

Bornmann, L., Guns, R., Thelwall, M., Wolfram. (2021) Which aspects of the Open Science agenda are most relevant to scientometric research and publishing? An opinion paper. Quantitative Science Studies. Advance Publication. https://doi.org/10.1162/qss_a_00121

Opinion paper

Which aspects of the Open Science agenda are most relevant to scientometric research and publishing? An opinion paper

Lutz Bornmann*, Raf Guns**,

Michael Thelwall***, Dietmar Wolfram****

* Division for Science and Innovation Studies
Administrative Headquarters of the Max Planck Society
Hofgartenstr. 8,
80539 Munich, Germany.
Email: bornmann@gv.mpg.de
Orcid: 0000-0003-0810-7091

** Centre for R&D Monitoring
Faculty of Social Sciences, University of Antwerp
Middelheimlaan 1,
2020 Antwerpen, Belgium.
Email: raf.guns@uantwerpen.be
Orcid: 0000-0003-3129-0330

*** Statistical Cybermetrics Research Group
School of Mathematics and Computer Science, University of Wolverhampton
Wulfruna Street,
Wolverhampton WV1 1LY, UK.
Email: m.thelwall@wlv.ac.uk
Orcid: 0000-0001-6065-205X

**** School of Information Studies
University of Wisconsin – Milwaukee
P.O. Box 413,
Milwaukee, WI, USA, 53201.
Email: dwolfram@uwm.edu
Orcid: 0000-0002-4991-276X

Copyright: © 2021 Lutz Bornmann, Raf Guns, Michael Thelwall, and Dietmar Wolfram. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

Bornmann, L., Guns, R., Thelwall, M., Wolfram. (2021) Which aspects of the Open Science agenda are most relevant to scientometric research and publishing? An opinion paper. Quantitative Science Studies. Advance Publication. https://doi.org/10.1162/qss_a_00121

Abstract

Open Science is an umbrella term that encompasses many recommendations for possible changes in research practices, management, and publishing with the objective to increase transparency and accessibility. This has become an important science policy issue that all disciplines should consider. Many Open Science recommendations may be valuable for the further development of research and publishing but not all are relevant to all fields. This opinion paper considers the aspects of Open Science that are most relevant for scientometricians, discussing how they can be usefully applied.

Key words

bibliometrics; scientometrics; Open Science; open data; open peer review; open access

1 Introduction

Whilst modern science has elements of openness at its core, such as the expectation that results are shared in some form rather than kept secret, there is currently a call for increased openness. Spellman, Gilbert and Corker (2018) characterize Open Science as a collection of actions that contribute to enhance the replicability, robustness, and accessibility of research. Increased accessibility may also support diversity in research. Related to this, sociologist Robert K. Merton formulated the ethos of science as follows: communalism, universalism, disinterestedness, and organized skepticism (CUDOS, Merton, 1942). Nevertheless, the PLACE counter-norms – proprietary, local, authority, commissioned, expert – prevail in the laboratory context of discovery (Latour & Woolgar, 1979; Latour, 1987; Ziman, 2000). Moreover, intellectual property rights have increasingly penetrated the core process of knowledge production and control in the emerging knowledge-based economies (Whitley, 1984). Thus, openness is often not achieved in practice.

Open Science is an important topic in science-policy debates, with many recommendations to improve transparency in research and publishing (Fecher & Friesike, 2014). As the Open Research Glossary (see https://figshare.com/articles/Open_Research_Glossary/1482094) suggests, the term Open Science relates to multiple phases in the research and publication process: from starting a study (e.g., by preregistering it) to assessing the value of the published outcomes (e.g., by using metrics other than traditional bibliometrics such as data citations). Open Science encourages multiple citable units for a given piece of research, including preregistration documents for studies, open notes, multiple manuscript versions made available through Open Access (OA), associated datasets and software, author responses to reviewer comments and the accompanying reviews that are public in the case of open peer-review sources.

Open Science is partly driven by a science-policy agenda (García-Peñalvo Francisco, García de Figuerola, & Merlo José, 2010). The emergence of this term and the related movements is intrinsically connected to new forms of knowledge production (post-academic science and Triple Helix) in the knowledge-based economy, and the opportunities web-based technologies offer. There would be no possibility for any of the ‘open the black box’ steps (as pre-registration of research, disentangling of the whole research process, open data) without the internet. At the same time, Open Science is hailed as the way to give a push to interdisciplinary research and to the re-use of data (perceived by science policy as one of the motors behind innovations). However, there exist also pushes towards Open Science from within the academic community. For example, the Open Journal Systems (OJS) software, developed by the Public Knowledge Project, has been embraced by research communities in countries which cannot afford scientific publishers (Willinsky, 2005).

Open Science shares many commonalities with the Free and Open Source Software (FOSS) movement (Tennant et al., 2020). FOSS is itself an umbrella term used to refer to both the Free Software movement, which started in the 1980s and the Open Source movement, which started around 1998. While Free Software focuses on social issues, embodied in four ‘fundamental freedoms’, Open Source is a more pragmatic reformulation of Free Software ideals, by focusing mainly on practical advantages of source code availability (Tozzi, 2017).

The term Open Science has many facets, and it is not clear which tools, initiatives, ideas, and recommendations are relevant and meaningful for good scientific practices in a given research field. For example, recommendations for Open Science collaborations (Gold, Ali-Khan, Allen, 2019) do not seem relevant for scientometrics due to the typically small-scale nature of research in this field (in terms of the number of co-authors associated with a paper). Furthermore, there is no general agreement on which facets are part of Open Science. It has even been argued that the Open Science movement is not necessary for the (current)

science system, since science is (always) open to some extent (Fecher & Friesike, 2014).

Watson (2015) argues that Open Science is ‘just (good) science’. Independent of the roots of the ideas which are discussed in the Open Science context, it might be worthwhile for every discipline to engage with Open Science recommendations and identify relevant ideas.

Every Open Science recommendation has the potential to improve research as well as an inherent cost, even if it is only the time taken to implement it. For example, data sharing can foster collaboration and more rapid development of a research topic, but can also require substantial effort to arrange data in a common format and create an effective set of metadata. Additional requirements that add to the burden of research publishing potentially disadvantage researchers (e.g., in less rich nations) who have little time, equipment and support to deal with them easily.

Scientometric research often analyzes collections of journal articles, and these have been affected by Open Science, such as with the introduction of open access general mega-journals. The possibilities of distributed digital publishing may change the nature or importance of journal articles and lead to a greater focus on outputs of all types (Guédon, Kramer, Laakso, et al., 2019). Journal articles have largely changed from static paper objects to electronic documents with embedded links to other documents (in the references) as well as data and visualizations, all of which increase the importance of open infrastructures. Bourne et al. (2012) call for further innovation in the forms and technologies of scholarly communication as well as the underlying business and access models. Such developments may have a substantial impact on future scientometric research, for instance, if datasets and software become recognized as first-class research objects in their own right.

Scientometrics research is affected by this new Open Science movement in two ways: as a research field itself, and as a research field monitoring the science or academic system. In this paper, we mainly focus on the former, occasionally touching also on the latter. This opinion paper identifies recommendations proposed in the Open Science context that may be

relevant and meaningful for scientometric research and publishing, although not necessarily in all or most circumstances. We discuss the most interesting aspects from the Open Science literature for scientometric research. We avoid to be normative: researchers in scientometrics should be informed about the various aspects, but should decide themselves whether they are of interest for their own research or not. The discussion of the Open Science issues in the following sections is ordered by the phases in typical research and publication processes (starting a study and publishing its outcome).

2 Relevance of open science to scientometrics

Scientometric research is characterized by heterogeneity. In general, there is a difference between the sociological ambition of studying the sciences and visualizing and mapping them (see, e.g., the contributions by Robert K. Merton) and the connected, but different objective of research evaluation. In the latter field, researchers and their institutes are often units of analysis, whereas the structuralist approaches share with information and library sciences a focus on document sets. Scientometrics can be considered as a Mode-2 science: in addition to its theoretical core mission of studying the sciences, there are applications in science studies considering scientometrics as offering tools and methods for the quantification, research evaluation on the applied side, and the development of expertise in practice. The call for ‘opening the black box’ is common for research evaluation practice.

Within this broader framework of distinguishing two broad directions of scientometric research, scientometric studies are very different in terms of their topics, methods and indicators used, as well as their scopes. Some studies are policy oriented (e.g., Rushforth & de Rijcke, 2015), others are mathematically oriented (e.g., Egghe & Rousseau, 2006). Some studies are based on huge datasets (e.g., Visser, van Eck, & Waltman, 2020), others focus on small datasets or report the results of one case study (e.g., Marx, 2014). Since visualizations of scientometric data have become very popular in recent years, some papers focus on the

explanation of tools or software (e.g., Bornmann, Stefaner, de Moya Anegón, & Mutz, 2014). Research on the *h* index revealed, for instance, that papers dealing with the same topic – this indicator – can be very heterogeneous: some studies mathematically investigated the indicator (e.g., Egghe & Rousseau, 2006), others addressed its convergent validity (e.g., Bornmann & Daniel, 2005), tried to develop new variants of the index (see Bornmann, Mutz, Hug, & Daniel, 2011), or explained tools for its calculation on the internet (e.g., Thor & Bornmann, 2011).

Against the backdrop of the intermingling of heterogeneous objectives and styles in scientometrics, Open Science recommendations may be relevant to different degrees for these studies. For example, for mathematically oriented studies, other recommendations would be relevant than for studies from the policy context. In this paper, therefore, we focus on Open Science recommendations that might be made relevant for studies from the scientometric area using (more extensive) data from literature databases such as Web of Science (Clarivate Analytics) or Scopus (Elsevier) for empirical investigations in the science of science area (Fortunato et al., 2018). These studies may focus, for instance, on certain structures in science (e.g., the existence of the Matthew effect) or the effectiveness of certain funding instruments for the promotion of science. In other words, we consider the study of data as the common denominator of the scientometric enterprise.

Authors of this core type of science of science studies may be interested in the Open Science issues raised in this opinion paper. Following the recommendations in scientometrics research might lead to research with a higher quality, accessibility, and transparency. Scientometricians can consider the Open Science recommendations addressed in the sections 3 and 4 of this paper. If so wished, this type of study can be pre-registered, the underlying data, used codes, and applied software can be made available, the contributions of the co-authors explained, and the paper can be published OA including reviews and other documents originating from the peer review process. We selected the relevant recommendations from the

Open Science literature based on our long-standing experiences in the field of scientometrics. These are mostly directly relevant to the robustness/replication Open Science goal by making the research process more transparent, with the tools and data openly available. They also support diversity since openly shared papers, data, or software may be used to produce new research.

3 Starting a study

3.1 Open preregistration of studies and planned analyses

(Publicly funded) Research is less valuable for research itself and society when researchers cannot share their findings with others, and publishing is one way in which results can be shared. Other ways include giving presentations, giving interviews, creating podcasts, including research outcomes in teaching materials, and participating in meetings with policy makers. Since various studies have shown that results which validate a previously formulated hypothesis are far more likely to get published (when compared with a refutation), researchers might be interested in producing such publishable results (see Marewski & Bornmann, 2018).

To reach the goal of publishing while using the strict model of hypothesis-testing (Hempel & Oppenheim, 1948), there is a danger that hypotheses are formulated based on the data at hand and from the perspective of hindsight (e.g., to achieve results which are statistically significant). Thus, the statistical analysis is not used to test certain previously formulated hypotheses, but to fish results from the data which might be more publishable than other results (see Cohen, 1994). This is sometimes called ‘harking’ (hypothesizing after the results are known) (Hand 2014; see also: Cumming and Calin-Jageman, 2016). Another term is ‘historicism’ for the rationalization of research results *ex post* (Popper, 1967). However, analytical research questions or hypotheses are more common as a stated objective in some fields of science than others; with vague terms, such as *aim*, *purpose* or *goal* being common

alternatives (Thelwall & Mas-Bleda, 2020). These alternatives suggest a more exploratory research approach that may be also relevant for many types of scientometric study.

To make the formulation of hypotheses from the perspective of hindsight more difficult and to demonstrate that hypotheses were formulated before analyzing the data, it is possible to register a study in an early stage based on a detailed research plan that can be later checked by reviewers, editors, and readers (Kupferschmidt, 2018; Nosek, Ebersole, DeHaven, & Mellor, 2018; see also Cumming & Calin-Jageman, 2016). Preregistration can be done, for instance, at the Center for Open Science (see <https://cos.io/prereg>). The PLOS journals *PLOS Biology* and *PLOS ONE* currently offer the possibility to peer review and publish preregistered research (see <https://www.plos.org/preregistration>).

Published analysis plans for upcoming studies can be an interesting source for researchers working on similar research questions. Young researchers may profit from detailed plans published by experienced researchers of how the study of certain phenomena can be tackled.

The journal *Psychological Science* uses a badge created by the Center for Open Science for acknowledging papers being preregistered (see <https://cos.io/our-services/open-science-badges>). Currently, no scientometric journal works with this badge to highlight this open practice. Neither does any scientometric journal at the moment offer the option to peer review a preregistered study. When designing (and potentially preregistering) a scientometric study, researchers must specify many parameters: Which data will be used to analyze the research questions such as funding data, peer review data, data on the scientific workforce, survey data, or publication data? Which database (e.g., Web of Science, WoS, or Scopus) will be used? How many papers will be analyzed? How will citation impact be measured? Will the median or geometric mean be calculated instead of the arithmetic mean? Should percentages or the original scores be preferred? To make these decisions wisely, pilot testing may be needed (Cumming & Calin-Jageman, 2016). However, the data used in pilot testing should

not be fished for ‘publishable’ results because doing so would undermine the purpose of preregistration.

It is an advantage of scientometric research that data are often available in large-scale databases so that pilot testing can be a practical step in many studies. In other social sciences, the generation of data can be effortful. However, the availability of high-quality data can also be a disadvantage. Before preregistering a study, the whole study could be completed, so that the outcomes are already known. This is a practical possibility for scientometric research because data collection is rarely a substantial public event, unlike a clinical trial or survey.

Working with pilot testing can be especially important in the process of formulating a hypothesis. Although the hypothesis is logically prior, its formulation does not have to be prior in time. On the contrary, the formulation of expectations may be circular and repetitive. In pilot-testing the hypothesis, it can be improved, and the research can be made more precise and sophisticated. The hypothesis can be improved, for example, by being theoretically informed and rooted in the previous literature. A hypothesis may be a prediction, but this is not necessarily the case. A hypothesis is not like a weather forecast; ‘prediction’ has the meaning of a theoretically informed expectation that can be tested against the data as observations. Based on theoretically informed hypotheses, pilot testing can be used to operationalize, for example, by specifying expectations which can perhaps be tested.

Basically, in this phase of a planned study, the ‘logic of discovery’ differs from the ‘logic of justification’. In the ‘logic of discovery’, the relevant theorizing, hypotheses, and operationalization can be changed. The empirical results can be discussed with colleagues in the ‘logic of justification’. This process may lead to further changes until a more formal research plan can be finalized. The difference between the two logics is analytical; in research practices, both momenta are important.

Suppose a scientometric researcher is interested in the growth of science based on scientometric data. Two obvious data options are numbers of papers and researchers. Previous

studies investigating the growth of science detail advantages and disadvantages of these options (e.g., Bornmann & Mutz, 2015a; Tabah, 1999). Suppose the scientometrician has decided to use numbers of publications, he/she has to make further choice such as the question of which database is used for this study. Several databases are available with respective advantages and possible disadvantages each (e.g., Dimensions from Digital Science, WoS, Scopus or Microsoft Academic). Decisions are also to be made on the range of publication years and document types included. Fractional or whole-number counting may make a difference if the study includes the country-perspective.

Based on experiences from the literature and information from the database providers, the data can be selected, and expectations specified about the growth curves: in which years, for example, is an increase or decrease of publication numbers expected and for which reasons? In pilot testing, a decision could be drawn whether the growth of science is studied for certain disciplines (in addition to the whole of science) and – if so – which field-categorization scheme can be used. A sample of publication numbers can be used to test the different statistical methods (e.g., regression models or effect-size measures) to analyze the data. The research plan of the study contains the necessary steps to analyze the data.

In scientometric evaluations, tests are also used to determine whether differences are statistically significant (e.g., the difference between two mean or median citation rates). However, the results of statistical significance testing are dependent on factors such as the sample size: the larger a sample, the more likely to receive statistically significant results. Thus, the sample size which is needed to detect a certain effect in a study should be considered before the study is conducted. This practice may prevent the increase of the sample size by the researchers in hindsight, i.e., after they have inspected the results, to receive statistically significant results. The danger of increasing the sample size in hindsight is present in scientometrics, since the data are available in large literature databases such as WoS and can be downloaded without any major problems.

When considering the above in the planning of a scientometric study, it is important to consider that research evaluation – an important subset of questions in our field – can deal with incomplete data and with goals (e.g., research quality or impact assessment) that usually do not match the data (e.g., citation counts) well. In addition, the large number of variables potentially influencing the data and continual changes over time make strong hypotheses often impossible. Instead, unless using simulation or pure mathematical modelling, empirical studies must make multiple simplifying assumptions. In this situation, research questions, when used, are likely to be accompanied by strong caveats and may be primarily devices to frame the analysis in a paper.

Nevertheless, statistical tests and other uses of algorithms, such as for modeling or machine learning, should not be misused for harking because this would undermine the validity of the results. If a study is not pre-registered, researchers might consider openly publishing (e.g., as online supplementary material) details of prior failed tests, or visualizations of the preparation stages that were associated with the project.

3.2 Open data

There is widespread support for data sharing in academia (e.g., Tenopir, Dalton, Allard, et al., 2015), including for the FAIR (Findable, Accessible, Interoperable, Reusable) principles (Wilkinson et al., 2016). Data might be ‘open’ if it is made available under a license that allows free use, modification, and redistribution, for example. The most well-known example of such licenses is the family of Creative Commons licenses (see <http://opendefinition.org/licenses> for a broader overview). In practice, datasets are often made publicly available – and hence considered open data in a broader sense – without an explicit license. From a scientometric perspective, open data is relevant to disciplinary practices (i.e., should we share our data openly) as well as providing an issue for study: how can the impact of shared data be evaluated. Here we focus primarily on the former.

Datasets from scientometric studies which can be made publicly available can be shared through repositories like FigShare (see <https://figshare.com>) or Zenodo (see <https://zenodo.org>). The Open Materials and Open Data badges provided by the Open Science center can be used to indicate that a paper provides data and other materials for the use by other researchers (see Cumming & Calin-Jageman, 2016) and data sharing is increasingly mandated by funders and journals. In general, it is very helpful to have access to the data of studies for their replication, checking for possible errors, and conducting meta-analyses (Glass, 1976).

Open datasets may also help in avoiding duplication in data collection (Fecher & Friesike, 2014). The data can be used for other research questions than those addressed by the producers of the datasets. Researchers undoubtedly welcome broader access to datasets for scientometric research. For example, the aggregated data used for the various releases of the Leiden Ranking have been published for many years (see <https://www.leidenranking.com/downloads>). Because of their transparency, they have already been used as data by several papers (e.g., Bornmann & Mutz, 2015b; Frenken, Heimeriks, & Hoekman, 2017; Leydesdorff, Mingers, & Bornmann, 2019). However, in research evaluations the sharing of data may be sensitive to strategic or privacy issues.

Data sharing has some common obstacles in scientometrics. Although data sharing and reuse are goals of Open Science, much bibliometric research is based on data obtained from proprietary sources (e.g., WoS or Scopus). This limits what researchers can share, even if required by a journal. Recently, Digital Science has made the Dimensions database available for scientometric research (see <https://www.digital-science.com/products/dimensions>). Elsevier's International Center for the Study of Research (ICSR) also provides free scientometrics data. Other initiatives of open (or free to access) literature databases are Microsoft Academic Search (see <https://academic.microsoft.com>) and the Initiative for Open Citations (see <https://i4oc.org>). Thus, scientometricians can use a range of bibliometric

datasets for scientometric research without paying fees (and without restrictions in data sharing).

Based on these developments and requirements, it seems that the primary data sharing goal in scientometrics may no longer be to publish the data, but to publish the procedures to extract and analyze the data. In other words, it may be less useful to have access to the indicator values for publications than to have access to the published procedure for calculating the indicator. The access to certain datasets might become superfluous with the large dataset sharing initiative of Digital Science, although access in this case is mediated by the company and, as such, can in principle be revoked at any time. This may not be possible with openly licensed data. Moreover, many scientometric studies test new sources of data (e.g., patents, altmetrics, webometrics) and these will not be served by shared common datasets.

Whilst some disciplines have their own data repositories, supporting specific file formats, metadata types, and legal requirements, scientometrics does not. Perhaps the closest is the Initiative for Open Citations, which may lead to almost all citations being available from one source in the same format. Similarly, some scientometricians and other stakeholders have advocated for making other (meta)data openly available through the Crossref infrastructure and coupled to DOIs (e.g., the recent Initiative for Open Abstracts, <https://i4oa.org/>). Hence, Crossref is increasingly a source of raw scientometric data, which can be further refined and enhanced by other commercial and noncommercial players. It is not clear that there is a systematic need for a separate database for any other kind of scientometric data for several reasons. Many studies use Scopus, Web of Science or Dimensions data, which is already either freely available to researchers or copyright protected and not sharable. Some studies have small datasets of particular interest, such as studies of specialist topic within a given country, that may be of limited value to others. Other studies link scientometric data to other types of data (such as university recruitment or economic indicators) that would

rarely be useful to other scientometricians. The data of such studies can be made available on an ad hoc basis through generic data repositories, like Zenodo. Thus, there does not seem to be a universal pressing need for a common disciplinary data type or data repository.

From the perspective of the FAIR (Findable, Accessible, Interoperable, Reusable) principles, large free datasets or data set access mechanisms presumably adequately satisfy the Findable requirement through sparsity and because the current main providers (Clarivate Analytics, Elsevier, Dimensions, and Microsoft Academic) are well known. Accessibility is a problem for the subscription services (Clarivate Analytics and Elsevier). Interoperability and reusability are relatively minor problems for these datasets because researchers typically work with a single data source. In contrast, FAIR seems to be a substantial problem for the many ad-hoc small-scale datasets generated in scientometric research which seem to be rarely shared openly, and which are presumably poorly documented and in a wide variety of formats. If these cannot be indirectly shared by publishing the procedures by which the data was extracted from a well-known bibliometric source, then FAIR data sharing may be difficult. Such datasets might include, for example, altmetrics, publication lists from departments or researcher CVs, selections from national current research information systems, publication lists from national research evaluation exercises, and national journal collections. It seems reasonable for scientometrics journals to encourage researchers to share their datasets, when possible, or to publish detailed instructions about how to replicate the study otherwise, such as the relevant queries.

4 Publishing the outcome of a study (e.g., papers or software)

4.1 Open code or shared software

The code or compiled software for analyzing data can be shared. Spellman, Gilbert and Corker (2018) point out that direct replication without involvement of the original researchers is very difficult as the methods are typically not described in sufficient detail. The

same problems were found in a small-scale exploration of reproducibility issues in scientometrics (Velden et al., 2018). If the data have been analyzed using code that is made available, it becomes far easier to reproduce the analysis and/or spot errors in it that may otherwise go unnoticed. It seems unlikely, however, that ‘computational reproducibility’ (Baker, 2016) can help to prevent outright fraud.

Several technical factors may make reproducing results more difficult, such as random seeding of non-deterministic processes like sampling, ongoing updates of the hardware and software, and the use of parameters that may not be fully communicated in the publication. While all of these can be overcome in principle, they provide major hurdles in practice. Large-scale models (e.g., topic models) are often not reproducible because the updating of the systems (for example, for safety reasons) is not under the control of the researchers (Hecking, & Leydesdorff, 2019). Nevertheless, the routines can be published in an open source repository like Github (see <https://github.com>).

The code for integrating the results in the paper or even writing the complete paper can also be made available. The Stata corporation has recently developed new commands which can be used for producing a Word document (or PDF) based on the data (see <https://www.stata.com/features/overview/create-word-documents>). In other words, the commands do not only refer to the process of producing tables and figures, but also the complete paper. If both commands and data are made publicly available, every researcher can exactly reproduce the paper. Other statistics software provides similar functions: R Markdown supports reproducible papers by interspersing R code and text in markdown format (Bauer, 2018).

Papers in scientometrics are often based on data from WoS or Scopus. This data might have been downloaded from the web interfaces of these data providers. If the search terms for compiling the publication sets, search limiters used and the export date of the data are made publicly available, not only the paper production beginning with the data and ending with the

text can be reproduced, but also the process of generating the dataset. Nevertheless, both bibliometric and historical data can change over time, undermining replications. Changes in the media, updates of hardware and software may make data and procedures irreproducible. The major databases also backtrack newly admitted data into their history. In general, the databases are dynamic, and cannot easily be reproduced. Internet data, however, might be archived by the Wayback Machine (Wouters, Hellsten, & Leydesdorff, 2004).

Another source of bibliometric data are in-house databases (e.g., at the Centre for Science and Technology Studies or Max Planck Society) which are based on data from WoS, Scopus, Dimensions etc. In this case, the SQL for producing the data could be made available for compiling the data (including explaining text what the single command lines mean). However, there may be property rights involved when using this data.

There are several caveats for code sharing. Scientometric research may have involved extensive data cleaning by the users, creating significant added value to a standard dataset (e.g., WoS) in a proprietary clean version. With this, any software applied to the standard version of the database would give different results. The solution would be to make the cleaned version open (at least for research purposes) instead of proprietary, which is not always legally or practically feasible. In addition, some applications may be complex, generating a substantial amount of work to put the code in a shareable format. Thus, the likelihood of reuse is important when deciding whether and how to share code.

4.2 Contributions of authors

Authors contribute differently to research projects and resulting publications. The American Psychological Association (APA) published a checklist which can be used by contributors to a research project to declare their contributions (see

<https://www.apa.org/science/leadership/students/authorship-determination-scorecard.pdf>).

The use of such lists (1) might lead to a more standardized consideration of authors on papers

in the long run and (2) might be a possible action to avoid bad practices like ghost authorship (substantial contribution made without being listed as a co-author) and honorary authorship (being listed as a co-author while having contributed little to nothing). In the scientometrics field, authors have to state authors' contributions in some journals, e.g., *Journal of Informetrics* and *Quantitative Science Studies*.

The contributions of authors to research projects and publications can be made publicly available in the process of publishing a paper. CASRAI has published CRediT (Contributor Roles Taxonomy) for common authorship roles (see <https://www.casrai.org/credit.html>). Dealing with author contributions in research projects is important for knowing and codifying 'who did what' in the project. In many fields, but not in all, researchers have traditionally used author order when assessing authorship credit. We can imagine that an initiative like CRediT provides a more objective way to assess author credit. Practices may also vary among countries.

4.3 Open access (OA)

OA, where authors or publishers make research articles or reports freely available to readers in traditional or OA journals, represents the most visible aspect of Open Science. The number of OA journals has grown quickly since the 1990s. The Directory of Open Access Journals (see <https://doaj.org/>) included more than 14,300 journals by early 2020. An increasing number of publishers also support forms of OA in their traditional journals by providing authors the option to make their submissions freely available, for a fee, once accepted. These articles appear alongside the closed access articles that require a subscription to access.

There is a cost associated with making publications freely available (van Noorden, 2013). This cost is borne by the publishers, authors and/or third parties. OA models are characterized by who is responsible for the associated costs and the user rights to the content.

Two common OA models are Gold OA and Green OA. Gold OA publication venues have an associated Article Processing Charge (APC) that makes articles freely available to readers once accepted. OA journals that do not charge authors are sometimes also referred to as Gold OA, although other terminology (Diamond/Platinum OA) is also used. Green OA venues allow authors to self-archive pre-publication (pre- and sometimes post-peer review) versions of their manuscripts in public repositories such as arxiv.org. The rationale for Green OA is that the results of publicly funded research should be made publicly available without cost to the reader.

Several journals that publish scientometric research support different forms of OA. For example, PLOS journals, *Frontiers in Research Metrics and Analytics* and *Quantitative Studies of Science* support Gold OA. More traditional journals such as the *Journal of Informetrics*, *Journal of the Association for Information Science and Technology*, and *Scientometrics* support hybrid OA, where authors may pay an APC to make their articles freely available online. These three journals also permit pre-print archiving of manuscript submissions.

Recently, the OA topic has a specific relevance for the scientometric field, since the chief editor and editorial board of an important scientometric journal (the *Journal of Informetrics*) decided to change the publisher (from Elsevier to MIT Press, see Waltman, 2019). The new journal at the new publisher is *Quantitative Science Studies* (see <https://www.mitpressjournals.org/qss>). One reason for this change was that the outcome of scientometric research could be made available for other researchers without any restrictions.

In 2018, a group of European funding agencies formed cOAlition S, which developed a strategy for mandated OA called Plan S. “Plan S requires that recipients of research funding from cOAlition S organisations make the resulting publications available immediately (without embargoes) and under open licenses, either in quality Open Access platforms or journals or through immediate deposit in open repositories that fulfil the necessary

conditions” (<https://www.scienceeurope.org/our-priorities/open-access>). More than 1,700 members of the scientific community have signed an open letter expressing concerns that Plan S takes OA too far and is too risky. Although Plan S has since been revised to address some of the letter writers’ objections, there are still concerns that the plan favors APC OA models and does not address the researchers’ comments about the quality of peer review and international collaborations (see <https://sites.google.com/view/plansopenletter/home>).

4.4 Open data and software citation

When the data used for a study have been made available on the internet (for example, at FigShare), the dataset can be cited in principle (see <https://datacite.org>). Thus, the work which has been invested for producing an interesting, complex or effortful dataset might result in receiving credit in the form of data citations. In some fields, sharing data has been shown to associate with more citations for a paper (Piwowar & Vision, 2013; Colavizza et al., 2020). The measurement of data citation impact can be supported by assigning DOIs to the dataset (as is done by FigShare, and others), to combinations of datasets (as done by the Global Biodiversity Information Facility, GBIF; see Khan, 2019), or by suggesting the format to be used in a reference list. The German National Library of Science and Technology (TIB) developed DOIs for datasets (Fecher & Friesike, 2014). With DataCite (see <https://datacite.org>), an institution exists with the goal to provide persistent identifiers for research.

At present, data citation is not widely practiced by authors overall (Robinson-García, Jiménez-Contreras, & Torres-Salinas, 2016) and in scientometrics. Unlike bibliographic sources that are cited and appear in reference sections of scholarly works, the data sources used in conducting research (including software) may not be granted the same level of acknowledgement. The prevalence of data citation also varies from one discipline to another (Zhao, Yan, & Li, 2018). When authors do acknowledge shared data they have reused, they

do not necessarily cite the data sources in a manner that allows a data citation indexing service such as Clarivate Analytics' Data Citation Index or DataCite to record instances of data reuse, thereby denying authors of the datasets formal credit for their contributions (Park, You, & Wolfram, 2018). Both DCI and DataCite are currently limited in their data repository coverage. Scientometricians who rely on citation data collected from these repositories are advised to be cautious about the conclusions they draw.

Scientometric studies seem to rarely generate datasets with general use that are separate from specific academic outputs. For example, all the datasets within the first 100 matches on Figshare for the query 'Scientometrics' seem to be associated with a paper (they almost always state this directly and associated papers can be found via Google for the exceptions), despite Figshare being a free open access repository supporting data sharing. In addition, many citation analyses use commercial data from the WoS or Scopus. Nevertheless, some open datasets have been used in scientometrics, such as the European Tertiary Education Register dataset (ETER; www.eter-project.com) and there are some datasets related to scientometrics on FigShare that are not associated with journal articles (e.g., https://figshare.com/articles/UK_university_Web_sites_June_July_2005/785775/1).

Thus, it seems that scientometric authors reusing data may prefer to cite the paper associated with the data rather than the data itself, which would explain the low rate of data citation (Robinson-García, Jiménez-Contreras, & Torres-Salinas, 2016). A search for data-related records (i.e., datasets, data studies, software) in Clarivate Analytics' Data Citation Index using the topic search "scientometr* OR bibliometr* OR informetr* OR altmetr*" on October 31, 2020 resulted in 2044 records across all disciplines for the period 2013 to 2020, and only 174 citations (2 records with 2 citations, 170 records with 1 citation) to these records. Even if a large percentage of the citations came from scientometric authors, this still indicates very limited data citation activity for metrics-related data. This contrasts with fields such as biodiversity, where organism prevalence data can be a primary research output, and so

data citations might be an important way to recognize the usefulness of a non-publishing scientist's work. In fields like genomics, however, data can be extremely time-consuming to collect and valuable, but may not be commonly shared (Thelwall, Munafo, Mas-Bleda, et al., 2020), which is needed to encourage reuse and ultimately data citation

There are many valuable uses of research data that do not lead to citations (Thelwall & Kousha, 2017), such as for verification and training. Hence, data citations reflect a possibly small proportion of the uses of shared research data. Also, as noted above, researchers in some fields are not regularly citing data sources in a way that allows these sources to be captured by data citation indexing services. Data citation is a relatively new development so that the tradition or expectation of citing datasets in a manner comparable to bibliographic sources is not yet commonly practiced (Park, You, and Wolfram, 2018).

4.5 Open review comments

Reviews, along with author rebuttals and editor comments, can be made publicly available (in anonymized or signed form; see Schmidt, Ross-Hellauer, van Edig, & Moylan, 2018). Signed open comments can be cited in principle. Thus, reviewers can receive citation impact for their contributions. Published reviews (in anonymized or signed form) might be an instrument of the journal to demonstrate that it cannot be categorized as a predatory journal. Predatory journals are OA journals that publish manuscripts for money (paid by the authors) without quality control (peer review) for what is published. Another benefit of open reviews is that the review process is no longer a black box. Readers can gain insight into how a publication came to be in its final form. Open reviews also can serve as an important learning tool for new scholars by providing exemplars of the review process. Currently, however, most scientists seem to believe that double-blind review is the most effective model for quality control (Moylan, Harold, O'Neill, & Kowalczyk, 2014; Mulligan, Hall, & Raphael, 2013; Rodriguez-Bravo et al., 2017).

No scientometric journal has published its peer review process (at the time of writing this paper). The website of the journal *Quantitative Science Studies* (QSS) indicates that reviewers may choose to identify themselves. QSS is currently running a transparent peer review pilot. When a manuscript participating in the pilot is accepted for publication in QSS, the reports of the reviewers, the responses of the authors, and the decision letters of the editor are made openly available in Publons (see <https://publons.com>). Participation in the pilot is voluntary (see https://www.mitpressjournals.org/journals/qss/peer_review). However, scientometricians can choose to publish in general journals or platforms that offer (complete) open peer review, if they believe that this is valuable.

The availability of review reports and reviewer identities may be optional or required by the journal policy. The journal *eLife* is publishing some ‘meta-research’ adopting (a kind of) open peer review process. Another example is the journal *Atmospheric Chemistry and Physics* (ACP) which was launched in 2001 and is freely accessible at <https://www.atmospheric-chemistry-and-physics.net> (publisher: Copernicus Publications). ACP has a two-stage publication process, with a peer review process which is different from processes use by traditional journals. The process is explained at https://www.atmospheric-chemistry-and-physics.net/peer_review/interactive_review_process.html.

The innovative peer review process at ACP has been evaluated by Bornmann, Marx, et al. (2010) and Bornmann, Schier, et al. (2011). The results of the study show that the process can reliably and validly select manuscripts for publication. A currently new example is the PLOS family of journals, for which open peer review became optional for authors in May 2019 (PLOS, 2019). Scientometric articles are also occasionally published in journals that reveal reviewer identities, but not the reviews, offering a different type of transparency (e.g., *Journal of Medical Internet Research*, see Eysenbach, 2011, or *Frontiers in Research Metrics and Analytics*). Similarly, the journal *PeerJ*, which also occasionally publishes scientometric

research, provides both authors the option to make their reviews available and reviewers the option to identify themselves. This journal started in 2013.

OA journals that provide open reports and/or reviewer identities make it possible for researchers to download or harvest data for scientometric and textual analysis. Because there is currently no standardized way that these journals currently provide access to open peer review data – which may be available in HTML, XML or PDF format – crawling and scraping routines need to be customized for each journal of interest, or at a minimum for each publisher. Current challenges include reviewer comment discovery and identification of review components. Reviewer comment discovery can be challenging if there is no standardized location for the reviews. Individual web pages must be crawled for review-related text.

Since standardization to provide access to open peer review data is ongoing, for instance by Publons and Crossref, data will become increasingly available.

5 Discussion and conclusions

A key element in the Open Science program is the demand for transparency (e.g., by preregistering a study or publishing the underlying data). This can help both replication/robustness and accessibility, if the final product is openly available. Transparency is needed in research evaluation both at the level of peer reviews and in terms of scientometric meta-evaluations. A number of studies have shown that bureaucracies are not necessarily able to identify the best performing researchers (Irvine & Martin, 1984; Bornmann, Leydesdorff, & van den Besselaar, 2010), although these studies have equated performance with bibliometric indicators rather than societal value or other types of impact. The process of priority programming and funding is continuously in need of legitimation.

In this opinion paper, we presented an overview of the Open Science program by discussing aspects from this movement which are most relevant for the scientometric field.

Many aspects of the Open Science movement have been triggered by specific scientific disciplines, for specific reasons. In this paper, we discuss several aspects towards their applicability for scientometrics. The outcomes of Open Science adoption might be, e.g., better reproducibility, better access, and more diversity. Although the Open Science program includes many interesting proposals that seem worth considering in scientometric research, there are also potentially problematic issues.

The developments which might result from the Open Science framework need to be scrutinized for both potentially positive and negative effects on science. For instance, publications in Gold OA journals enhance the accessibility of research to a wider audience and to practitioners and scholars in less affluent institutions. At the same time, the costs associated with publishing in OA journals arising from Article Publication Charges may be a barrier for researchers with limited financial resources and cannot always be resolved through APC waivers (e.g., in the case of retired scientists). Possible solutions include publishing in diamond (no fee) OA journals and/or making a pre-print version of the manuscript available through pre-print servers and institutional repositories, if permitted by the publisher's OA policies.

Authors may not be willing or feel obliged by funding agencies to expend the time and effort needed for research study preregistration and the sharing of data and software in formats that make them usable by other researchers (see Nosek et al., 2015; Nosek, Ebersole, DeHaven, & Mellor, 2018). The current reward system in science does not foster these activities explicitly. The 'steering of science', however, is a policy process which can be analyzed by means of policy analysis. The track record of science-policy interventions, however, is poor (van den Daele & Weingart, 1975). Institutional interests are always important in the background and inclusiveness and accessibility (e.g., for minorities) may be more important than transparency.

At the end of this opinion paper, there are some take home messages resulting from the issues discussed in sections 3 and 4. The target group for these messages are scientometricians conducting their own (empirical) studies.

6 Take home messages

- A scientometric study can be registered early based on a detailed research plan that can be later checked by reviewers, editors, and readers in order to make the formulation of hypotheses in hindsight more difficult and to demonstrate that hypotheses have been formulated before analyzing the data.
- Data used for a scientometric study can be made ‘open’ if this data is available under a license that allows free use, modification, and redistribution.
- The code or compiled software for analyzing scientometric data can be shared under an open source license.
- The contributions of authors to research projects and publications can be made publicly available in the process of publishing a scientometric paper.
- Authors can make scientometric papers or reports freely available to readers in traditional or Open Access journals. Where permitted by journal policies, pre-prints of articles can be self-archived or made available in public repositories to increase their availability.
- Users of ‘open’ scientometric datasets can give credit to the developers of the datasets in the form of data citations.
- Reviews of scientometric manuscripts, along with author rebuttals and editor comments, can be made publicly available, making the process of the refinement of the research reporting more transparent.

Acknowledgements

We thank Ludo Waltman and Loet Leydesdorff for the discussions of previous versions of the manuscript and detailed suggestions for improvements.

Competing interests

We have no competing interests to declare.

Funding information

L.B, M.T. and D.W. have received no funding for their research. The work of R.G. was supported by the Flemish Government through its funding of the Flemish Centre for R&D Monitoring (ECOOM).

References

- Baker, M. (2016). Why scientists must share their research code. Retrieved February 18, 2020, from <https://www.nature.com/news/why-scientists-must-share-their-research-code-1.20504> doi: 10.1038/nature.2016.20504.
- Bauer, P. C. (2018). Writing a reproducible paper in R markdown (SSRN scholarly paper no. ID 3175518). Retrieved February 18, 2020, from <https://papers.ssrn.com/abstract=3175518>.
- Bornmann, L., & Daniel, H.-D. (2005). Does the *h*-index for ranking of scientists really work? *Scientometrics*, 65(3), 391-392. doi: 10.1007/s11192-005-0281-4.
- Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, 66(11), 2215-2222. doi: 10.1002/asi.23329.
- Bornmann, L., & Mutz, R. (2015). How well does a university perform in comparison with its peers? The use of odds, and odds ratios, for the comparison of institutional citation impact using the Leiden Rankings. *Journal of the Association for Information Science and Technology*, 66(12), 2711-2713. doi: 10.1002/asi.23451.
- Bornmann, L., Leydesdorff, L., & van den Besselaar, P. (2010). A meta-evaluation of scientific research proposals: Different ways of comparing rejected to awarded applications. *Journal of Informetrics*, 4(3), 211-220. doi: 10.1016/j.joi.2009.10.004.
- Bornmann, L., Stefaner, M., de Moya Anegon, F., & Mutz, R. (2014). What is the effect of country-specific characteristics on the research performance of scientific institutions? Using multi-level statistical models to rank and map universities and research-focused institutions worldwide. *Journal of Informetrics*, 8(3), 581-593. doi: 10.1016/j.joi.2014.04.008.
- Bornmann, L., Mutz, R., Hug, S., & Daniel, H. (2011). A multilevel meta-analysis of studies reporting correlations between the *h* index and 37 different *h* index variants. *Journal of Informetrics*, 5(3), 346-359. doi: 10.1016/j.joi.2011.01.006.
- Bornmann, L., Schier, H., Marx, W. and Daniel, H.-D. (2011), Is interactive open access publishing able to identify high-impact submissions? A study on the predictive validity of *Atmospheric Chemistry and Physics* by using percentile rank classes. *Journal of the Association for Information Science and Technology*, 62, 61-71. doi:10.1002/asi.21418.
- Bornmann, L., Marx, W., Schier, H., Thor, A., & Daniel, H.-D. (2010). From black box to white box at open access journals: Predictive validity of manuscript reviewing and editorial decisions at *Atmospheric Chemistry and Physics*, *Research Evaluation*, 19(2), 105-118. doi: 10.3152/095820210X510089.
- Bourne, P. E., Clark, T. W., Dale, R., De Waard, A., Herman, I., Hovy, E. H., & Shotton, D. (2012). Improving the future of research communications and e-scholarship (Dagstuhl Perspectives Workshop 11331). doi: 10.4230/DAGMAN.1.1.41.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49(12), 997-1003.
- Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PLOS ONE*, 15(4), e0230416. doi: 10.1371/journal.pone.0230416.
- Cumming, G., & Calin-Jageman, R. (2016). *Introduction to the new statistics: Estimation, open science, and beyond*. Milton Park, UK: Taylor & Francis.
- van den Daele, W., & Weingart, P. (1975). Resistenz und Rezeptivität der Wissenschaft zu den Entstehungsbedingungen neuer Disziplinen durch wissenschaftliche und politische Steuerung. *Zeitschrift für Soziologie*, 4(2), 146-164.

- Egghe, L., & Rousseau, R. (2006). An informetric model for the Hirsch-index. *Scientometrics*, 69(1), 121-129.
- Elzinga, A. (1976). *Notes on the life and works of Christiaan Huygens (1629-1695)*, Gothenburg, Sweden: Department of Theory of Science.
- Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research*, 13(4), e123.
- Fecher, B., & Friesike, S. (2014). Open science: One term, five schools of thought. In S. Bartling & S. Friesike (Eds.), *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing* (pp. 17-47). Cham, Switzerland: Springer International Publishing.
- Frenken, K., Heimeriks, G. J., & Hoekman, J. (2017). What drives university research performance? An analysis using the CWTS Leiden Ranking data. *Journal of Informetrics*, 11(3), 859-872. doi: 10.1016/j.joi.2017.06.006.
- Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., . . . Barabási, A.-L. (2018). Science of science. *Science*, 359(6379). doi: 10.1126/science.aao0185.
- García-Peñalvo Francisco, J., García de Figuerola, C., & Merlo José, A. (2010). Open knowledge: Challenges and facts. *Online Information Review*, 34(4), 520-539. doi: 10.1108/14684521011072963.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis. *Educational Researcher*, 5, 3-8.
- Gold, E. R., Ali-Khan, S. E., Allen, L., Ballell, L., Barral-Netto, M., Carr, D., & Cook-Deegan, R. (2019). An open toolkit for tracking open science partnership implementation and impact. *Gates Open Research*, 3, 1442. doi: 10.12688/gatesopenres.12958.2.
- Guédon, J. C., Kramer, B., Laakso, M., Schmidt, B., Šimukovič, E., Hansen, J. . . . Patterson, M. (2019). Future of scholarly publishing and scholarly communication: Report of the expert group to the European Commission. Retrieved July 20, 2020, from <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1098&context=scholcom>
- Hand, D. J. (2014). *The improbability principle: Why coincidences, miracles, and rare events happen every day*. New York, NY, USA: Farrar, Straus and Giroux.
- Hecking, T., & Leydesdorff, L. (2019). Can topic models be used in research evaluations? Reproducibility, validity, and reliability when compared with semantic maps. *Research Evaluation*, 28(3), 263-272.
- Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, 15, 135-175.
- Huygens, Christiaan; Blackwell, Richard J., trans. (1986 [1673]). *Horologium oscillatorium (the pendulum clock, or geometrical demonstrations concerning the motion of pendula as applied to clocks)*. Ames, Iowa: Iowa State University Press.
- Irvine, J., & Martin, B. R. (1984). *Foresight in science: Picking the winners*. London, UK: Frances Pinter.
- Khan, N. (2019). Does data sharing influence data reuse in biodiversity? A citation analysis. Retrieved February 18, 2020, from <http://www.diva-portal.org/smash/get/diva2:1267704/FULLTEXT01.pdf>.
- Kupferschmidt, K. (2018). More and more scientists are preregistering their studies. Should you? Retrieved October 16, 2019, from <https://www.sciencemag.org/news/2018/09/more-and-more-scientists-are-preregistering-their-studies-should-you>.
- Latour, B. (1987). *Science in action*. Milton Keynes, UK: Open University Press.
- Latour, B., & Woolgar, S. (1979). *Laboratory life: The social construction of scientific facts*. Thousand Oaks, CA, USA: Sage.

- Leydesdorff, L., Bornmann, L., & Mingers, J. (2019). Statistical significance and effect sizes of differences among research universities at the level of nations and worldwide based on the Leiden rankings. *Journal of the Association for Information Science and Technology*, 70(5), 509-525. doi: 10.1002/asi.24130.
- Machiavelli, N. (1532). *Il principe*. Retrieved February 18, 2020, from http://www.ibiblio.org/ml/libri/m/MachiavelliNB_IlPrincipe/main_h.html.
- Marewski, J. N., & Bornmann, L. (2018). *Opium in science and society: Numbers*. Retrieved February 18, 2020, from <https://arxiv.org/abs/1804.11210>.
- Marx, W. (2014). The Shockley-Queisser paper – A notable example of a scientific sleeping beauty. *Annalen der Physik*, 526(5-6), A41-A45. doi: 10.1002/andp.201400806.
- Merton, R. K. (1942). Science and technology in a democratic order. *Journal of Legal and Political Sociology*, 1, 115-126.
- Moylan, E. C., Harold, S., O'Neill, C., & Kowalczyk, M. K. (2014). Open, single-blind, double-blind: Which peer review process do you prefer? *BMC Pharmacology & Toxicology*, 15. doi: 10.1186/2050-6511-15-55.
- Mulligan, A., Hall, L., & Raphael, E. (2013). Peer review in a changing world: An international study measuring the attitudes of researchers. *Journal of the American Society for Information Science and Technology*, 64(1), 132-161. doi: 10.1002/asi.22798.
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America*, 115(11), 2600-2606. doi: 10.1073/pnas.1708274114.
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., . . . Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422-1425. doi: 10.1126/science.aab2374.
- Park, H., You, S., & Wolfram, D. (2018). Informal data citation for data sharing and reuse is more common than formal data citation in biomedical fields. *Journal of the Association for Information Science and Technology*, 69(11), 1346-1354. doi: 10.1002/asi.24049.
- Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175.
- PLOS (2019). PLOS journals now open for published peer review. Retrieved February 18, 2020, from <https://blogs.plos.org/plos/2019/05/plos-journals-now-open-for-published-peer-review>.
- Popper, K. R. (1967). *The Poverty of Historicism*. London, UK: Routledge and Kegan Paul
- Robinson-García, N., Jiménez-Contreras, E., & Torres-Salinas, D. (2016). Analyzing data citation practices using the data citation index. *Journal of the Association for Information Science and Technology*, 67(12), 2964-2975.
- Rushforth, A., & de Rijcke, S. (2015). Accounting for impact? The journal impact factor and the making of biomedical research in the Netherlands. *Minerva*, 53(2), 117-139. doi: 10.1007/s11024-015-9274-5.
- Schmidt, B., Ross-Hellauer, T., van Edig, X., & Moylan, E. (2018). Ten considerations for open peer review [version 1; peer review: 2 approved]. *F1000Research*, 7(969). doi: 10.12688/f1000research.15334.1.
- Spellman, B.A., Gilbert, E.A. and Corker, K.S. (2019). *Open science*. In J. T. Wixted (Ed.), *Stevens' handbook of experimental psychology and cognitive neuroscience*. doi:10.1002/9781119170174.epcn519.
- Tabah, A. N. (1999). Literature dynamics: Studies on growth, diffusion, and epidemics. *Annual Review of Information Science and Technology*, 34, 249-286.
- Tennant, J., Agarwal, R., Baždarić, K., Brassard, D., Crick, T., Dunleavy, D. J., Evans, T. R., Gardner, N., Gonzalez-Marquez, M., Graziotin, D., Greshake Tzovaras, B.,

- Gunnarsson, D., Havemann, J., Hosseini, M., Katz, D. S., Knöchelmann, M., Madan, C. R., Manghi, P., Marocchino, A. . . . Yarkoni, T. (2020). A tale of two ‘opens’: Intersections between free and open source software and open scholarship. Retrieved July 30, 2020, from <https://doi.org/10.31235/osf.io/2kxq8>
- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., ... & Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLOS ONE*, *10*(8), e0134826.
- Thelwall, M., & Kousha, K. (2017). Do journal data sharing mandates work? Life sciences evidence from Dryad. *Aslib Journal of Information Management*, *69*(1), 36-45.
- Thelwall, M. & Mas-Bleda, A. (2020). How common are explicit research questions in journal articles? *Quantitative Science Studies*, *1*(2), 730-748.
- Thelwall, M., Munafò, M., Mas-Bleda, A., Stuart, E., Makita, M., Weigert, V., & Kousha, K. (2020). Is useful research data usually shared? An investigation of genome-wide association study summary statistics. *PLOS ONE*, *15*(2), e0229578. doi: 10.1371/journal.pone.0229578.
- Thor, A., & Bornmann, L. (2011). The calculation of the single publication h index and related performance measures: A web application based on Google Scholar data. *Online Information Review*, *35*(2), 291-300. doi: 10.1108/14684521111128050.
- Tozzi, C. (2017). *For fun and profit: A history of the free and open source software revolution*. Cambridge, MA, USA: MIT Press.
- Velden, T., Hinze, S., Scharnhorst, A., Schneider, J.W., & Waltman, L. (2018). Exploration of reproducibility issues in scientometric research. In P. Wouters, *Proceedings of the 23rd International Conference on Science and Technology Indicators* (pp. 612-624). Leiden University: Leiden, the Netherlands. From <https://openaccess.leidenuniv.nl/handle/1887/65315>.
- Van Noorden, R. (2013). Open Access: The true cost of science publishing. *Nature*, *495*(7442), 426-429. doi: 10.1038/495426a.
- Visser, M., van Eck, N. J., & Waltman, L. (2020). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. Retrieved May 22, 2020, from <https://arxiv.org/abs/2005.10732>.
- Waltman, L. (2019). From *Journal of Informetrics* to *Quantitative Science Studies*. Retrieved March 9, 2020, from <http://www.issi-society.org/blog/posts/2019/april/from-journal-of-informetrics-to-quantitative-science-studies>.
- Watson, M. (2015). When will ‘open science’ become simply ‘science’? *Genome Biology*, *16*(1), 101. doi: 10.1186/s13059-015-0669-2.
- Whitley, R. D. (1984). *The intellectual and social organization of the sciences*. Oxford, UK: Oxford University Press.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., . . . Mons, B. (2016). Comment: The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*. doi: 10.1038/sdata.2016.18.
- Willinsky, J. (2005). Open journal systems: An example of open source software for journal management and publishing. *Library Hi Tech*, *23*(4), 504-519.
- Wouters, P., Hellsten, I., & Leydesdorff, L. (2004). Internet time and the reliability of search engines. *First Monday*, *9*(10), at http://www.firstmonday.org/issues/issue9_10/wouters/index.html.
- Zhao, M. N., Yan, E. J., & Li, K. (2018). Data set mentions and citations: A content analysis of full-text publications. *Journal of the Association for Information Science and Technology*, *69*(1), 32-46. doi: 10.1002/asi.23919.
- Ziman, J. (2000). *Real science: What it is and what it means*. Cambridge, UK: Cambridge University Press.