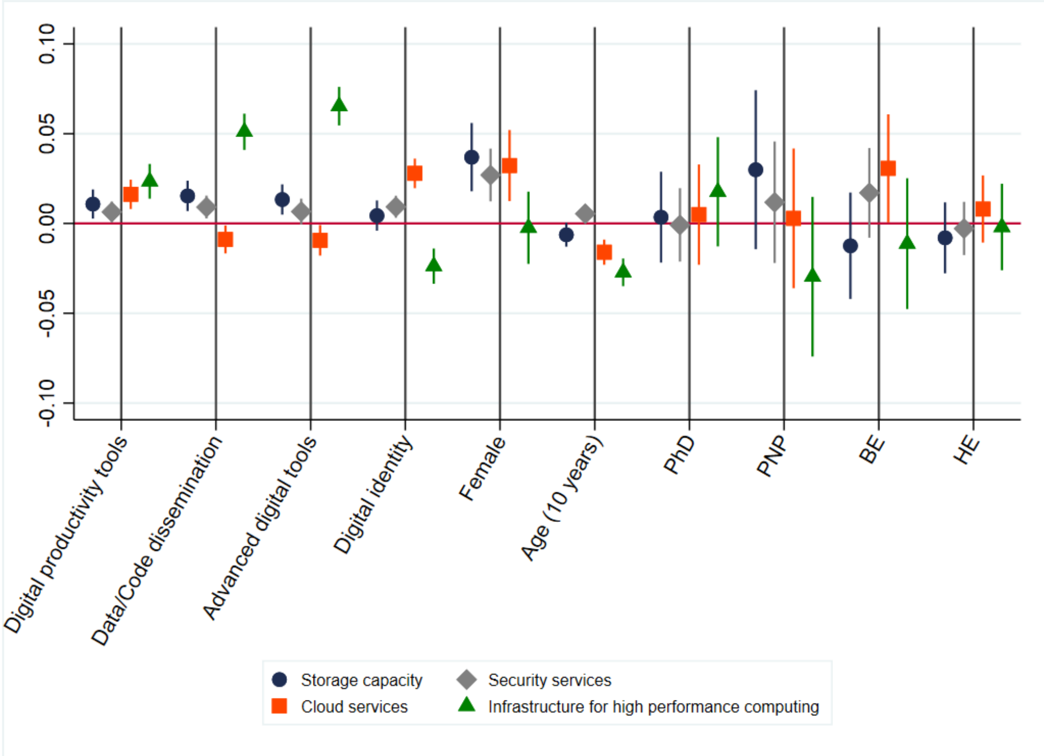


Note: Regression full results are provided in Table D.4. PNP stands for private, non-profit sector. BE stands for business enterprise sector. HE stands for higher education sector. Government (GOV) is the baseline of PNP, BE and HE coefficients.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Similar analysis on the importance of specific types of infrastructure highlights the higher importance of cloud services for scientists in the business sector. High performance computing is much more important to authors using advanced digital tools than it is to the others (Figure 5.9). In addition, the infrastructure storage capacity seems to be essential to all authors with high digitalisation scores for advanced digital tools and data and code dissemination.

Figure 5.9. Most important infrastructure and digitalisation patterns

Least square regression coefficients and confidence intervals



Note: Complete regression results are available in the Annex. PNP stands for private, non-profit sector. BE stands for business enterprise sector. HE stands for higher education sector. Government (GOV) is the baseline of PNP, BE and HE coefficients.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

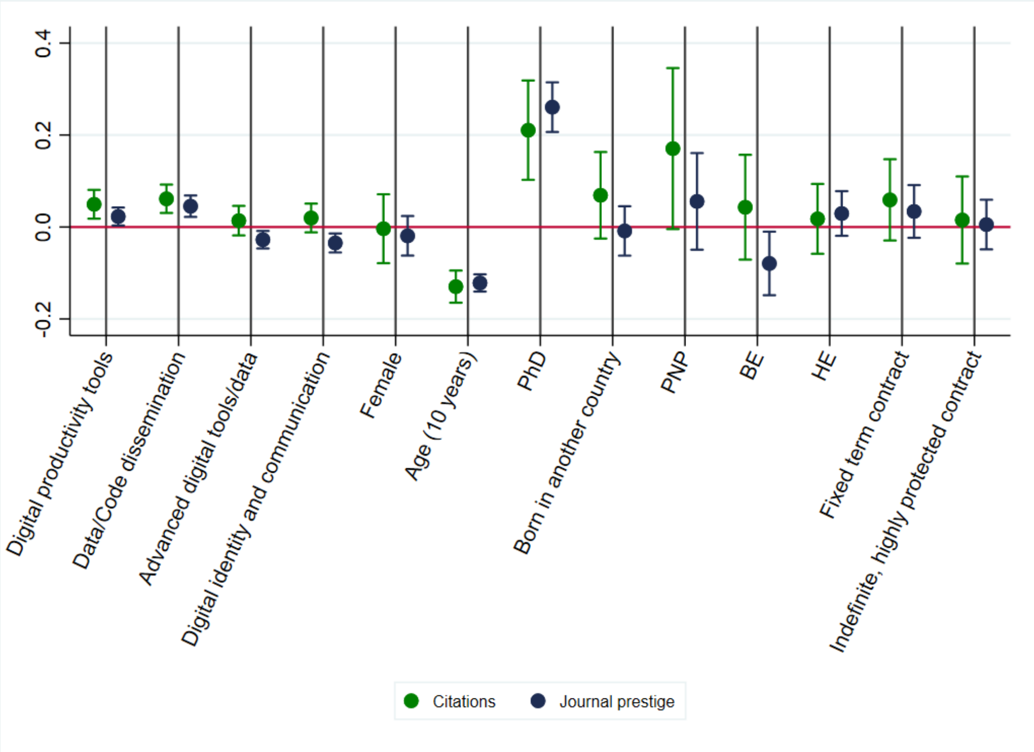
5.2. Digitalisation and the impact of research activity

Citation influence and journal prestige

The data collected allows to explore the extent to which often used measures of research influence, such as normalised citations or journal prestige, are associated with measures of digitalisation intensity. The results of the analysis suggest that both proxy measures of research quality are positively correlated with the use of digital productivity tools and data/code sharing (Figure 5.10). This might be suggestive of positive effects of collaboration and openness that may be enhanced by such digital practices. The propensities to use advanced digital tools and engage in digital identity and communication practices are negatively correlated with the measure of journal prestige while they are positively but not significantly correlated with the measure of citation impact. Analysis of the role of authors' personal characteristics indicates that there is no gender difference with respect to any of the two measures considered, while holding a doctorate degree has a significant positive impact. Scientific authors born in another country tend to attain a higher average citation impact, without any difference in terms of journal prestige. ICT graduates and those in the business sector appear to be more likely to publish documents featured in lower prestige journals, although their citation impact is comparable to that of their peers if not higher. Finally, the unexpected, negative correlation between age and research quality is likely to be caused by the inclusion of a variable referring to the total number of publications in the regressions, which appears to capture most of the correlation between age and the two measures of research influence adopted.

Figure 5.10. Drivers of lifetime citation impact and journal prestige

Least square regression estimates and confidence intervals



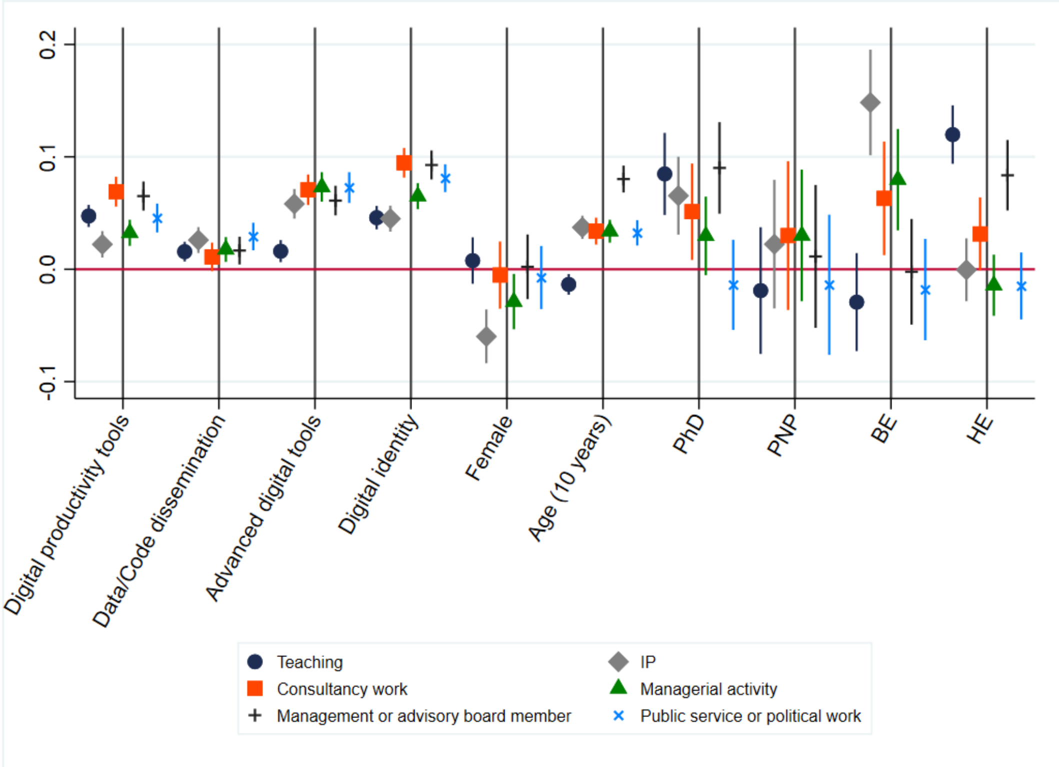
Note: The citation measure is the average of the field (and cohort/type of document)-normalised citation rate per document over the 1996-2017 period. The journal prestige measure is the average of an author's Scimago's SJR score over her publications over the same period. Full regression results are provided in Table A D.7 PNP stands for private, non-profit sector. BE stands for business enterprise sector. HE stands for higher education sector. Government (GOV) is the baseline of PNP, BE and HE coefficients.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Other influences of research

Since the impacts of research and related activity can extend beyond the immediate influence of publications on the scientific community, it is important to explore to what extent digitalisation relates to broader measures of potential impact, as implied by activities potentially incurred by scientists. Analysis of ISSA2 data suggests that authors reporting a more intensive use of digital tools tend to engage in more activities that can result in broader impact mechanisms (Figure 5.11). Higher digitalisation scores on the use of advanced digital tools and digital identity and communication are associated with a higher probability of reporting the registration of IPRs, engagement in business management activity, provision of research services and consultancy work. The results also show that women and younger researchers are less likely to engage in most of these activities. This suggests some degree of caution when recommending the use of measures of broader engagement as incentives to research.

Figure 5.11. Engagement in activities beyond core research and digitalisation patterns

Least square regression estimates and confidence intervals



Note: Regression full results are provided in Table D.6. PNP stands for private, non-profit sector. BE stands for business enterprise sector. HE stands for higher education sector. Government (GOV) is the baseline of PNP, BE and HE coefficients.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Once again, the results point to systematic differences in the motivations and engagement of scientists working in different sectors. As expected, commercially-oriented engagement is more common in the Business sector and less so in Higher education compared to Government. However, scientists in HEIs engage more frequently in providing consultancy services and advisory roles than the latter, implying a degree of engagement with demand for scientific and technical services which may come from different parts of the economy, not only the business sector. Personal development activities such as mentoring and teaching are as expected more common among authors in Higher education.

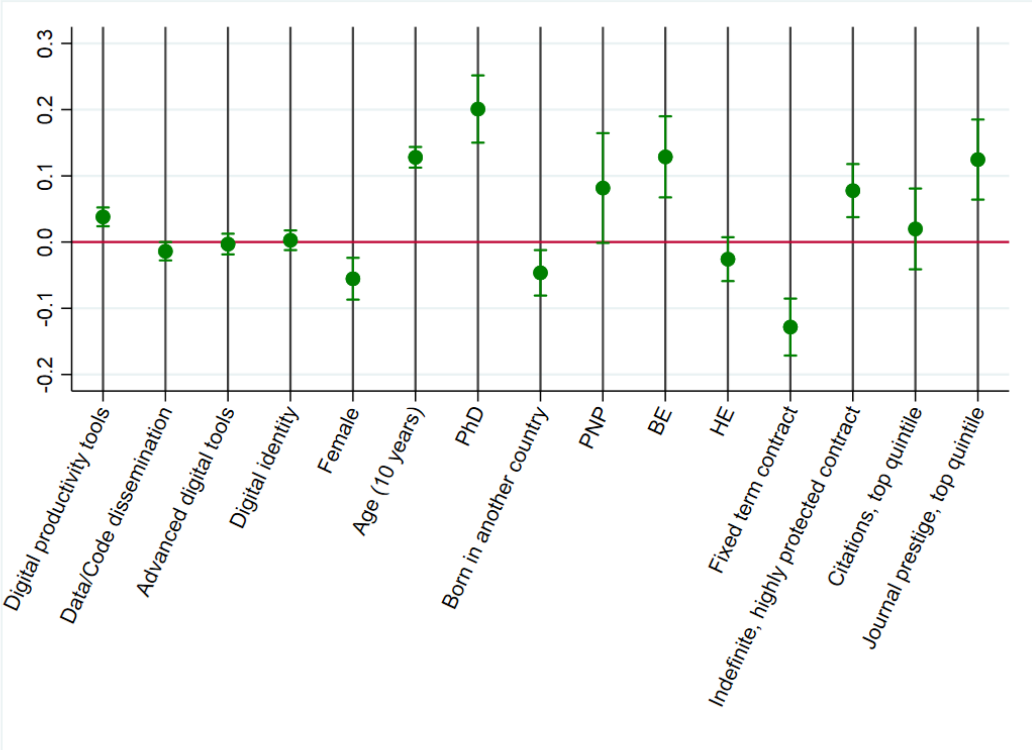
Digitalisation and research careers

How does the market price scientists' traits and behaviours?

In order to provide evidence on the link between digitalisation patterns and authors' career performance, Figure 5.12 reports the coefficient of a regression analysis of the relationship between authors' annual income gross of tax and her digital profile and other personal characteristics, including the average number of hours worked. These estimates are therefore interpreted as indicative of the way in which the market rewards the competences and efforts of scientific authors or appears to penalise certain traits.

Figure 5.12. Authors' earnings and their profiles

Least square regression estimates and confidence intervals



Note: Regression full results are provided in the annex. PNP stands for private, non-profit sector. BE stands for business enterprise sector. HE stands for higher education sector. Government (GOV) is the baseline of PNP, BE and HE coefficients. Earnings are included in the regression in logarithmic form. The survey questionnaire asks for the range of an author's annual earnings. Estimates used in the regressions are based on mid points for the income bands, with an adjustment for lower and upper bands. National differences in purchasing power are absorbed into country-specific fixed effects.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Among the four digitalisation latent factors considered, only the intensity in the use of productivity and collaboration tools appears to be slightly associated with higher income levels. In contrast, after controlling for a long list of characteristics, there is no evidence of a premium from displaying behaviours associated to the use of advanced digital tools or data and code sharing practices. Consistently with other general labour economics results, women earn significantly less than men by around 5 to 6 percent. The same type of implied penalty applies to individuals born in a different country. Experience is rewarded in the scientific marketplace, as implied by age and number of publications, while a doctorate qualification is associated with a 20% higher level of pay[13]. Earnings are also higher for the authors in the Business enterprise sector and Private non-profit compared to those in Government whereas those in the Higher Education sector are the least well paid. Controlling for all these characteristics, there does not appear to be an implied compensation for lack of job security. Individuals holding a secured (tenure) indefinite contract are better paid than those with less secure indefinite contracts, while those with fixed term contracts are the least well paid.

[13] Differences by cohorts have not yet been explored.

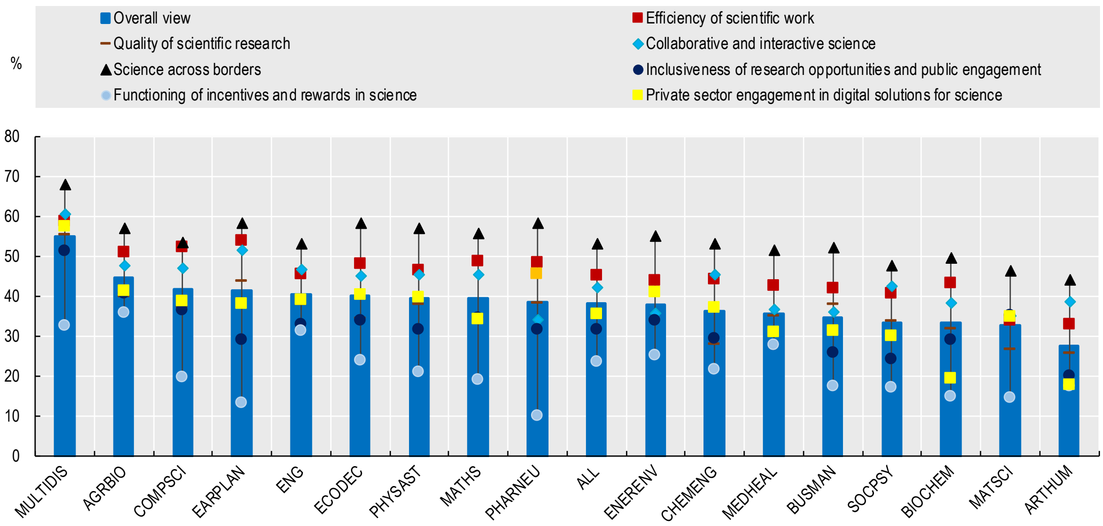
Scientists’ income is related to measures of scientific productivity, suggesting that the metrics reportedly used (as highlighted in section 3.4) do bear implications for research careers. What is particularly revealing is that the marketplace appears to confer a stronger value on measures of journal prestige than on measures of normalised actual citations received by authors, even when both are averaged over their research careers. An author in the sample’s citation top quintile earns a non-statistically significant 2% more than one in the lower quintile, whereas on in the sample’s top prestige quintile can expect to earn on average 13% more. This finding is very relevant for the design of policies and research assessment practices, as currently authors appear to have very strong incentives to pursue the publication in their work in outfits with higher citation influence ratings.

5.3. Scientific authors’ views on the impact of digitalisation in science

Conducting a survey provides the opportunity to ask scientists directly about their views on digitalisation along a number of relevant dimensions. ISSA2 specifically asked respondents to position themselves between two statements which highlighted somewhat opposing statements about digitalisation trends and their impact on the world of scientific research. The dimensions covered related to opportunities for increasing the efficiency and productivity of work (opposed to rising burdens); addressing hitherto intractable problems (vs encouraging low quality research); facilitating inclusion and collaboration (vs promoting exclusion and excessive competition); opening science to society (vs celebrity science and external pressures); and improving incentives (vs distorting behaviour).

Figure 5.13. Scientific authors’ views on the digitalisation of science and its potential impacts

Average sentiment towards “positive” digitalisation scenarios, as percentage deviation from mid-viewpoint



Note: Survey respondents were asked to rate opposing scenarios on different dimensions from (1 = fully agree with a negative view) to (10 = fully agree with a positive view). For interpretability, weighted average scores on each dimension and the general summary view (weighted average across dimensions) are presented as percentage deviations from the midpoint. This means, for example, that with respect to the subject of “Science across borders” (triangle label), the average score differs from the midpoint by around 50%, corresponding to authors being oriented towards the positive outcome, relative to the neutral perspective. Weighted average scores take into account the sample design and nonresponse.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

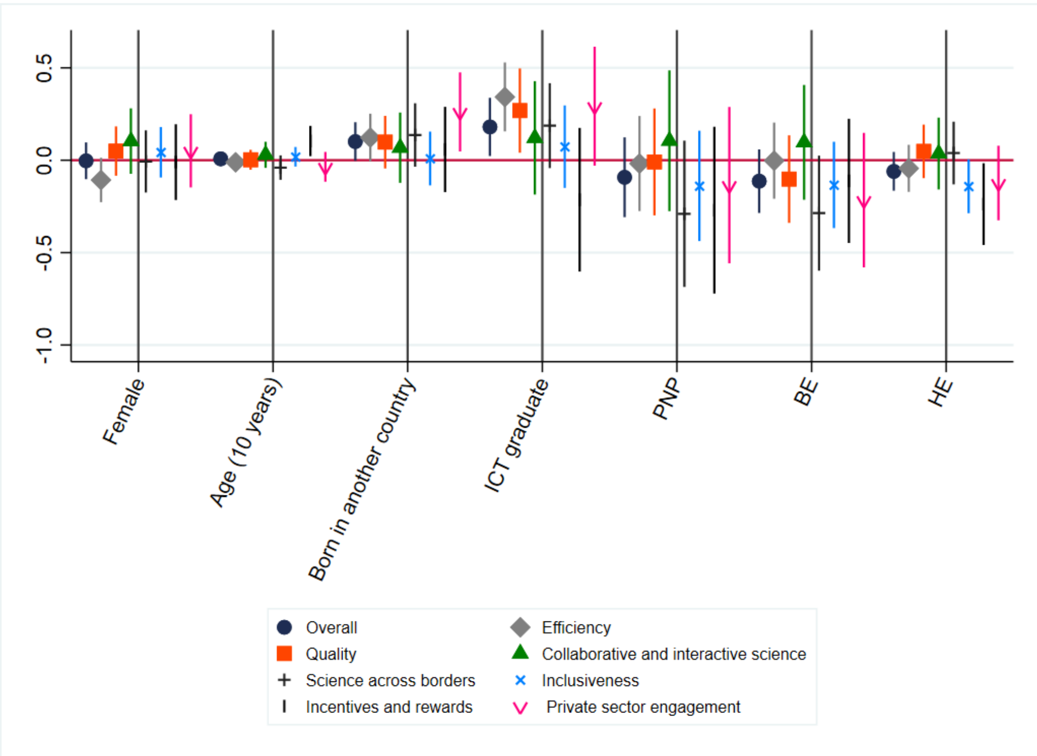
As shown in Figure 5.13, scientists lean on average towards having a positive attitude to digitalisation, especially those in Agricultural and Biological sciences and Computer sciences, namely the areas with greater overall digital scores. In all fields, scientists seem to expect that the use of digital tools in science will be coupled with higher collaboration in research, especially across borders, and greater efficiency of research work. While still showing an overall positive attitude, authors seem to be more uncertain regarding the impact of digitalisation on how research is evaluated and incentivised.

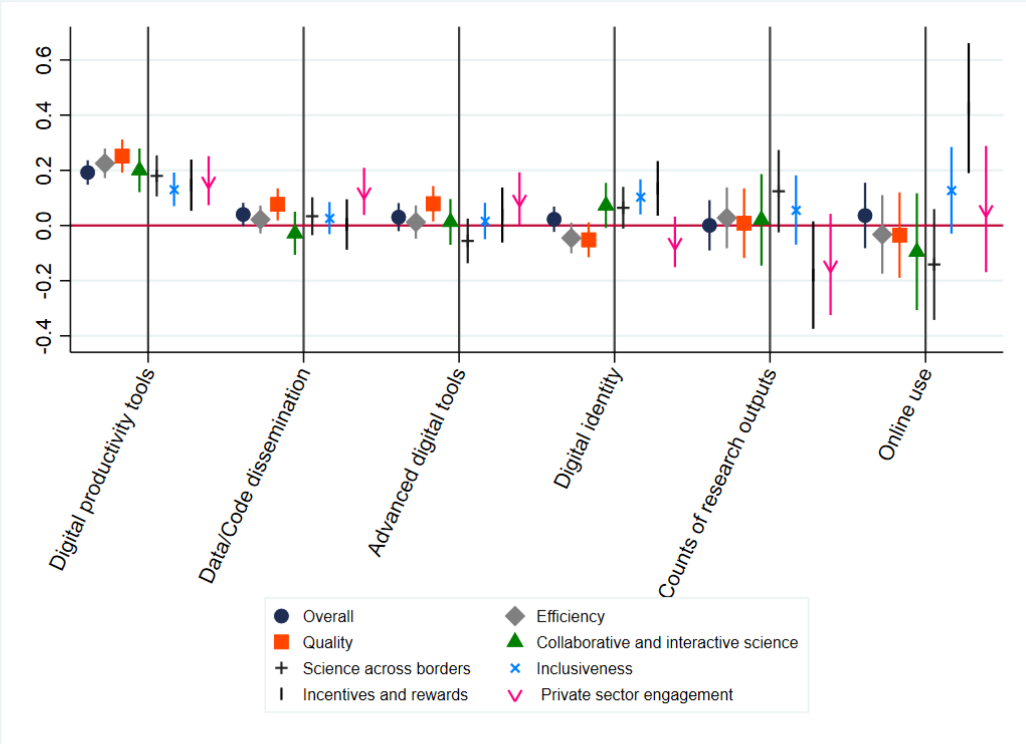
While there is no evidence of a significant difference in attitudes between women and men towards the impact of digitalisation on science, ICT graduates tend to be more positive in relation to the impact of digitalisation on the quality and efficiency of research (Figure 5.14).

Authors making intensive use of digital productivity and collaboration tools are more positive than their peers with regard to all science dimensions and more so in relation to its impact on the efficiency and quality of research. Direct experience of digitalisation appears to make respondents more inclined to have positive views. There is also evidence of a positive attitude among those engaged in advanced digital practices and data sharing with respect to the impact of digitalisation on the quality of research and the role of the private sector in providing digital solutions to assist their work. Authors with high digital identity and communication scores are more likely to have a positive view of the impact of digitalisation on the incentives system and inclusiveness of research work.

Figure 5.14. Authors' attitudes towards digitalisation and their profiles

Least square regression coefficients and confidence intervals





Note: Regression full results are provided in Table D.9 and Table D.10. Survey respondents were asked to rate opposing scenarios regarding the impact of digitalisation on different dimensions of science from 1 (fully agree with negative view) to 10 (fully agree with positive view). Responses by dimension were used as dependent variables in the regressions. *Counts of research outputs* and *online use* refer to the use of research metrics based on the counts of outputs or the online use of research outputs, respectively, in an author's field of research. *ICT graduate* is a dummy variable, which is equal to 1 if an author holds a degree in computer and information sciences, electrical engineering, electronic engineering, or information engineering, and 0 otherwise. PNP stands for private, non-profit sector. BE stands for business enterprise sector. HE stands for higher education sector. Government (GOV) is the baseline of PNP, BE and HE coefficients.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

6. Concluding remarks

This paper has provided an overview of the key findings of the OECD ISSA2 study on the digitalisation of scientific research. The survey results provide a rich, global snapshot picture of the multi-faceted nature of science and digitalisation, providing a baseline for charting its digital transformation and the mechanisms through which it influences scientific research and its impacts on society.

The evidence presented in this document shows that although digital activity is pervasive, the transformation has been uneven across fields and sectors, influenced by factors such as norms, experience, skills and data availability. The potential impacts of digitalisation have been explored, combining different indicators of research impact as well as examining subjective views about the state of play. Overall, scientists appear to be optimist about the possibilities brought about by digitalisation, and especially in relation to the efficiency of research and collaboration across national borders.

The ISSA study also provides a response to the call by OECD member countries and partners at the OECD Ministerial Meeting in Daejeon, Korea for evidence-based responses to the rapid evolution of digital technologies and its impact on STI (OECD, 2015c). By delivering a digitally enabled solution to the human centred measurement challenge identified at the OECD Blue Sky Forum (OECD, 2018; http://oe.cd/blue-sky), this innovation in the way the OECD collects and reports data on science complements the tools currently available for informing international analysis and comparisons.

This paper also showcases the potential of the ISSA approach to collect on timely basis micro data on scientific authors linked to bibliographic information. This results in an unique global set of information on the characteristics, behaviour and preferences of scientists. This evidence adds itself to that already gathered in 2016 when a first pilot of ISSA was carried out. Overall, the ISSA instrument has the potential to become a valuable resource for measuring the impacts of scientific work and investigate how scientific process takes place, complementing information that is already captured in other statistical, administrative and commercial data sources. The ability to collect linked, granular data on scientists, coupled with the possibilities to adapt this instrument based on emerging research and policy needs, represents a major opportunity for policy analysis.

Taking into account the room for further technical improvements, the ISSA approach can provide a basis for distributed data collection within countries, which could also be extended to target a broader population, or the development of a more consolidated survey data infrastructure within the OECD to use for the statistical analysis of major science and research policy questions. Finally, the process leading to the identification and prioritisation of the survey topics can also provide an important opportunity to strengthen communication within the OECD, between the policy and statistical communities working on science issues, as well as between the OECD and the global scientific community. The dissemination of the anonymised microdata for research purposes and the development of additional web functionalities for personal use of the data could in principle support this vision.

References

Bello M. and F. Galindo-Rueda (2020), "The 2018 OECD International Survey of Scientific Authors", *OECD Science, Technology and Industry Working Papers*, OECD Publishing, Paris. https://doi.org/10.1787/18151965

Bloom T, Ganley E, Winker M (2014) Data Access for the Open Access Literature: PLOS's Data Policy. PLoS Biol 12(2): e1001797. https://doi.org/10.1371/journal.pbio.1001797

Bloom, N., C. I. Jones, J.V.Reenen, and M.Webb (2017), "Are Ideas Getting Harder to Find?", NBER Working Paper No. 23782, September 2017, https://www.nber.org/papers/w23782.

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.

Choudhury S, Fishman JR, McGowan ML, Juengst ET. Big data, open science and the brain: lessons learned from genomics. Frontiers in Human Neuroscience. 2014; 8: 239. doi:10.3389/fnhum.2014.00239.

DAMVAD (2014), "Sharing and archiving of publicly funded research data", Report to the Research Council of Norway. Accessed through https://www.usit.uio.no/om/organisasjon/itf/saker/forskningsdata/bakgrunn/sharing-and-archiving-research-data.pdfDigital 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 https://ec.europa.eu/info/sites/info/files/research_and_innovation/open_science_monitor_methodological_note_april_2019.pdf

Elsevier (2019b). Research Futures. Researcher survey results www.elsevier.com/__data/assets/pdf_file/0011/847334/ResearchFutures-Survey-Results-February-2019.pdf

European Commission (2013), Digital science in Horizon 2010, https://ec.europa.eu/digital-single-market/en/news/digital-science-horizon-2020

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

Gannon, D. and D. Reed (2009), Parallelism and the Cloud, in: Hey, T., Tansley, S. & Tolle, K. (eds.)(2009). The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond, Washington: Microsoft Research.

Grice, J. W. (2001). Computing and evaluating factor scores. Psychological Methods, 6(4), 430-450

Hey, T., Tansley, S. & Tolle, K. (eds.)(2009). The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond, Washington: Microsoft Research.

Kaye J, Heeney C, Hawkins N, De Vries, Boddington. Data sharing in genomics—re-shaping scientific practice. Nature Reviews Genetics. 2009; 10(5): 331-335.

Market Research, Wiley (2017): Wiley Open Science Researcher Survey 2016 Infographic. figshare. Journal contribution. https://doi.org/10.6084/m9.figshare.4910714.v1

Meijer, I. , Berghmans, S., Cousijn, H., Tatum, C., Deakin, G., Plume, A., Rushforth, A., Mulligan, A., de Rijcke, S., Tobin, S., Van Leeuwen, T. and Waltman, L. (2017). Open Data: the researcher perspective. 10.17632/bwrnfb4bvh.1.

OECD (2019), Measuring the Digital Transformation: A Roadmap for the Future, OECD Publishing, Paris, https://doi.org/10.1787/9789264311992-en.

OECD (2020), The Digitalisation of Science, Technology and Innovation: Key Developments and Policies, OECD Publishing, Paris, https://doi.org/10.1787/b9e4a2c0-en

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, https://doi.org/10.1787/sti_in_outlook-2018-19-en.

OECD (2015a), 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, https://doi.org/10.1787/9789264239012-en.

OECD (2015b), Making Open Science a Reality, OECD Publishing, Paris.

OECD (2015c), "Daejeon Declaration on Science, Technology, and Innovation Policies for the Global and Digital Age", webpage, www.oecd.org/sti/daejeon-declaration-2015.htm (accessed 1 June 2019).

OECD / SCImago Research Group (CSIC) (2016), Compendium of Bibliometric Science Indicators.

Priem, J. & Hemminger, B. M. (2010). Scientometrics 2.0: Toward new metrics of scholarly impact on the social Web. First Monday, Volume 15, Number 7 - 5 July 2010 https://firstmonday.org/ojs/index.php/fm/article/view/2874/2570

Toronto International Data Release Workshop Authors, Birney E, Hudson TJ, et al. Prepublication data sharing. Nature. 2009;461(7261):168–170. doi:10.1038/461168a

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

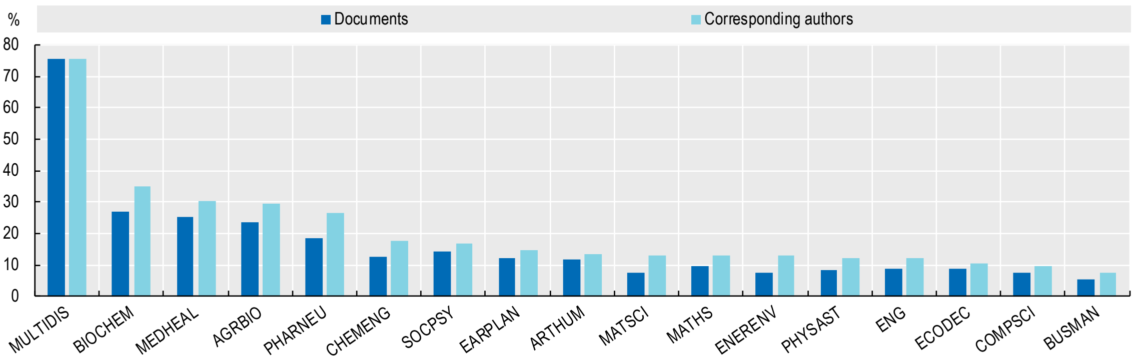
Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology.PloSone,8(7), e67332, p. 15.

Annex A. Benchmark bibliometric indicators on digitalisation

This section presents a number of benchmark digitalisation indicators on open access and AI content based on bibliometric sources for comparison with survey-based indicators presented in the main section of this paper. These bibliometric indicators are used to control for potential response biases collated with digital activity in the analysis. Results are presented for documents and (corresponding) authors.

Figure A.1. Open access journal publishing by science fields, 2017, ISSA2 sampling frame

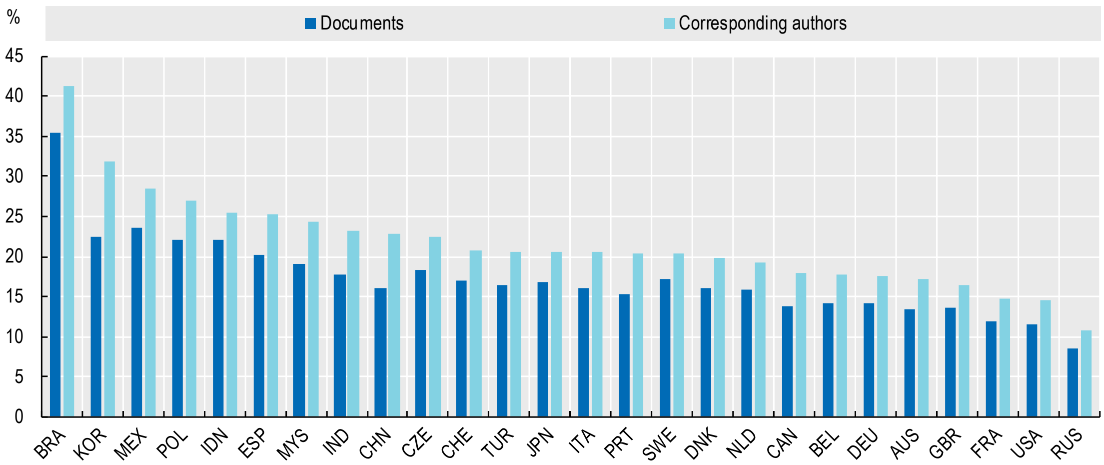
OA-J (gold) documents and corresponding authors with at least one OA publication as percentages of total number of documents and corresponding authors, respectively



Note: Figures relating to corresponding authors are derived from fractional counts.
Source: OECD International Survey of Scientific Authors (ISSA), 2018 and Scopus. http://oe.cd/issa.

Figure A.2. Open access journal publishing, selected economies, 2017, ISSA2 sampling frame

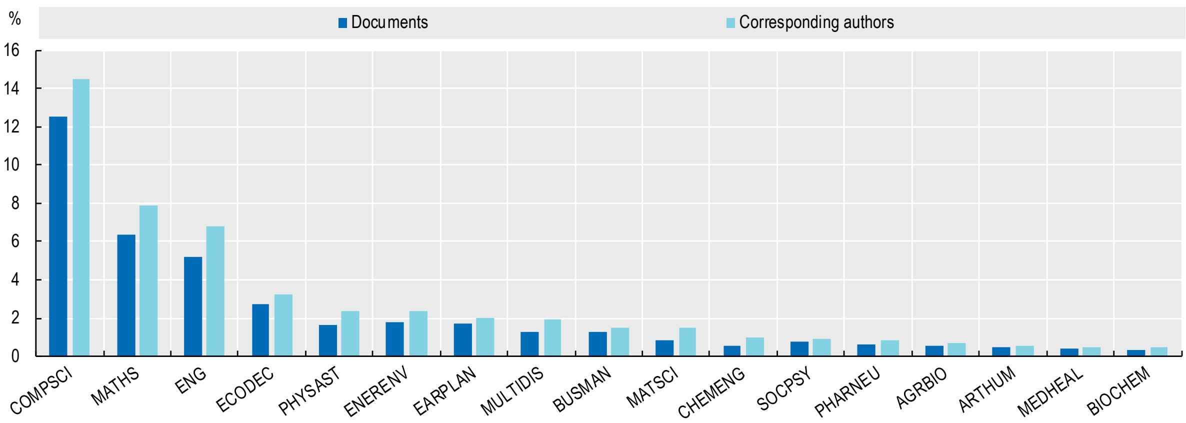
Economies with the largest share of OA documents and corresponding authors with at least one OA publication on the total number of documents and corresponding authors, respectively



Note: Figures relating to corresponding authors are derived from fractional counts. Only economies with more than 7,000 authors are shown in the chart.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure A.3. AI-related scientific document content, by fields, 2017, ISSA2 sampling frame

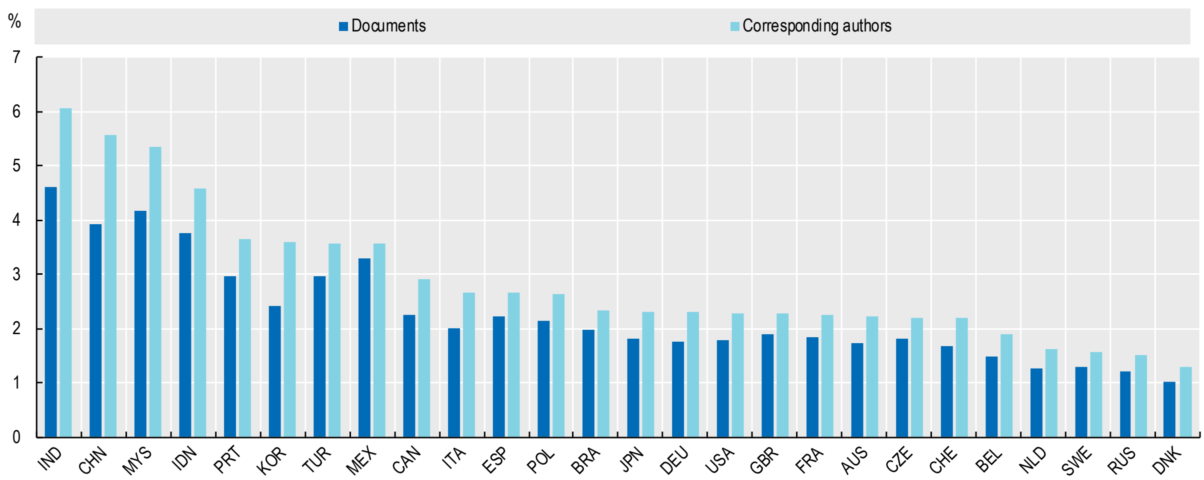
AI-related documents and corresponding authors with at least one AI-related publication as percentages of total number of documents and corresponding authors, respectively



Note: Figures relating to corresponding authors are derived from fractional document counts. 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) and Scopus, 2018. http://oe.cd/issa.

Figure A.4. AI-related scientific document content, selected economies, 2017, ISSA2 sampling frame

Economies with the largest share of AI-related documents and corresponding authors with at least one AI-related publication on total number of documents and corresponding authors, respectively



Note: Figures relating to corresponding authors are derived from fractional counts. 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). Only economies with more than 7,000 authors are shown in the chart.
Source: OECD International Survey of Scientific Authors (ISSA) and Scopus, 2018. http://oe.cd/issa.

Annex B. Factor analysis of digitalisation patterns

Table B.1. Description of the questionnaire-based variables used in the factor analysis

Variable	Question number	Description
Open access to publication	Q6	Open access status of the survey reference publication
Use/development of code	Q8	Use or development of code as part of the work for the survey reference publication
Use/development of data	Q8	Use or development of data as part of the work for the survey reference publication
Data compliance with standards	Q12	Compliance of the data or code resulting from the work for the publication with standards that facilitate combining with other data sources
Information on data/code online	Q12	It is possible for interested users to search online for information about the data or code resulting from the work for the survey reference publication
Metadata	Q12	The data or code resulting from the work for the survey reference publication are accompanied by detailed and comprehensive metadata or explanations
Object identifiers for data/code	Q12	Data or code resulting from the work for the survey reference publication have been assigned unique and permanent digital object identifiers
Free access to data	Q12	Interested users do not have to subscribe or pay a fee to access any of the data or code resulting from the work for the survey reference publication
Standard data access mechanism	Q12	The data resulting from the work for the survey reference publication comply with standards that facilitate combining with other data sources
Usage licence for data/code	Q12	A clear usage licence was applied to the data or code resulting from the work for the survey reference publication
Archive manuscripts	Q14	Use of online platforms or related tools to archive manuscripts for review and publication
Communicate to peers	Q14	Use of online platforms or related tools to communicate research findings to peers
Communicate to the public	Q14	Use of online platforms or related tools to communicate research findings to the broader public
Connect with other researchers	Q14	Use of online platforms or related tools to connect with other researchers
Connect with stakeholders	Q14	Use of online platforms or related tools to connect with other stakeholders
Crowd-funding	Q14	Use of online platforms or related tools for crowd-funding
Data analysis	Q14	Use of online platforms or related tools for data analysis
Data collection	Q14	Use of online platforms or related tools to carry out data collection and processing
Disseminate/archive code	Q14	Use of online platforms or related tools to disseminate or archive code
Disseminate/archive data	Q14	Use of online platforms or related tools to disseminate or archive data
Online research profiles	Q14	Use of online platforms or related tools to develop personal or team profiles, associating research-related activities and outputs
Project management	Q14	Use of online platforms or related tools to manage projects
Review peers' work	Q14	Use of online platforms or related tools to review work undertaken by peers
Search database/code	Q14	Use of online platforms or related tools to search for and retrieve databases or computer codes
Search for funding	Q14	Use of online platforms or related tools to search for funding opportunities and submit applications
Search for research materials	Q14	Use of online platforms or related tools to search for and order materials for research
Share documents	Q14	Use of online platforms or related tools to archive/ share documents other than articles, e.g. presentations
Virtual online meetings	Q14	Use of online platforms or related tools to participate in virtual, online meetings
Autonomous machines	Q16	Use or development of programmable machines that executes tasks autonomously, such as robots

Variable	Question number	Description
Big data	Q16	Use or development of data with size, complexity and heterogeneity features that can be handled with unconventional tools and approaches
Computational methods	Q16	Use or development of computational methods, processes and systems to extract knowledge or insights from structured or unstructured data
Conventional data	Q16	Use or development of data with size, complexity and heterogeneity features that can be handled with conventional tools and approaches
Participative networks	Q16	Use or development of participative networks for securing data from a range of individual actors, for example for crowdsourcing research inputs
Sensors	Q16	Use or development of connected sensors to collect information from environment and systems in an automated fashion
Identifiers for research work	Q21	Use of identifiers to connect an author to his/her research work
Webpages on research work	Q22	Use of individual or specialised webpages or networking sites to provide information on research-related work online

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

Table B.2. Factor variance

Factor	Variance	Difference	Proportion	Cumulative
Digital productivity tools	5.288	0.544	0.147	0.147
Data/code dissemination	4.744	1.093	0.132	0.279
Advanced digital tools/data	3.651	0.644	0.101	0.380
Digital identity and communication	3.007		0.084	0.464

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

Table B.3. Factor score evaluation tests

Indeterminacy/determinacy indices and validity coefficients

	Multiple R	R-squared	Minimum correlation	Validity coefficients
Digital productivity tools	0.872	0.760	0.519	0.872
Data/code dissemination	0.808	0.652	0.304	0.808
Advanced digital tools/data	0.837	0.701	0.403	0.837
Digital identity and communication	0.847	0.718	0.435	0.847

Univocality

Factor scores/ Factor	Digital productivity tools	Data/code dissemination	Advanced digital tools/data	Digital identity and communication
Digital productivity tools	-	-0.029	-0.031	-0.119
Data/code dissemination	-0.027	-	-0.068	-0.019
Advanced digital tools/data	-0.030	-0.070	-	-0.050
Digital identity and communication	-0.116	-0.020	-0.051	-

Factor correlation

Factor/Factor	Digital productivity tools	Data/code dissemination	Advanced digital tools/data	Digital identity and communication
Digital productivity tools	1.000	0.000	0.000	0.000
Data/code dissemination	0.000	1.000	0.000	0.000
Advanced digital tools/data	0.000	0.000	1.000	0.000
Digital identity and communication	0.000	0.000	0.000	1.000

Correlation accuracy test: factor score correlation

Factor scores/Factor scores	Digital productivity tools	Data/code dissemination	Advanced digital tools/data	Digital identity and communication
Digital productivity tools	1.000			
Data/code dissemination	-0.030	1.000		
Advanced digital tools/data	-0.021	-0.087	1.000	
Digital identity and communication	-0.141	-0.030	-0.090	1.000

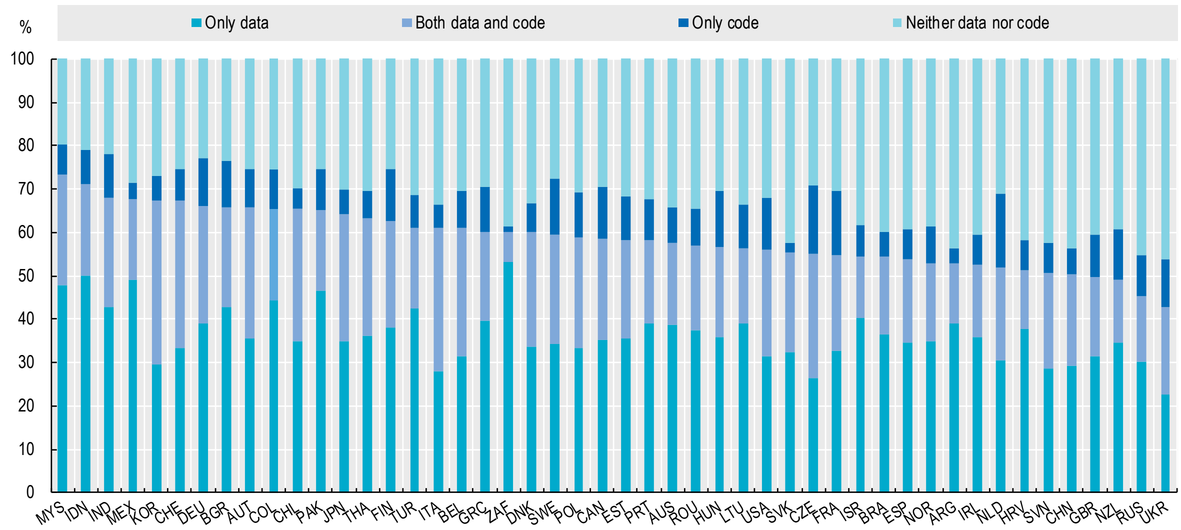
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Annex C. Economy and country-level indicators

This annex provides a series of illustrative indicators derived from the ISSA2 results at the economy or country level. These should be used with great caution as no standard errors are provided and differences may not be statistically significant. Furthermore, their interpretation should be assisted by understanding of the population (scientific corresponding authors) and sample obtained, as detailed in the accompanying technical paper (Bello and Galindo-Rueda, 2020).

Figure C.1. Scientific production resulting in new data or code, by country or economy of residence

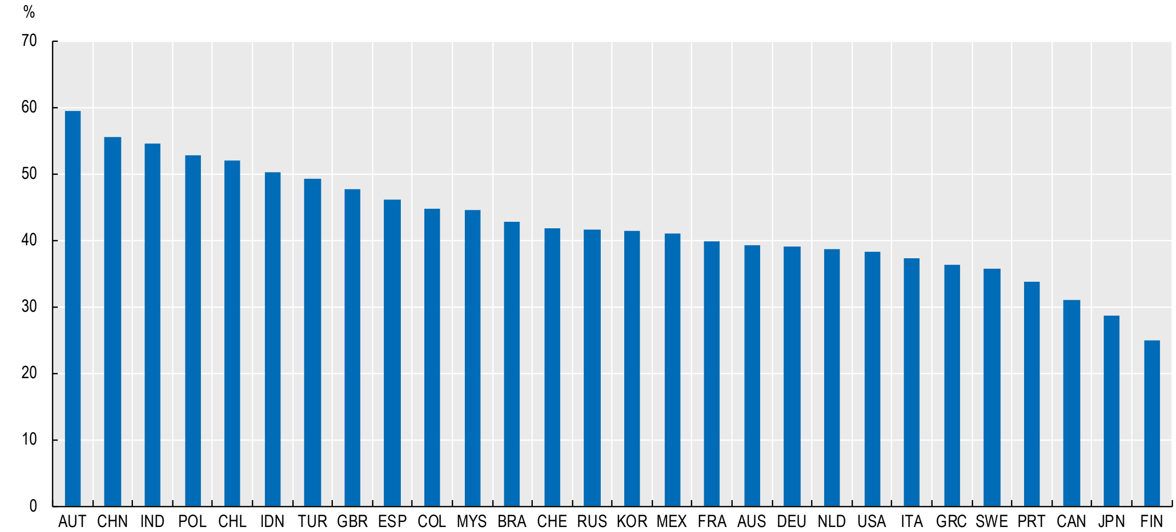
Percentage of respondents with outputs in each category



Note: Weighted estimates based on sampling weights adjusted for nonresponse. Only countries/economies with more than 70 observations are shown in the chart.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure C.2. Absence of fee payment requirements to access to data or code

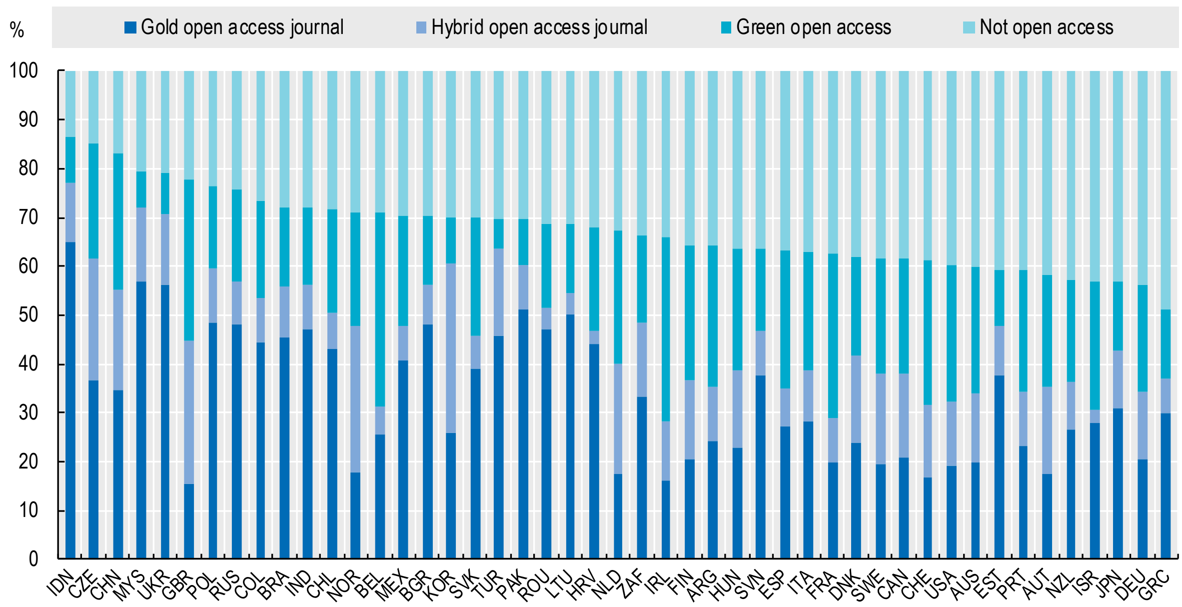
As a percentage of authors with data or code outputs, by country/economy of residence



Note: Weighted estimates based on sampling weights adjusted for nonresponse. Only countries/economies with more than 70 observations are shown in the chart. "Data or code" refer to the data or code developed as part of the work undertaken for the reference publication.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure C.3. Open access to scientific publications, by an author's country or economy of residence

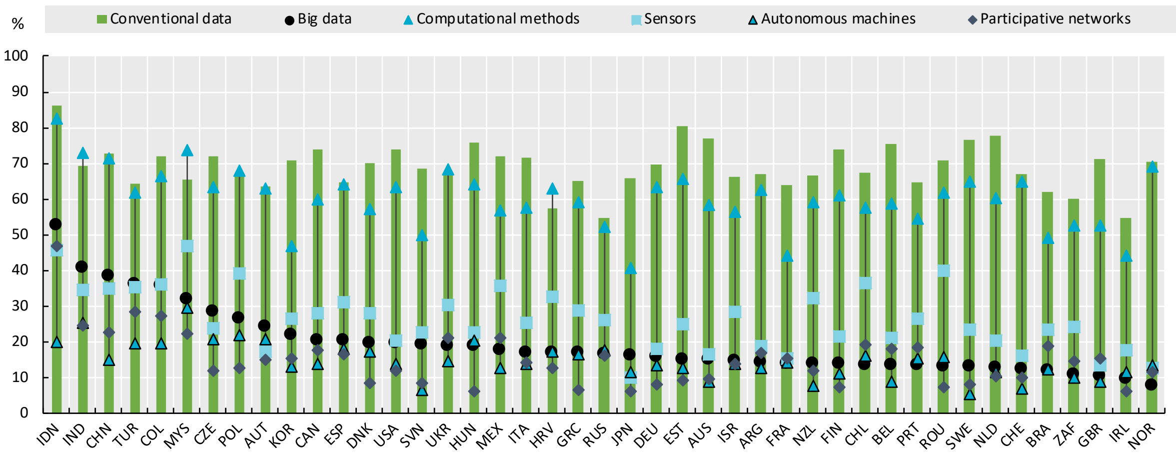
Percentage of authors



Note: Weighted estimates based on sampling weights adjusted for nonresponse. Only countries/economies with more than 70 observations are shown in the chart.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure C.4. Use and/or development of advanced digital practices, by country or economy of residence

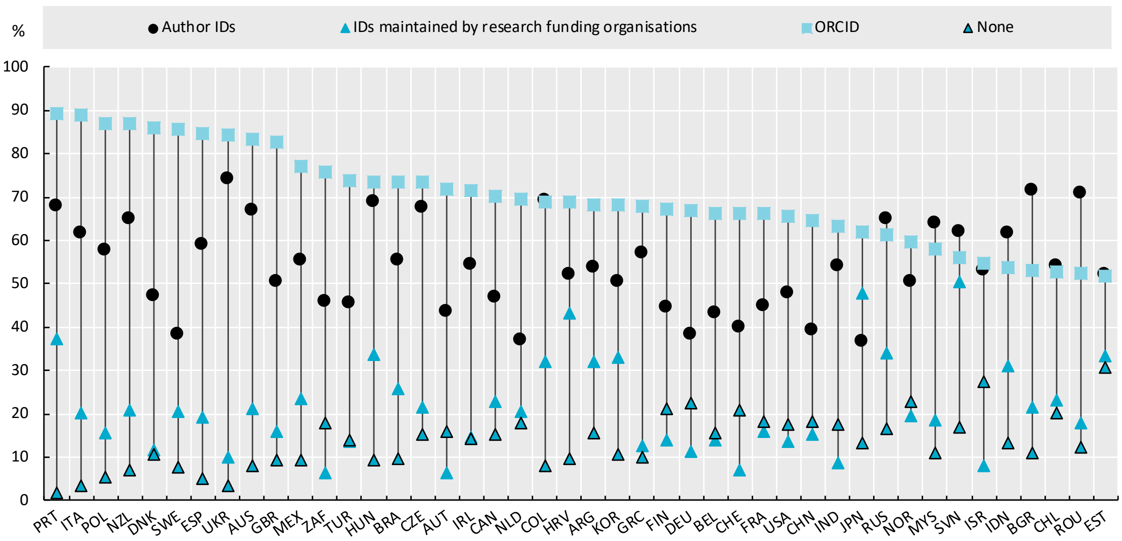
Percentage of authors using or developing a given tool or method



Note: Weighted estimates based on sampling weights adjusted for nonresponse. Only countries/economies with more than 70 observations are shown. Based on the wording adopted in the survey questionnaire, *big data* refer to data with size, complexity and heterogeneity features that can only be handled with unconventional tools and approaches; *computational methods* include all computational methods, processes and systems to extract knowledge or insights from structured or unstructured data; *sensors* include all connected sensors to collect information from environment and systems in an automated fashion; *autonomous machines* refer to programmable machines that execute tasks autonomously, such as robots.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure C.5. Use of identifiers (IDs) by scientific authors to track his/her research work, by country or economy of residence

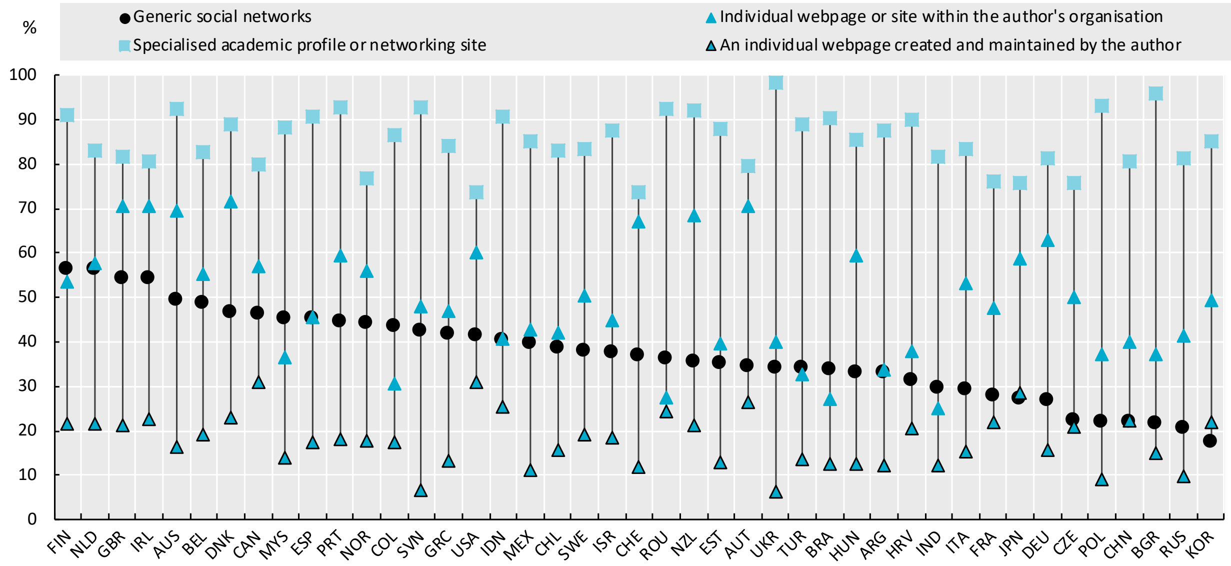
Percentage of authors that use a given identifier



Note: Weighted estimates based on sampling weights adjusted for nonresponse. Only countries/economies with more than 70 observations are shown.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure C.6. Channels used for disseminating information about research, by country or economy of residence

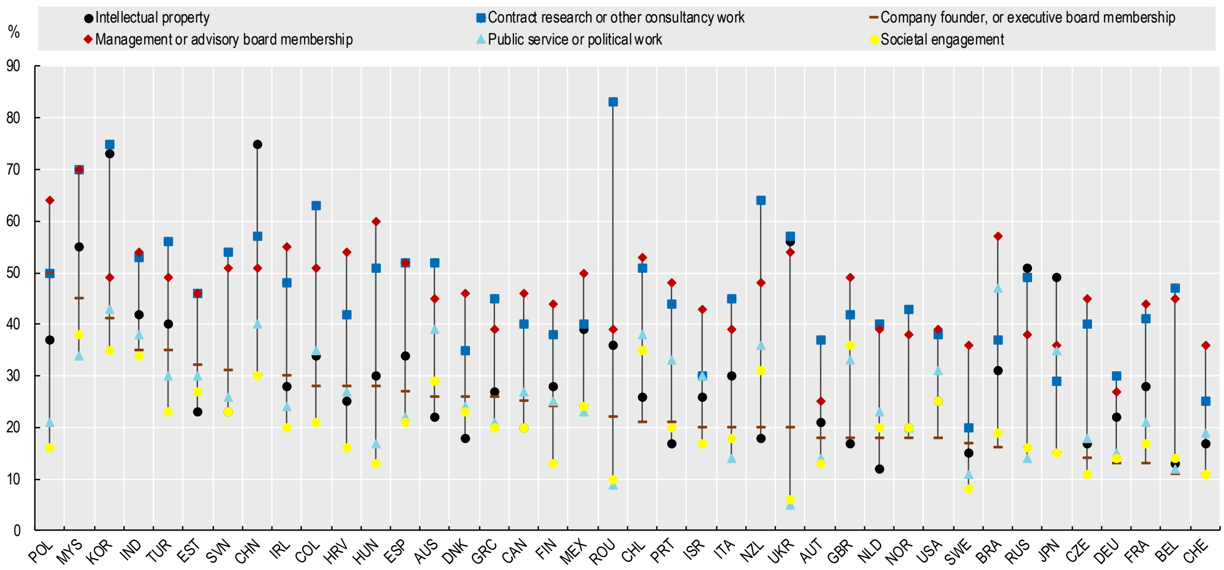
Percentage of authors using a given channel



Note: Weighted estimates based on sampling weights adjusted for nonresponse. Only countries/economies with more than 70 observations are shown.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure C.7. Engagement in research-related activities by country or economy of residence

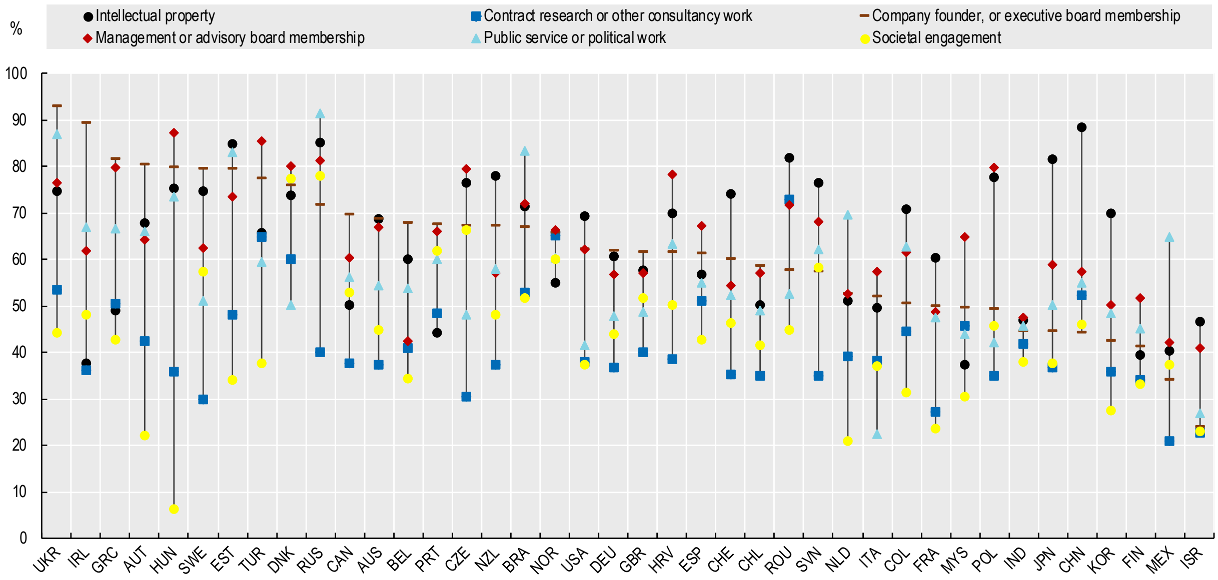
Percentage of authors engaged in a given activity



Note: Weighted estimates based on sampling weights adjusted for nonresponse. Only countries/economies with more than 70 observations are shown.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure C.8. Online availability of information on research-related activities by country or economy of residence

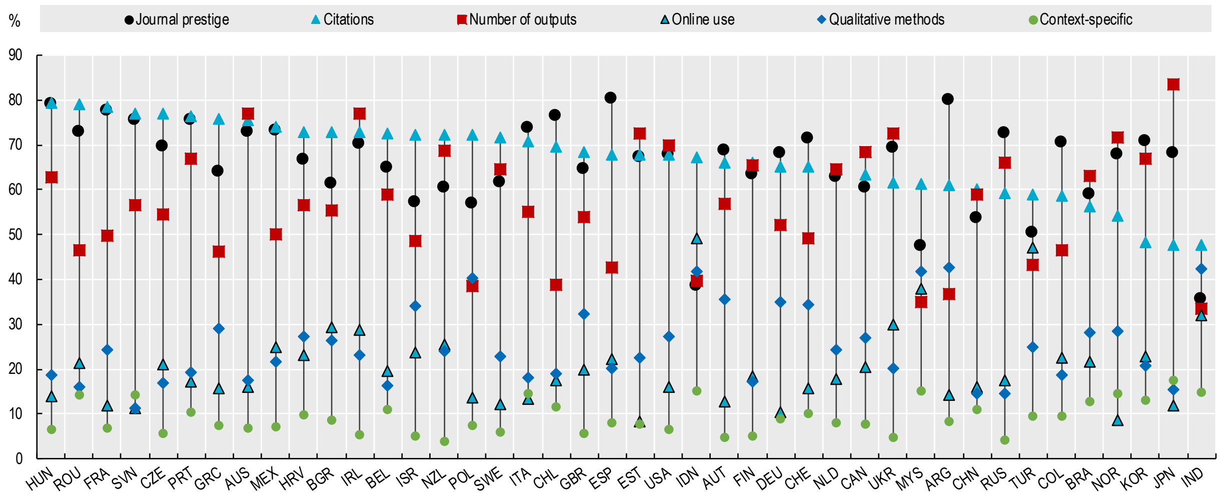
As a percentage of authors engaged in a given activity



Note: Weighted estimates based on sampling weights adjusted for nonresponse. Only countries/economies with more than 70 observations are shown.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure C.9. Use of research metrics, by type and country or economy of residence

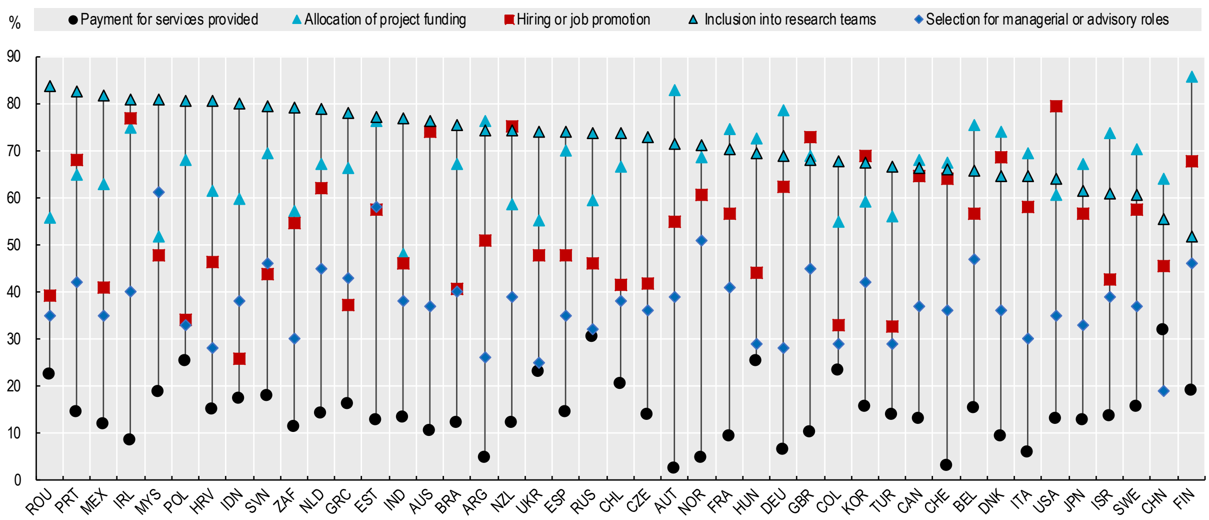
Percentage of authors reporting the use of a given type of research metrics



Note: Weighted estimates based on sampling weights adjusted for nonresponse. Only countries/economies with more than 70 observations are shown.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure C.10. Decisions affected by the use of quantitative research indicators, by country or economy of affiliation

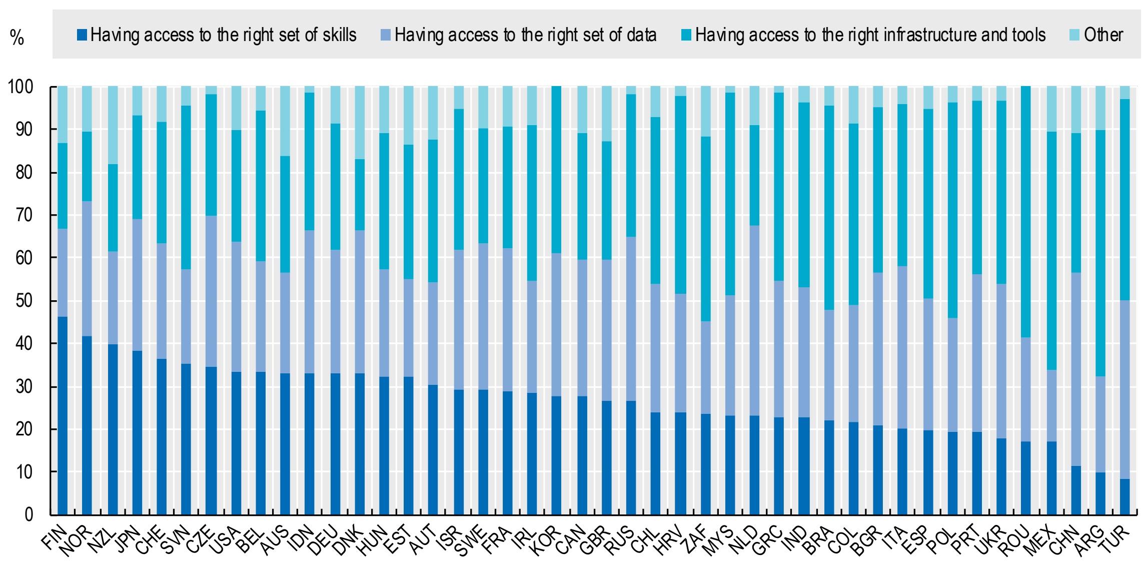
Percentage of authors indicating different channels of indicator-based influence on their careers



Note: Weighted estimates based on sampling weights adjusted for nonresponse. Only countries/economies with more than 70 observations are shown.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure C.11. Challenges faced by scientific authors in their research work

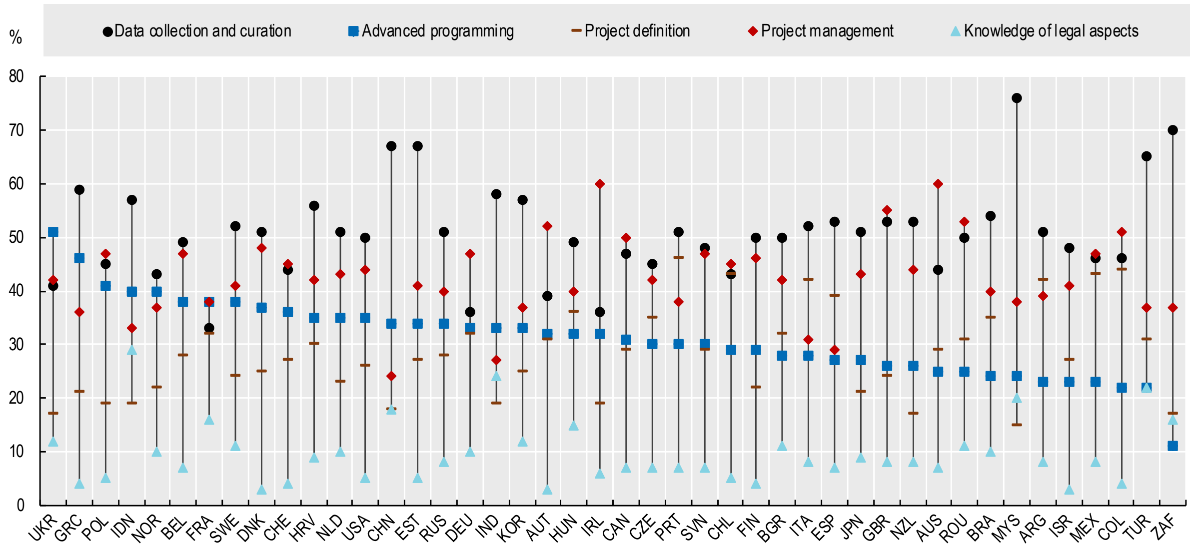
Percentage of authors



Note: Weighted estimates based on sampling weights adjusted for nonresponse. Only countries/economies with more than 70 observations are shown.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure C.12. Most important skills for scientific authors' research work

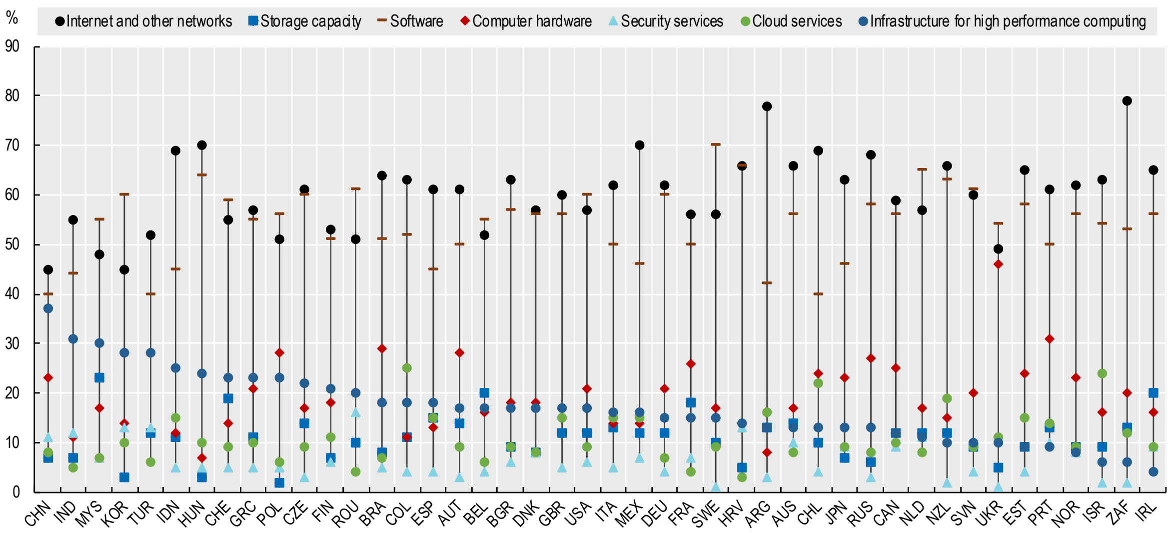
Percentage of authors deeming given types of skills as important



Note: Weighted estimates based on sampling weights adjusted for nonresponse. Only countries/economies with more than 70 observations are shown.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Figure C.13. Most important infrastructure for scientific authors' research work

Percentage of authors deeming a given type of infrastructure as important



Note: Weighted estimates based on sampling weights adjusted for nonresponse. Only countries/economies with more than 70 observations are shown.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Annex D. Detailed regression analysis results

Table A D.1. Links between digitalisation patterns and author profiles

Least square regression estimates

	Digital productivity tools			Data/Code dissemination			Advanced digital tools/data			Digital identity and communication		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Female	0.014	0.018	0.014	-0.063*	-0.062*	-0.071*	-0.132***	-0.105***	-0.110***	0.150***	0.157***	0.152***
	(0.620)	(0.530)	(0.636)	(0.023)	(0.032)	(0.014)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age	-0.007***	-0.009***	-0.008***	-0.002	-0.002	-0.000	-0.007***	-0.005***	-0.004***	-0.004***	-0.005***	-0.005***
	(0.000)	(0.000)	(0.000)	(0.056)	(0.084)	(0.895)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
PhD	0.229***	0.142***	0.144***	0.077*	0.093*	0.107**	-0.137***	-0.056	-0.054	0.282***	0.189***	0.172***
	(0.000)	(0.001)	(0.001)	(0.036)	(0.022)	(0.010)	(0.001)	(0.185)	(0.205)	(0.000)	(0.000)	(0.000)
Born in another country	0.093**	0.096**	0.092**	0.067*	0.065*	0.051	0.002	0.014	0.006	0.124***	0.122***	0.112***
	(0.002)	(0.003)	(0.004)	(0.033)	(0.046)	(0.118)	(0.932)	(0.618)	(0.838)	(0.000)	(0.000)	(0.000)
ICT graduate		0.069	0.034		0.136**	0.122*		0.396***	0.349***		-0.325***	-0.315***
		(0.177)	(0.518)		(0.005)	(0.016)		(0.000)	(0.000)		(0.000)	(0.000)
Baseline = GOV												
PNP		-0.099	-0.112		-0.035	-0.029		0.014	0.040		0.116	0.124*
		(0.149)	(0.103)		(0.598)	(0.661)		(0.830)	(0.552)		(0.059)	(0.046)
BE		-0.021	-0.011		-0.132**	-0.137**		0.115*	0.106*		-0.021	-0.020
		(0.671)	(0.831)		(0.006)	(0.004)		(0.030)	(0.046)		(0.652)	(0.680)
HE		0.080*	0.087**		-0.045	-0.039		-0.107***	-0.103***		0.114***	0.115***
		(0.011)	(0.006)		(0.167)	(0.241)		(0.000)	(0.001)		(0.000)	(0.000)
Number of publications (1996-2017, in log)		0.054***	0.037***		-0.000	-0.013		-0.019	-0.032**		0.033***	0.038***
		(0.000)	(0.001)		(0.973)	(0.248)		(0.059)	(0.003)		(0.001)	(0.000)

	Digital productivity tools			Data/Code dissemination			Advanced digital tools/data			Digital identity and communication		
Average of field-normalised citations per document (1996-2017)			0.034***			0.031**			-0.000			0.005
			(0.001)			(0.004)			(0.964)			(0.556)
Collaboration (1996-2017)			0.193***			0.323***			0.408***			-0.106
			(0.000)			(0.000)			(0.000)			(0.056)
Share of open access publications (1996-2017)			-0.064			0.403***			0.092			0.184**
			(0.277)			(0.000)			(0.100)			(0.001)
AI-related reference publication			-0.000			0.136			0.503***			-0.109
			(0.999)			(0.098)			(0.000)			(0.284)
Constant	0.352**	0.427***	0.216	0.685***	0.673***	0.083	1.175***	1.125***	0.626***	1.044***	1.078***	1.112***
	(0.003)	(0.000)	(0.116)	(0.000)	(0.000)	(0.570)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	6 855	6 389	6 301	6 855	6 389	6 301	6 855	6 389	6 301	6 855	6 389	6 301
R-squared	0.078	0.086	0.090	0.078	0.082	0.098	0.123	0.137	0.149	0.139	0.152	0.153

Note: P-values in parentheses. Country- and science field- fixed effects included in all regressions. * refers to p<0.05; ** refers to p<0.01; *** refers to p<0.001. *ICT graduate* is a dummy variable, which is equal to 1 if an author holds a degree in computer and information sciences, electrical engineering, electronic engineering, or information engineering, and 0 otherwise. C*ollaboration* is measured by the proportion of an author's documents involving co-authorship and indexed in Scopus in the period 1996-2017. *Share of open access publications* refers to the number of an author's open access-publications indexed in Scopus on the author's total number of publications. *AI-related reference publication* is a dummy variable, which is equal to one if the reference 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 D.2. Links between digitalisation patterns and author research methods

Least square regression estimates

	Digital productivity tools			Data/Code dissemination			Advanced digital tools/data			Digital identity and communication		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Female	0.043	0.043	0.041	-0.038	-0.040	-0.047	-0.085***	-0.065*	-0.070**	0.123***	0.132***	0.130***
	(0.101)	(0.118)	(0.142)	(0.170)	(0.167)	(0.104)	(0.001)	(0.013)	(0.007)	(0.000)	(0.000)	(0.000)
Age	-0.006***	-0.008***	-0.007***	-0.001	-0.001	0.001	-0.006***	-0.004***	-0.003*	-0.004***	-0.006***	-0.006***
	(0.000)	(0.000)	(0.000)	(0.396)	(0.562)	(0.551)	(0.000)	(0.001)	(0.015)	(0.000)	(0.000)	(0.000)
PhD	0.201***	0.126**	0.128**	0.064	0.086*	0.097*	-0.171***	-0.079	-0.078	0.268***	0.182***	0.164***
	(0.000)	(0.001)	(0.002)	(0.078)	(0.035)	(0.019)	(0.000)	(0.054)	(0.058)	(0.000)	(0.000)	(0.000)
Born in another country	0.049	0.054	0.051	0.060	0.059	0.044	-0.023	-0.009	-0.017	0.119***	0.119***	0.110***
	(0.094)	(0.078)	(0.094)	(0.057)	(0.071)	(0.173)	(0.401)	(0.739)	(0.555)	(0.000)	(0.000)	(0.000)
Gathering and curating information	0.304***	0.321***	0.322***	0.079**	0.071**	0.071**	0.164***	0.148***	0.146***	0.047	0.056*	0.041
	(0.000)	(0.000)	(0.000)	(0.002)	(0.009)	(0.009)	(0.000)	(0.000)	(0.000)	(0.058)	(0.031)	(0.108)
Testing hypotheses in experimental settings	0.242***	0.222***	0.213***	0.133***	0.141***	0.123***	0.283***	0.288***	0.268***	0.077**	0.068**	0.080**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.010)	(0.003)
Testing hypotheses in non-experimental settings	0.264***	0.266***	0.262***	0.070**	0.061*	0.054*	0.121***	0.136***	0.128***	0.099***	0.081**	0.070**
	(0.000)	(0.000)	(0.000)	(0.005)	(0.021)	(0.042)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.007)
Formulating theories	0.225***	0.220***	0.224***	-0.037	-0.035	-0.021	0.092***	0.093***	0.094***	0.109***	0.106***	0.109***
	(0.000)	(0.000)	(0.000)	(0.145)	(0.186)	(0.426)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Using computational modelling	0.296***	0.289***	0.286***	0.284***	0.293***	0.283***	0.421***	0.403***	0.381***	-0.323***	-0.318***	-0.310***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ICT graduate		0.018	-0.008		0.085	0.082		0.318***	0.287***		-0.280***	-0.278***
		(0.723)	(0.872)		(0.079)	(0.109)		(0.000)	(0.000)		(0.000)	(0.000)
Baseline = GOV												
PNP		-0.065	-0.081		-0.019	-0.018		0.046	0.064		0.113	0.125*
		(0.308)	(0.201)		(0.781)	(0.782)		(0.476)	(0.320)		(0.063)	(0.041)
BE		-0.018	-0.008		-0.143**	-0.149**		0.107*	0.098		-0.016	-0.013
		(0.709)	(0.878)		(0.003)	(0.002)		(0.037)	(0.059)		(0.729)	(0.792)
HE		0.071*	0.077*		-0.046	-0.041		-0.118***	-0.115***		0.110***	0.110***
		(0.018)	(0.010)		(0.157)	(0.208)		(0.000)	(0.000)		(0.000)	(0.000)

	Digital productivity tools			Data/Code dissemination			Advanced digital tools/data			Digital identity and communication		
Number of publications (1996-2017, in log)		0.050***	0.037***		-0.008	-0.016		-0.031**	-0.038***		0.036***	0.039***
		(0.000)	(0.001)		(0.450)	(0.149)		(0.002)	(0.000)		(0.000)	(0.000)
Average of field-normalised citations per document (1996-2017)			0.034***			0.031**			-0.000			0.005
			(0.000)			(0.003)			(0.979)			(0.586)
Collaboration (1996-2017)			0.099			0.238***			0.280***			-0.080
			(0.056)			(0.000)			(0.000)			(0.152)
Share of open access publications (1996-2017)			-0.100			0.397***			0.083			0.190***
			(0.081)			(0.000)			(0.126)			(0.001)
AI-related reference publication			-0.080			0.065			0.404***			-0.036
			(0.391)			(0.418)			(0.000)			(0.721)
Constant	-0.451***	-0.365**	-0.438**	0.198	0.169	-0.286	0.358**	0.286*	-0.013	1.174***	1.235***	1.230***
	(0.000)	(0.004)	(0.001)	(0.133)	(0.215)	(0.053)	(0.002)	(0.017)	(0.918)	(0.000)	(0.000)	(0.000)
Observations	6 834	6 370	6 282	6 834	6 370	6 282	6 834	6 370	6 282	6 834	6 370	6 282
R-squared	0.171	0.177	0.180	0.101	0.105	0.118	0.193	0.204	0.210	0.163	0.175	0.175

Note: P-values in parentheses. Country- and science field- fixed effects included in all regressions. * refers to p<0.05; ** refers to p<0.01; *** refers to p<0.001. *ICT graduate* is a dummy variable, which is equal to 1 if an author holds a degree in computer and information sciences, electrical engineering, electronic engineering, or information engineering, and 0 otherwise. C*ollaboration* is measured by the proportion of an author's documents involving co-authorship and indexed in Scopus in the period 1996-2017. *Share of* o*pen access publications* refers to the number of an author's open access-publications indexes in Scopus on the author's total number of publications. *AI-related reference publication* is a dummy variable, which is equal to one if the reference 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 D.3. Drivers of authors' perceived challenges in their research work

Least square regression estimates

	Access to the right set of skills		Access to the right set of data		Access to the right infrastructure and tools	
	(1)	(2)	(1)	(2)	(1)	(2)
Female	-0.003	-0.004	-0.028*	-0.034*	0.024	0.030*
	(0.827)	(0.760)	(0.039)	(0.017)	(0.068)	(0.032)
Age	-0.001*	-0.001*	-0.000	0.000	0.000	0.000
	(0.010)	(0.018)	(0.973)	(0.762)	(0.838)	(0.843)
PhD	0.023	0.017	-0.049**	-0.040	0.027	0.024
	(0.179)	(0.382)	(0.010)	(0.059)	(0.138)	(0.235)
Born in another country	-0.024	-0.028	0.010	0.005	0.010	0.019
	(0.104)	(0.073)	(0.503)	(0.768)	(0.494)	(0.201)
Measures of digital intensity						
Digital productivity tools	-0.000	-0.002	0.004	0.007	0.002	-0.000
	(0.932)	(0.712)	(0.512)	(0.297)	(0.780)	(0.975)
Data/Code dissemination	0.006	0.003	-0.000	-0.001	0.008	0.011
	(0.277)	(0.576)	(0.979)	(0.912)	(0.171)	(0.069)
Advanced digital tools/data	-0.014*	-0.015*	0.026***	0.025***	0.000	0.000
	(0.019)	(0.018)	(0.000)	(0.000)	(0.971)	(0.965)
Digital identity and communication	0.007	0.014*	-0.022***	-0.023***	0.015*	0.011
	(0.227)	(0.029)	(0.001)	(0.000)	(0.012)	(0.080)
ICT graduate		0.067**		-0.027		-0.034
		(0.007)		(0.280)		(0.168)
PNP, *Baseline = GOV*		-0.045		0.028		-0.018
		(0.125)		(0.368)		(0.545)
BE, *Baseline = GOV*		-0.054*		0.043		0.018
		(0.014)		(0.077)		(0.452)
HE, *Baseline = GOV*		0.005		0.007		-0.023
		(0.737)		(0.659)		(0.131)
Number of publications (1996-2017, in log)		0.003		-0.017**		0.014*
		(0.607)		(0.001)		(0.011)
Average of field-normalised citations per document (1996-2017)		0.012*		0.000		-0.011**
		(0.033)		(0.957)		(0.009)
Collaboration (1996-2017)		0.013		0.002		0.016
		(0.619)		(0.937)		(0.540)
Share of open access publications (1996-2017)		0.006		-0.011		0.004
		(0.830)		(0.699)		(0.891)
Observations	6 484	5 964	6 484	5 964	6 484	5 964
R-squared	0.054	0.063	0.089	0.100	0.096	0.107

Note: P-values in parentheses. Country- and science field- fixed effects included in the regressions. * refers to p<0.05; ** refers to p<0.01; *** refers to p<0.001. *ICT graduate* is a dummy variable, which is equal to 1 if an author holds a degree in computer and information sciences, electrical engineering, electronic engineering, or information engineering, and 0 otherwise. C*ollaboration* is measured by the proportion of an author's documents involving co-authorship and indexed in Scopus in the period 1996-2017. *Share of* open *access publications* refers to the number of an author's open access-publications indexed in Scopus on the author's total number of publications.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Table D.4. Factors explaining respondent's reporting of most important research skills

Least square regression estimates

	Data collection and curation		Advanced programming		Project definition		Project management		Knowledge of legal aspects	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Female	0.037**	0.026	-0.052***	-0.052***	-0.004	-0.012	0.034*	0.047**	0.027**	0.022*
	(0.009)	(0.083)	(0.000)	(0.000)	(0.786)	(0.387)	(0.015)	(0.001)	(0.001)	(0.011)
Age	-0.000	0.000	-0.004***	-0.004***	0.001*	0.002**	-0.001**	-0.002**	0.001**	0.001**
	(0.549)	(0.764)	(0.000)	(0.000)	(0.041)	(0.005)	(0.007)	(0.003)	(0.002)	(0.007)
PhD	-0.059**	-0.056**	-0.010	-0.009	0.030	0.036	0.032	0.002	-0.027*	-0.009
	(0.002)	(0.009)	(0.558)	(0.639)	(0.087)	(0.068)	(0.093)	(0.933)	(0.021)	(0.460)
Born in another country	-0.025	-0.026	0.001	0.002	0.024	0.025	-0.032*	-0.031	-0.001	-0.005
	(0.104)	(0.114)	(0.933)	(0.871)	(0.096)	(0.098)	(0.037)	(0.051)	(0.878)	(0.489)
Measures of digital intensity										
Digital productivity tools	0.014*	0.020**	0.038***	0.038***	-0.001	-0.004	0.042***	0.036***	0.001	0.002
	(0.026)	(0.003)	(0.000)	(0.000)	(0.808)	(0.467)	(0.000)	(0.000)	(0.710)	(0.566)
Data/Code dissemination	0.019**	0.021**	0.061***	0.056***	-0.017**	-0.017**	-0.021***	-0.020**	0.009*	0.008*
	(0.002)	(0.001)	(0.000)	(0.000)	(0.004)	(0.006)	(0.001)	(0.002)	(0.015)	(0.031)
Advanced digital tools/data	0.019**	0.024***	0.075***	0.071***	-0.015*	-0.013*	-0.012	-0.011	0.015***	0.010*
	(0.003)	(0.001)	(0.000)	(0.000)	(0.014)	(0.038)	(0.061)	(0.119)	(0.000)	(0.030)
Digital identity and communication	-0.010	-0.008	-0.073***	-0.074***	0.020***	0.021***	0.083***	0.083***	0.012**	0.013***
	(0.122)	(0.270)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
ICT graduate		-0.084***		0.100***		0.017		-0.011		0.025
		(0.001)		(0.000)		(0.470)		(0.665)		(0.089)
Baseline = GOV										
PNP		-0.076*		0.020		0.055		0.000		-0.005
		(0.021)		(0.500)		(0.078)		(0.998)		(0.794)
BE		-0.029		-0.025		0.025		0.009		0.013
		(0.256)		(0.265)		(0.270)		(0.706)		(0.400)
HE		-0.010		0.008		0.022		-0.010		-0.017
		(0.544)		(0.578)		(0.130)		(0.516)		(0.056)
Number of publications (1996-2017, in log)		-0.022***		0.007		-0.005		0.020***		-0.002

	Data collection and curation		Advanced programming		Project definition		Project management		Knowledge of legal aspects	
		(0.000)		(0.151)		(0.393)		(0.000)		(0.566)
Average of field-normalised citations per document (1996-2017)		-0.003		0.005		0.005		0.006		-0.007***
		(0.589)		(0.274)		(0.286)		(0.166)		(0.000)
Collaboration (1996-2017)		-0.017		0.058*		-0.028		0.107***		-0.021
		(0.561)		(0.015)		(0.294)		(0.000)		(0.221)
Share of open access publications (1996-2017)		-0.002		0.023		0.026		-0.041		0.022
		(0.949)		(0.351)		(0.339)		(0.153)		(0.250)
Constant	0.090	0.109	1.141***	1.106***	-0.128*	-0.156*	0.026	-0.079	0.936***	0.945***
	(0.129)	(0.119)	(0.000)	(0.000)	(0.015)	(0.013)	(0.656)	(0.245)	(0.000)	(0.000)
Observations	6 663	6 128	6 663	6 128	6 663	6 128	6 663	6 128	6 663	6 128
R-squared	0.087	0.094	0.184	0.190	0.052	0.056	0.092	0.098	0.068	0.074

Note: P-values in parentheses. Country- and science field- fixed effects included in the regressions. * refers to p<0.05; ** refers to p<0.01; *** refers to p<0.001. *ICT graduate* is a dummy variable, which is equal to 1 if an author holds a degree in computer and information sciences, electrical engineering, electronic engineering, or information engineering, and 0 otherwise. C*ollaboration* is measured by the proportion of an author's documents involving co-authorship and indexed in Scopus in the period 1996-2017. *Share of open access publications* refers to the number of an author's open access-publications indexes in Scopus on the author's total number of publications.

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

Table D.5. Factors explaining reporting of most important research infrastructure

Least square regression estimates

	Internet and network connectivity		Storage capacity		Software		Computer hardware		Security services		Cloud services		Infrastructure for high performance computing	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Female	0.012	0.012	0.038***	0.037***	-0.070***	-0.065***	-0.020	-0.018	0.034***	0.027***	0.035***	0.032**	0.000	-0.002
	(0.378)	(0.413)	(0.000)	(0.000)	(0.000)	(0.000)	(0.073)	(0.120)	(0.000)	(0.000)	(0.000)	(0.001)	(0.973)	(0.816)
Age	0.002***	0.002***	-0.001	-0.001	0.000	-0.000	-0.000	-0.000	0.000	0.001*	-0.002***	-0.002***	-0.003***	-0.003***
	(0.000)	(0.000)	(0.057)	(0.064)	(0.983)	(0.708)	(0.623)	(0.681)	(0.177)	(0.044)	(0.000)	(0.000)	(0.000)	(0.000)
PhD	-0.013	-0.027	0.006	0.004	-0.003	-0.025	0.017	0.024	-0.008	-0.001	-0.006	0.005	0.014	0.018
	(0.479)	(0.193)	(0.583)	(0.785)	(0.895)	(0.243)	(0.260)	(0.139)	(0.388)	(0.940)	(0.614)	(0.728)	(0.326)	(0.256)
Born in another country	0.005	0.009	-0.003	-0.003	-0.005	-0.006	-0.016	-0.015	-0.009	-0.009	-0.003	-0.003	0.021	0.021
	(0.738)	(0.564)	(0.779)	(0.754)	(0.766)	(0.703)	(0.208)	(0.242)	(0.170)	(0.159)	(0.750)	(0.783)	(0.051)	(0.066)
Digital intensity measures														
Digital productivity tools	0.003	0.005	0.011**	0.011**	0.008	0.007	-0.001	-0.001	0.007*	0.006*	0.014***	0.016***	0.027***	0.023***
	(0.614)	(0.461)	(0.005)	(0.009)	(0.211)	(0.278)	(0.842)	(0.924)	(0.011)	(0.029)	(0.000)	(0.000)	(0.000)	(0.000)
Data/Code dissemination	-0.032***	-0.036***	0.015***	0.015***	0.006	0.005	-0.002	-0.001	0.010**	0.009**	-0.010*	-0.009*	0.050***	0.051***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.330)	(0.408)	(0.723)	(0.796)	(0.002)	(0.005)	(0.012)	(0.026)	(0.000)	(0.000)
Advanced digital tools/data	-0.045***	-0.046***	0.013**	0.013**	0.010	0.012	0.006	0.008	0.008*	0.007	-0.007	-0.009*	0.064***	0.065***
	(0.000)	(0.000)	(0.001)	(0.002)	(0.112)	(0.087)	(0.228)	(0.140)	(0.023)	(0.071)	(0.092)	(0.030)	(0.000)	(0.000)
Digital identity and communication	0.029***	0.027***	0.008	0.004	-0.024***	-0.027***	-0.020***	-0.016**	0.009**	0.009**	0.027***	0.028***	-0.025***	-0.024***
	(0.000)	(0.000)	(0.061)	(0.306)	(0.000)	(0.000)	(0.000)	(0.004)	(0.004)	(0.003)	(0.000)	(0.000)	(0.000)	(0.000)
ICT graduate		-0.006		-0.027		-0.033		0.037		-0.003		0.016		0.009
		(0.807)		(0.054)		(0.209)		(0.096)		(0.800)		(0.362)		(0.680)
PNP		-0.026		0.030		-0.004		0.030		0.012		0.003		-0.030
		(0.413)		(0.185)		(0.909)		(0.263)		(0.493)		(0.886)		(0.191)

	Internet and network connectivity		Storage capacity		Software		Computer hardware		Security services		Cloud services		Infrastructure for high performance computing	
BE		-0.038		-0.012		-0.054*		0.023		0.017		0.031*		-0.011
		(0.122)		(0.412)		(0.031)		(0.278)		(0.180)		(0.045)		(0.545)
HE		-0.006		-0.008		0.022		-0.019		-0.003		0.008		-0.002
		(0.704)		(0.429)		(0.172)		(0.153)		(0.711)		(0.397)		(0.873)
Number of publications (1996-2017, in log)		-0.005		0.005		0.011*		0.007		-0.007*		-0.009**		-0.001
		(0.323)		(0.166)		(0.047)		(0.137)		(0.020)		(0.009)		(0.839)
Average of field-normalised citations per document (1996-2017)		0.008		0.000		0.002		- 0.009**		0.001		-0.002		-0.000
		(0.065)		(0.919)		(0.629)		(0.009)		(0.571)		(0.410)		(0.934)
Collaboration (1996-2017)		-0.084**		0.004		0.058*		0.016		0.027*		-0.011		0.027
		(0.002)		(0.828)		(0.043)		(0.468)		(0.041)		(0.563)		(0.110)
Share of open access publications (1996-2017)		0.032		0.023		-0.008		0.016		-0.008		-0.029		-0.001
		(0.259)		(0.219)		(0.777)		(0.494)		(0.595)		(0.117)		(0.959)
Constant	0.056	0.145*	1.020***	1.007***	0.950***	0.896***	-0.021	-0.075	-0.053*	-0.076**	-0.003	0.024	0.021	-0.008
	(0.313)	(0.025)	(0.000)	(0.000)	(0.000)	(0.000)	(0.660)	(0.177)	(0.027)	(0.009)	(0.934)	(0.560)	(0.654)	(0.869)
Observations	6 694	6 160	6 694	6 160	6 694	6 160	6 694	6 160	6 694	6 160	6 694	6 160	6 694	6 160
R-squared	0.064	0.070	0.040	0.044	0.045	0.053	0.045	0.050	0.059	0.057	0.064	0.068	0.126	0.128

Note: P-values in parentheses. Country- and science field- fixed effects included in the regressions. * refers to p<0.05; ** refers to p<0.01; *** refers to p<0.001. *ICT graduate* is a dummy variable, which is equal to 1 if an author holds a degree in computer and information sciences, electrical engineering, electronic engineering, or information engineering, and 0 otherwise. C*ollaboration* is measured by the proportion of an author's documents involving co-authorship and indexed in Scopus in the period 1996-2017. *Share of open access publications* refers to the number of an author's open access-publications indexes in Scopus on the author's total number of publications.

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

Table D.6. Drivers of probability of engagement in activities beyond core research

Least square regression estimates

	Teaching activity		Intellectual property applied for or granted		Consultancy work		Company founder or executive board membership		Management or advisory board membership		Public service or political work	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Female	0.008	0.008	-0.069***	-0.060***	-0.006	-0.005	-0.029*	-0.029*	-0.008	0.002	-0.010	-0.007
	(0.407)	(0.464)	(0.000)	(0.000)	(0.694)	(0.735)	(0.015)	(0.022)	(0.561)	(0.883)	(0.438)	(0.605)
Age	-0.001*	-0.001**	0.004***	0.004***	0.004***	0.003***	0.004***	0.003***	0.009***	0.008***	0.003***	0.003***
	(0.032)	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
PhD	0.118***	0.085***	0.074***	0.065***	0.082***	0.051*	0.029	0.030	0.146***	0.090***	0.003	-0.014
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.019)	(0.075)	(0.096)	(0.000)	(0.000)	(0.889)	(0.498)
Born in another country	0.008	0.000	0.035**	0.035*	0.003	-0.005	-0.003	-0.008	-0.029	-0.027	-0.090***	-0.091***
	(0.477)	(0.969)	(0.009)	(0.013)	(0.871)	(0.743)	(0.812)	(0.556)	(0.065)	(0.098)	(0.000)	(0.000)
Digital productivity tools	0.054***	0.047***	0.022***	0.022***	0.071***	0.069***	0.032***	0.032***	0.072***	0.065***	0.048***	0.046***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Data/Code dissemination	0.016***	0.016***	0.029***	0.026***	0.008	0.011	0.016**	0.018**	0.018**	0.017**	0.027***	0.029***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.193)	(0.090)	(0.003)	(0.002)	(0.004)	(0.009)	(0.000)	(0.000)
Advanced digital tools/data	0.012*	0.016**	0.067***	0.058***	0.071***	0.071***	0.074***	0.073***	0.060***	0.061***	0.068***	0.073***
	(0.012)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Digital identity and communication	0.050***	0.046***	0.041***	0.045***	0.090***	0.095***	0.061***	0.065***	0.096***	0.093***	0.083***	0.081***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ICT graduate		0.000		0.052*		-0.039		-0.047*		-0.047		-0.068**
		(0.983)		(0.042)		(0.142)		(0.038)		(0.056)		(0.005)
Baseline = GOV												
PNP		-0.019		0.022		0.030		0.030		0.011		-0.014
		(0.510)		(0.444)		(0.375)		(0.314)		(0.725)		(0.663)
BE		-0.029		0.148***		0.063*		0.080***		-0.002		-0.018
		(0.189)		(0.000)		(0.014)		(0.001)		(0.926)		(0.431)
HE		0.120***		-0.000		0.031		-0.014		0.084***		-0.015
		(0.000)		(0.974)		(0.058)		(0.306)		(0.000)		(0.327)
		0.013**		0.017**		0.019**		0.007		0.038***		-0.002

	Teaching activity		Intellectual property applied for or granted		Consultancy work		Company founder or executive board membership		Management or advisory board membership		Public service or political work	
Number of publications (1996-2017 in log)		(0.004)		(0.001)		(0.001)		(0.187)		(0.000)		(0.660)
Average of field-normalised citations per document (1996-2017)		0.003		0.002		-0.009*		-0.012***		-0.007		0.003
		(0.352)		(0.607)		(0.047)		(0.001)		(0.155)		(0.512)
Collaboration (1996-2017)		0.024		0.092***		0.077**		0.030		0.014		-0.049
		(0.270)		(0.000)		(0.010)		(0.224)		(0.612)		(0.088)
Share of open access publications (1996-2017)		0.031		-0.032		-0.036		-0.044		-0.010		-0.030
		(0.154)		(0.195)		(0.234)		(0.076)		(0.726)		(0.286)
Constant	0.817***	0.820***	-0.443***	-0.524***	-0.547***	-0.596***	-0.328***	-0.334***	-0.623***	-0.587***	-0.265***	-0.190**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)
Observations	6 204	5 721	5 970	5 530	5 976	5 533	5 927	5 488	6 103	5 659	5 812	5 387
R-squared	0.111	0.141	0.224	0.242	0.144	0.152	0.120	0.135	0.173	0.186	0.133	0.135

Note: P-values in parentheses. Country- and science field- fixed effects included in the regressions. * refers to p<0.05; ** refers to p<0.01; *** refers to p<0.001. *ICT graduate* is a dummy variable, which is equal to 1 if an author holds a degree in computer and information sciences, electrical engineering, electronic engineering, or information engineering, and 0 otherwise. C*ollaboration* is measured by the proportion of an author's documents involving co-authorship and indexed in Scopus in the period 1996-2017. *Share of* o*pen access publications* refers to the number of an author's open access-publications indexes in Scopus on the author's total number of publications.

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

Table A D.7. Link between research quality and author profile

Least square regression estimates

	Average of field-normalised citations per document (1996-2017)			Journal prestige (Average SJR, 1996-2017)		
	(1)	(2)	(3)	(1)	(2)	(3)
Female	-0.064	0.015	-0.004	-0.009	-0.015	-0.019
	(0.139)	(0.695)	(0.921)	(0.676)	(0.471)	(0.381)
Age (10 years)	-0.046**	-0.124***	-0.130***	-0.116***	-0.132***	-0.122***
	(0.008)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
PhD	0.396***	0.185***	0.211***	0.330***	0.247***	0.261***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Born in another country	0.035	0.064	0.069	0.006	-0.004	-0.009
	(0.504)	(0.180)	(0.150)	(0.827)	(0.878)	(0.753)
Digital productivity tools	0.086***	0.053***	0.049**	0.046***	0.023*	0.023*
	(0.000)	(0.001)	(0.002)	(0.000)	(0.014)	(0.022)
Data/Code dissemination	0.073***	0.057***	0.061***	0.055***	0.051***	0.045***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Advanced digital tools/data	0.027	0.022	0.014	-0.027**	-0.019*	-0.028**
	(0.119)	(0.187)	(0.401)	(0.003)	(0.046)	(0.005)
Digital identity and communication	0.042*	0.022	0.020	-0.011	-0.032**	-0.035***
	(0.013)	(0.184)	(0.221)	(0.250)	(0.001)	(0.001)
ICT graduate		0.119	0.083		-0.331***	-0.348***
		(0.072)	(0.210)		(0.000)	(0.000)
Baseline = GOV						
PNP		0.137	0.171		0.053	0.056
		(0.098)	(0.056)		(0.283)	(0.299)
BE		0.014	0.043		-0.092**	-0.079*
		(0.809)	(0.459)		(0.005)	(0.025)
HE		0.011	0.018		0.032	0.029
		(0.794)	(0.649)		(0.185)	(0.236)
Average number of working hours per week (in log)		-0.007	0.005		0.064**	0.061**
		(0.877)	(0.890)		(0.003)	(0.006)
Percentage of weekly working time spent on research		0.002**	0.002**		0.004***	0.004***
		(0.001)	(0.004)		(0.000)	(0.000)
Number of publications (1996-2017 in log)		0.225***	0.231***		0.041***	0.033***
		(0.000)	(0.000)		(0.000)	(0.001)
Collaboration (1996-2017)			0.302***			0.290***
			(0.000)			(0.000)
Work modality (baseline='indefinite,' not protected)						
Fixed term contract			0.059			0.034
			(0.191)			(0.247)
Indefinite, highly protected contract			0.015			0.005
			(0.753)			(0.847)
Constant	0.058	0.065	-0.210	1.606***	1.087***	0.797***
	(0.689)	(0.764)	(0.326)	(0.000)	(0.000)	(0.000)
Observations	6 838	6 320	5 934	6 838	6 320	5 934
R-squared	0.110	0.165	0.184	0.293	0.326	0.332

Note: P-values in parentheses. Country- and science field- fixed effects included in the regressions. * refers to p<0.05; ** refers to p<0.01; *** refers to p<0.001. *ICT graduate* is a dummy variable, which is equal to 1 if an author holds a degree in computer and information sciences, electrical engineering, electronic engineering, or information engineering, and 0 otherwise. C*ollaboration* is measured by the proportion of an author's documents involving co-authorship and indexed in Scopus in the period 1996-2017.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Table D.8. Link between scientific authors' income and their profiles

Least square regression estimates on log gross income

	(1)	(2)	(3)
Female	-0.111***	-0.049**	-0.055***
	(0.000)	(0.003)	(0.001)
Age (10 years)	0.162***	0.127***	0.128***
	(0.000)	(0.000)	(0.000)
PhD	0.311***	0.219***	0.201***
	(0.000)	(0.000)	(0.000)
Born in another country	-0.068***	-0.057**	-0.046**
	(0.000)	(0.001)	(0.008)
Digital productivity tools	0.051***	0.038***	0.038***
	(0.000)	(0.000)	(0.000)
Data/Code dissemination	-0.017*	-0.012	-0.014
	(0.016)	(0.080)	(0.051)
Advanced digital tools/data	-0.005	-0.007	-0.003
	(0.510)	(0.405)	(0.706)
Digital identity and communication	0.015*	0.004	0.003
	(0.048)	(0.580)	(0.714)
ICT graduate		-0.055*	-0.015
		(0.049)	(0.599)
Baseline = GOV			
PNP		0.091*	0.082
		(0.029)	(0.054)
BE		0.101**	0.129***
		(0.001)	(0.000)
HE		-0.026	-0.026
		(0.123)	(0.127)
Average number of working hours per week (in log)		0.186***	0.146***
		(0.000)	(0.000)
Percentage of weekly working time spent on research		-0.002***	-0.001***
		(0.000)	(0.000)
Number of publications (1996-2017 in log)		0.103***	0.091***
		(0.000)	(0.000)
Collaboration (1996-2017)			-0.009
			(0.778)
Work modality (baseline='indefinite,' not protected)			
Fixed term contract			-0.128***
			(0.000)
Indefinite, highly protected contract			0.078***
			(0.000)
Average of field-normalised citations per document (1996-2017)			
2th quintile			-0.048
			(0.082)
3rd quintile			-0.031
			(0.289)

		(1)	(2)	(3)
	4th quintile			-0.036
				(0.235)
	5th quintile			0.020
				(0.523)
Journal prestige (Average SJR, 1996-2017)				
	2th quintile			-0.002
				(0.946)
	3rd quintile			0.047
				(0.091)
	4th quintile			0.057*
				(0.048)
	5th quintile			0.125***
				(0.000)
Constant		2.265***	1.795***	1.920***
		(0.000)	(0.000)	(0.000)
Observations		5 755	5 381	5 073
R-squared		0.598	0.638	0.666

Note: P-values in parentheses. Country- and science field- fixed effects included in the regressions. * refers to p<0.05; ** refers to p<0.01; *** refers to p<0.001. *ICT graduate* is a dummy variable, which is equal to 1 if an author holds a degree in computer and information sciences, electrical engineering, electronic engineering, or information engineering, and 0 otherwise. C*ollaboration* is measured by the proportion of an author's documents involving co-authorship and indexed in Scopus in the period 1996-2017.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Table D.9. Authors' attitudes towards digitalisation and its potential impacts

Least square regression estimates

	General views (combined index)			Efficiency of scientific work			Quality of scientific research			Collaborative and interactive science		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Female	-0.019	-0.006	-0.003	-0.117*	-0.112	-0.107	0.007	0.049	0.050	0.070	0.110	0.104
	(0.693)	(0.900)	(0.953)	(0.042)	(0.064)	(0.081)	(0.916)	(0.468)	(0.464)	(0.410)	(0.218)	(0.251)
Age	-0.001	0.001	0.001	-0.004	-0.002	-0.001	0.000	0.000	0.000	0.002	0.003	0.003
	(0.595)	(0.647)	(0.686)	(0.085)	(0.487)	(0.657)	(0.994)	(0.973)	(0.917)	(0.622)	(0.463)	(0.409)
PhD	-0.128*	-0.089	-0.070	-0.154*	-0.133	-0.108	-0.094	-0.120	-0.115	0.060	0.065	0.070
	(0.045)	(0.208)	(0.335)	(0.048)	(0.116)	(0.217)	(0.265)	(0.203)	(0.232)	(0.619)	(0.626)	(0.608)
Born in another country	0.081	0.076	0.101	0.114	0.103	0.123	0.074	0.076	0.098	0.032	0.040	0.068
	(0.104)	(0.144)	(0.058)	(0.066)	(0.109)	(0.061)	(0.278)	(0.285)	(0.177)	(0.718)	(0.675)	(0.483)
Digital productivity tools	0.197***	0.197***	0.192***	0.229***	0.230***	0.225***	0.254***	0.249***	0.252***	0.213***	0.207***	0.200***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Data/Code dissemination	0.071***	0.045*	0.040	0.045	0.025	0.022	0.110***	0.078**	0.077**	-0.001	-0.028	-0.028
	(0.001)	(0.034)	(0.060)	(0.067)	(0.318)	(0.396)	(0.000)	(0.008)	(0.010)	(0.982)	(0.480)	(0.482)
Advanced digital tools/data	0.042	0.037	0.031	0.030	0.019	0.013	0.092**	0.085**	0.079*	0.034	0.015	0.013
	(0.087)	(0.152)	(0.239)	(0.300)	(0.522)	(0.684)	(0.003)	(0.009)	(0.016)	(0.390)	(0.716)	(0.757)
Digital identity and communication	0.021	0.022	0.023	-0.047	-0.043	-0.045	-0.057	-0.060	-0.052	0.085*	0.071	0.073
	(0.330)	(0.337)	(0.329)	(0.072)	(0.122)	(0.110)	(0.058)	(0.058)	(0.110)	(0.030)	(0.081)	(0.081)
ICT graduate		0.187*	0.181*		0.325***	0.343***		0.279*	0.269*		0.133	0.121
		(0.018)	(0.025)		(0.000)	(0.000)		(0.015)	(0.021)		(0.384)	(0.439)
Baseline = GOV												
PNP		-0.080	-0.092		-0.006	-0.017		0.025	-0.009		0.145	0.106
		(0.457)	(0.404)		(0.962)	(0.894)		(0.864)	(0.950)		(0.445)	(0.587)
BE		-0.125	-0.113		-0.001	-0.002		-0.126	-0.102		0.041	0.097
		(0.137)	(0.198)		(0.991)	(0.983)		(0.281)	(0.399)		(0.788)	(0.542)
HE		-0.081	-0.060		-0.060	-0.044		0.036	0.048		0.025	0.036
		(0.124)	(0.263)		(0.350)	(0.499)		(0.624)	(0.516)		(0.793)	(0.715)
Number of publications (1996-2017, in log)		-0.028	-0.024		-0.035	-0.036		0.019	0.027		0.026	0.021

	General views (combined index)			Efficiency of scientific work			Quality of scientific research			Collaborative and interactive science		
		(0.136)	(0.212)		(0.119)	(0.110)		(0.451)	(0.293)		(0.400)	(0.514)
Collaboration (1996-2017)		0.262**	0.264**		0.328**	0.315*		0.187	0.167		0.438*	0.445*
		(0.008)	(0.009)		(0.007)	(0.011)		(0.154)	(0.214)		(0.019)	(0.020)
Share of open access publications (1996-2017)		0.203*	0.219*		0.104	0.084		0.254	0.263		0.319	0.344
		(0.050)	(0.036)		(0.376)	(0.483)		(0.059)	(0.053)		(0.086)	(0.067)
Counts of research outputs			0.001			0.028			0.009			0.021
			(0.987)			(0.621)			(0.894)			(0.807)
Online use			0.037			-0.032			-0.035			-0.095
			(0.546)			(0.657)			(0.662)			(0.379)
Constant	4.751***	4.329***	4.294***	4.453***	4.028***	4.005***	3.355***	3.066***	3.060***	9.888***	6.171***	8.821***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	6 816	6 266	6 071	6 810	6 261	6 066	6 787	6 241	6 048	3 395	3 120	3 021
R-squared	0.079	0.087	0.085	0.071	0.080	0.078	0.062	0.068	0.067	0.091	0.103	0.103

Note: P-values in parentheses. Country- and science field- fixed effects included in the regressions. * refers to p<0.05; ** refers to p<0.01; *** refers to p<0.001. Survey respondents were asked to rate opposing scenarios regarding the impact of digitalisation on different dimensions of science from 1 (fully agree with negative view) to 10 (fully agree with positive view). Responses by dimension were used as dependent variables in the regressions. *ICT graduate* is a dummy variable, which is equal to 1 if an author holds a degree in computer and information sciences, electrical engineering, electronic engineering, or information engineering, and 0 otherwise. C*ollaboration* is measured by the proportion of an author's documents involving co-authorship and indexed in Scopus in the period 1996-2017. *Share of open access publications* refers to the number of an author's open access-publications indexes in Scopus on the author's total number of publications.

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

Table D.10. Authors' attitudes towards digitalisation and their profiles, continued

Least square regression estimates

	Science across borders			Inclusiveness of research opportunities			Functioning of incentives and rewards in science			Private sector engagement in digital solutions for science		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Female	0.008	0.003	-0.006	0.034	0.043	0.043	-0.049	-0.002	-0.010	0.041	0.036	0.051
	(0.921)	(0.976)	(0.941)	(0.598)	(0.534)	(0.535)	(0.618)	(0.988)	(0.925)	(0.660)	(0.714)	(0.613)
Age	-0.006*	-0.004	-0.004	-0.001	0.002	0.002	0.009*	0.012**	0.010*	-0.004	-0.002	-0.004
	(0.036)	(0.218)	(0.235)	(0.582)	(0.422)	(0.489)	(0.020)	(0.004)	(0.013)	(0.253)	(0.651)	(0.386)
PhD	-0.144	-0.103	-0.103	-0.148	-0.049	-0.041	-0.190	-0.129	-0.089	-0.194	-0.098	-0.062
	(0.168)	(0.373)	(0.378)	(0.088)	(0.610)	(0.681)	(0.163)	(0.401)	(0.570)	(0.120)	(0.486)	(0.666)
Born in another country	0.110	0.078	0.137	0.001	-0.007	0.009	0.057	0.074	0.058	0.222*	0.213*	0.262*
	(0.186)	(0.366)	(0.119)	(0.990)	(0.926)	(0.900)	(0.604)	(0.520)	(0.620)	(0.029)	(0.048)	(0.017)
Digital productivity tools	0.217***	0.201***	0.180***	0.134***	0.136***	0.131***	0.123**	0.145**	0.146**	0.171***	0.157***	0.163***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.006)	(0.002)	(0.002)	(0.000)	(0.000)	(0.000)
Data/Code dissemination	0.071*	0.048	0.034	0.060*	0.035	0.027	0.032	0.009	0.004	0.152***	0.122**	0.124**
	(0.029)	(0.168)	(0.336)	(0.034)	(0.230)	(0.367)	(0.475)	(0.849)	(0.933)	(0.000)	(0.005)	(0.005)
Advanced digital tools/data	-0.051	-0.049	-0.056	0.012	0.020	0.016	0.044	0.041	0.038	0.117**	0.107*	0.097*
	(0.192)	(0.231)	(0.178)	(0.709)	(0.555)	(0.625)	(0.354)	(0.414)	(0.453)	(0.010)	(0.027)	(0.046)
Digital identity and communication	0.073*	0.060	0.064	0.091**	0.095**	0.104**	0.131**	0.149**	0.135**	-0.063	-0.064	-0.059
	(0.043)	(0.117)	(0.096)	(0.002)	(0.003)	(0.001)	(0.005)	(0.002)	(0.008)	(0.143)	(0.164)	(0.206)
ICT graduate		0.171	0.187		0.091	0.074		-0.157	-0.214		0.269	0.293
		(0.144)	(0.110)		(0.415)	(0.517)		(0.421)	(0.281)		(0.097)	(0.075)
Baseline = GOV												
PNP		-0.333	-0.290		-0.112	-0.139		-0.203	-0.270		-0.166	-0.135
		(0.096)	(0.151)		(0.450)	(0.362)		(0.361)	(0.240)		(0.435)	(0.533)
BE		-0.291	-0.286		-0.139	-0.134		-0.088	-0.111		-0.215	-0.216
		(0.058)	(0.072)		(0.225)	(0.262)		(0.594)	(0.516)		(0.230)	(0.245)
HE		0.034	0.039		-0.156*	-0.141		-0.271*	-0.238*		-0.156	-0.123
		(0.688)	(0.650)		(0.033)	(0.057)		(0.015)	(0.035)		(0.121)	(0.231)

	Science across borders			Inclusiveness of research opportunities			Functioning of incentives and rewards in science			Private sector engagement in digital solutions for science		
Number of publications (1996-2017, in log)		-0.080**	-0.082**		-0.062*	-0.059*		-0.024	-0.016		-0.042	-0.023
		(0.006)	(0.006)		(0.012)	(0.021)		(0.514)	(0.685)		(0.263)	(0.553)
Collaboration (1996-2017)		0.109	0.070		0.151	0.180		0.495*	0.595**		0.257	0.226
		(0.487)	(0.653)		(0.241)	(0.175)		(0.015)	(0.005)		(0.179)	(0.245)
Share of open access publications (1996-2017)		0.159	0.247		0.252	0.271*		0.080	0.074		0.252	0.282
		(0.370)	(0.153)		(0.062)	(0.048)		(0.700)	(0.722)		(0.219)	(0.171)
Counts of research outputs			0.124			0.056			-0.180			-0.141
			(0.103)			(0.382)			(0.071)			(0.131)
Online use			-0.141			0.128			0.426***			0.060
			(0.168)			(0.111)			(0.000)			(0.610)
Constant	3.656***	3.501***	3.454***	6.373***	6.001***	5.954***	9.305***	8.554***	8.656***	7.029***	6.538***	6.475***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	3 392	3 122	3 027	6 764	6 219	6 028	3 395	3 121	3 024	3 372	3 105	3 012
R-squared	0.075	0.081	0.087	0.063	0.068	0.069	0.122	0.129	0.136	0.084	0.088	0.089

Note: P-values in parentheses. Country- and science field- fixed effects included in the regressions. * refers to p<0.05; ** refers to p<0.01; *** refers to p<0.001. Survey respondents were asked to rate opposing scenarios regarding the impact of digitalisation on different dimensions of science from 1 (fully agree with negative view) to 10 (fully agree with positive view). Responses by dimension were used as dependent variables in the regressions. *ICT graduate* is a dummy variable, which is equal to 1 if an author holds a degree in computer and information sciences, electrical engineering, electronic engineering, or information engineering, and 0 otherwise. C*ollaboration* is measured by the proportion of an author's documents involving co-authorship and indexed in Scopus in the period 1996-2017. *Share of open access publications* refers to the number of an author's open access-publications indexes in Scopus on the author's total number of publications.

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