

Article

Quantification 2.0? Bibliometric Infrastructures in Academic Evaluation

Anne K. Krüger^{1,2}¹ Berlin Brandenburg Academy of Sciences and Humanities, 10117 Berlin, Germany; E-Mail: anne.krueger@bbaw.de² Department of Science Studies, Humboldt University Berlin, 10099 Berlin, Germany

Submitted: 22 October 2019 | Accepted: 3 February 2020 | Published: 9 April 2020

Abstract

Due to developments recently termed as ‘audit,’ ‘evaluation,’ or ‘metric society,’ universities have become subject to ratings and rankings and researchers are evaluated according to standardized quantitative indicators such as their publication output and their personal citation scores. Yet, this development is not only based on the rise of new public management and ideas on ‘the return on public or private investment.’ It has also profited from ongoing technological developments. Due to a massive increase in digital publishing corresponding with the growing availability of related data bibliometric infrastructures for evaluating science are continuously becoming more differentiated and elaborate. They allow for new ways of using bibliometric data through various easily applicable tools. Furthermore, they also produce new quantities of data due to new possibilities in following the digital traces of scientific publications. In this article, I discuss this development as quantification 2.0. The rise of digital infrastructures for publishing, indexing, and managing scientific publications has not only made bibliometric data become a valuable source for performance assessment. It has triggered an unprecedented growth in bibliometric data production turning freely accessible data about scientific work into edited databases and producing competition for its users. The production of bibliometric data has thus become decoupled from their application. Bibliometric data have turned into a self-serving end while their providers are constantly seeking for new tools to make use of them.

Keywords

bibliometrics; big data; digital infrastructures; higher education; quantification

Issue

This article is part of the issue “Quantifying Higher Education: Governing Universities and Academics by Numbers” edited by Maarten Hillebrandt (University of Helsinki, Finland) and Michael Huber (University of Bielefeld, Germany).

© 2020 by the author; licensee Cogitatio (Lisbon, Portugal). This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

1. Introduction

Current observations—discussed as the ‘audit’ (Power, 1999), ‘evaluation’ (Dahler-Larsen, 2012), or ‘metric society’ (Mau, 2019)—indicate a lack of societal trust in the performance of public organizations and the individuals working in them. Power describes this “audit explosion” as a need for more control through constant performance measurement that is based on “a certain set of attitudes or cultural commitments to problem solving” (Power, 1999, p. 4) which have been transferred from private companies to public organizations—such as universities. With the rise of new public management and economic calculations about the return on public (see Schimank, 2005) or—in the case of the US (see Espeland & Sauder, 2016)—private investment into science and

higher education, universities have become subject to internal evaluations (see Hillebrandt, 2020; Huber, 2020; Matthies & Simon, 2008) and external ratings and rankings (see Brankovic, Ringel, & Werron, 2018; Hazelkorn, 2011; Espeland & Sauder, 2016) according to their performance in research and increasingly also in teaching (see Times Higher Education, 2018). Consequently, this has also led to an increasing evaluation of the performance of individual researchers building on standardized quantitative indicators such as their publication output and their personal citation scores (see de Rijcke, Wouters, Rushforth, Franssen, & Hammarfelt, 2016; Waltman, van Eck, Visser, & Wouters, 2016).

Yet, this development towards performance measurement as problem solving is not only based on a shift in political attitudes and societal orders of worth but has

profited from ongoing technological developments that provide new ways of producing and assessing data about performances. The aim of the article therefore is to highlight the significant changes in data production and assessment that can be witnessed, in particular, in the context of bibliometric research evaluation. Already in 1964, Eugene Garfield started the Science Citation Index that was based on automated data processing (with handwritten punch cards) and the use of the first IBM computers (see Wouters, 1999, pp. 26–27). Due to a massive increase in digital publishing and the growing availability of publication metadata such as author names, reference lists, author affiliations, or funding organizations, such bibliometric infrastructures are increasingly becoming more differentiated and elaborate. They allow for new ways of using these data through various easily applicable tools. Furthermore, bibliometric infrastructures also produce new quantities of data due to new possibilities in following the digital traces scientific publications leave when people within or outside of academia engage with them, for instance, by viewing an article or downloading it from a journal website. The counts of such metadata about article usage are discussed as alternative metrics or ‘altmetrics’ (see Franzen, 2015; Haustein, Sugimoto, & Larivière, 2015).

In this article, I suggest to address these new developments from a critical data perspective (see boyd & Crawford, 2012) to highlight how new ways of data production affect questions of data application and usage. First, I provide some insights into the development of bibliometrics from research information to research evaluation. I highlight how the “competition for expertise” (de Rijcke & Rushforth, 2015, p. 1955) on research evaluation has brought new providers of bibliometric infrastructures onto the stage. Second, I discuss the increasing possibilities of bibliometric infrastructures from a sociomaterial perspective (see Orlikowski & Scott, 2008). I argue that the assemblage of infrastructures, their providers, and their users influences how science and research practice are understood and thus how ‘research performance’ is measured. And third, drawing on insights from critical data studies, I suggest discussing these developments as quantification 2.0. I argue that in applying these infrastructures, it has not only become a standardized practice to turn qualitative characteristics into quantitative metrics by making different things such as the individual work of researchers commensurable (see Espeland & Stevens, 2008). Moreover, I highlight that the rise of digital infrastructures for publishing, indexing, and managing scientific publications has triggered an unprecedented growth in bibliometric data production turning freely accessible data about scientific work into edited databases. Bibliometric data production has thereby become decoupled from questions of utility and usability. Instead, bibliometric data has turned into a self-serving end while their providers are constantly seeking for new tools to make use of them.

2. Quantifying Science as ‘Research Performance’

In 2008, Wendy Espeland and Mitchell Stevens called for a “sociology of quantification” because they witnessed “the spread of quantification” defined as “the production and communication of numbers” and “the significance of new regimes of measurement” (Espeland & Stevens, 2008, p. 402; for a detailed overview on the different lines of study on quantification, see Mennicken & Espeland, 2019). University ratings and rankings either for distributing public money or for informing prospective students have played a significant role in transforming “qualities into quantities” and “difference into magnitude” by reducing and simplifying “disparate information into numbers that can be easily compared” (Espeland & Stevens, 1998, p. 316). One of the commonly used indicators of such rankings is the publication output and the respective citation score (see Hazelkorn, 2011, pp. 32–37; Taubert, 2013). Even in the Times Higher Education Europe Teaching Ranking, which was released for the first time in 2018, “papers-to-staff ratio” adds to the overall ranking result of the teaching quality of a university by 7.5 percent (Times Higher Education, 2018, p. 29).

Yet, first ideas behind the collection of metadata of journal articles such as author names and references were not about research evaluation but about research information (for an encompassing history of bibliometrics, see Wouters, 1999). As a reaction towards the profound increase of scientific publications (see Wouters, 1999, pp. 59–60), in the 1960s, Eugene Garfield started to build up the Science Citation Index with a public grant funded by the US–American National Institutes of Health. The goal behind the Science Citation Index was to facilitate the search for publications in medicine and natural sciences among researchers. It was thus designed as a tool for enabling researchers to collect information about latest work for keeping track with new developments.

When information scientists like Derek de Solla Price (1963) but also Eugene Garfield himself (1979) recognized that bibliometrics provided possibilities for studying science as such they began to develop this field of research further. Based on the Science Citation Index, they started to explore the development of the science system and within particular disciplinary fields. As Taubert describes it:

Not the single citation but instead the emerging patterns from the analysis of masses of citations were of interest that allowed for insights into the importance of particular research groups, institutions, national research systems or the dynamic of research fields. (Taubert, 2013, p. 183, translation by the authors)

The search for patterns furthermore allowed identifying the journals that were deemed most relevant in a particular research field (see Garfield, 2007). This information was summarized in the *Journal Citation Report* which was

designed for helping librarians in choosing the journals that scientists cite the most.

Initially designed as a tool for researchers and librarians and for research about science, bibliometrics furthermore began to feature more and more prominently in science policy. Bibliometric calculations became a relevant 'judgment device' (Karpik, 2010) for evaluating research organizations and individual researchers and, consequentially, for allocating resources within the science system. De Rijcke and Rushforth argue that:

[The bibliometric] field had managed to create a demand for their measures—not only by supplying on-demand data and data-handling techniques but also by making sure their products were promoted as policy relevant information that decision makers could use strategically. (2015, p. 1955)

Publication output and citation scores thus turned into indicators for measuring research performance with different intensity according to the national and/or regional research system.

However, since the 1990s, concerns have constantly grown among the bibliometrics community about how these data are used calling for 'responsible metrics' (see, for an overview, Aksnes, Langfeldt, & Wouters, 2019; Ràfols, 2019; Wilsdon, 2015; Wilsdon et al., 2017). The Leiden Manifesto for research metrics (see Hicks, Wouters, Waltman, de Rijcke, & Rafols, 2015) is but one of the prominent examples (see also San Francisco Declaration on Research Assessment, 2013, or the Hong Kong Manifesto for Assessing Researchers, 2019) of how the bibliometrics community struggles with the problem that bibliometric indicators for research evaluation are no longer only used by experts from the bibliometrics community but often by research organizations and policy institutions themselves (see Hammarfelt & Rushforth, 2017; Leydesdorff, Wouters, & Bornmann, 2016).

3. The 'Competition for Expertise'

Despite critical debates within the bibliometrics community, "an increasingly intense globalized competition for expertise" (de Rijcke & Rushforth, 2015, p. 1955) on research evaluation has grown. The provision of bibliometric data for evaluation purposes has turned into "a crowded marketplace" (de Rijcke & Rushforth, 2015, p. 1956). The Institute of Scientific Information, which was founded by Garfield in 1960, is no longer the only provider of bibliometric data. Currently, there is an increasing number of databases that provide bibliometric information because the conditions for data production have changed. While Garfield and his colleagues still had to select and process data on journals, articles, authors and references manually (see Wouters, 1999), bibliometric data production has become facilitated through digital publishing and automated data collection. Due to these developments new databases have

developed. There are open initiatives such as Crossref, which seeks to provide information on links between publications, funders, preprints or datasets across different publishers. Furthermore, there are community-based subject-specific databases such as Astrophysics Data System or MathSciNet that operate as bibliographic databases for the search of relevant literature, but also include citation information. Yet, there is also a growing field of databases that is provided by commercial providers. Besides Web of Science, which was initiated by Garfield and is currently owned by Clarivate Analytics, commercial databases such as Scopus and Dimensions have emerged. They also belong to major companies, namely Elsevier and Digital Science. These companies do not only provide bibliometric databases, but furthermore own a large infrastructure of tools to generate information from these data about people, organizations or even entire countries regarding their research activities. However, there are significant differences between these databases.

Web of Science is a collection of databases that has emerged from the Science Citation Index and the Social Science Citation Index that were launched in 1964 and 1965 by the Institute of Scientific Information. From 1975 on, the Arts and Humanities Citation Index was added. Since 1990, Web of Science also contains the Conference Proceedings Citation Index and since 2005 it enlists also books in the Book Citation Index. Only recently, from 2015 onwards, Web of Science has also integrated the Emerging Sources Citation Index to "make content important to funders, key opinion leaders, and evaluators visible...even if it has not yet demonstrated citation impact on an international audience" (Clarivate Analytics, 2017a). Web of Science therefore comprises a plurality of datasets thereby attempting to capture different disciplines with their distinct styles of publishing research as journal articles, books, or conference proceedings.

Scopus was launched by the publisher Elsevier in 2004. In contrast to Web of Science, which indexes only a specific set of supposedly key journals in their field, Scopus attempts to index the largest number of peer-reviewed literature possible. Yet, only publications that have a Digital Object Identifier are included. Elsevier also claims to control the selected journals for their quality through the Content Selection & Advisory Board that is made up of "an international group of scientists, researchers and librarians who represent the major scientific disciplines" which "is comprised of 17 Subject Chairs, each representing a specific subject field" (Elsevier, 2020). Scopus consists of only one database that, however, contains information on items that range from journals, books, and conference proceedings to patents and trade publications from all common disciplines.

The Holtzbrinck Publishing Group, who holds the majority of shares of Springer Nature and thus owns—similar to Elsevier—more than 2,500 English language journals, has recently become a new player in the bibliometric field. They own the company Digital Science.

Digital Science describes itself as “a technology company serving the needs of scientific and research communities at key points along the full cycle of research” (Bode, Herzog, Hook, & McGrath, 2018). In 2018, Digital Science launched Dimensions as another bibliometric database that seeks to cover a broader range of sources than Web of Science or even Scopus. Dimensions is described as “transcend[ing] existing tools and databases” by “bringing together...grants, publications, clinical trials and patents, consistently linked and contextualized” (Bode et al., 2018, p. 9). Dimensions is thus becoming a new prominent player in bibliometric data production and assessment.

Yet, due to technological developments in data collection, other players beyond academia and academic publishers have entered the bibliometric scene. Sweeping the entire academic web from digitally available journals to books and even websites in all languages and all countries, Google uses its search engine not only for supporting the search for academic literature (see Martín-Martín, Orduña-Malea, Ayllón, & López-Cózar, 2015). With GoogleScholar, which was set up in 2004, Google is furthermore able to extract and analyze the mentions of publications and its authors from all these sources. Disciplines like the social sciences and arts and humanities where coverage in the established databases like Web of Science or Scopus is still fragmentary and incomplete are thus discussed to have become more visible through it (see Bornmann, Thor, Marx, & Schier, 2016; Harzing & Alakangas, 2016). GoogleScholar is furthermore freely accessible. It has thereby become a convenient source for researchers and administrators to search for research but also for information on research performance.

The recent growth in the provision of bibliometric data thus highlights that there is already an ongoing competition for producing and providing data on academic publications from journal articles to conference reports as well as their analysis. Yet, while the coverage of these databases is continuously getting broader due to digitalization, quality issues arise that are prominently discussed in the bibliometrics community. The most prominent problem is that the providers of these databases give only very limited insights into how they collect and edit their data because this kind of information is their actual business secret. Dimensions and GoogleScholar claim to capture as much data as possible. This holds particular problems for the validity and reliability of their data. GoogleScholar, for instance, can only draw on publications that are available online and in contrast to Web of Science or Scopus it is impossible to acquire access to its database to get insights into which kinds of data are actually included. Prins, Costas, van Leeuwen, and Wouters (2016, p. 267) therefore highlight that “restrictions to the use of GoogleScholar are the intensive manual data handling and cleaning, necessary for a feasible and proper data collection” (see also Mingers & Meyer, 2017).

Web of Science and Scopus are instead more selective (see Stahlschmidt, Stephen, & Hinze, 2019, Chapter 3).

Besides scientific quality, they also apply more formal criteria for selecting their data sources. Peer review and “ethical publication practices” as well as bibliographic information in English language are important selection criteria (Clarivate Analytics, 2017b; Elsevier, n.d.-a; Testa, n.d.). Yet, these selection criteria still leave some room for interpretation. How decisions are actually made on choosing journals and indexing them within which discipline remains opaque. In the case of journals categorized in the Emerging Sources Citation Index in Web of Science, they can pass “an initial editorial evaluation and can continue to be considered for inclusion in products such as the Science Citation Index Expanded, the Social Science Citation Index, and the Arts and Humanities Citation Index, which have rigorous evaluation processes and selection criteria” (Clarivate Analytics, 2017a). What this evaluation looks like, however, is not openly discussed. For Scopus, Taşkin, Doğan, Akça, Şencan, and Akbulut have shown that journals indexed in this database in some cases fail to have information e.g., about their publication ethics, malpractice management, or editorial policies publicly available although this information comprises part of Scopus’ defined selection criteria (2015, as cited in Stahlschmidt et al., 2019). Journals might also be excluded from Scopus based on decisions made by the Scopus Content Selection and Advisory Board. These decisions on indexing journals have a significant impact on the results these databases produce. Despite attempts to integrate more journals and other publication formats these databases are “still less accurate for the social sciences and humanities than for other fields, and for certain regions such as Africa, Oceania and Central and South America” (Stahlschmidt et al., 2019, p. 10) due to an overrepresentation of English language publications and a predominant focus on journal publications (see also Mongeon & Paul-Hus, 2016).

4. Bibliometric Databases as Digital Infrastructures

In the bibliometrics community, these issues are mainly discussed as a methodological problem in terms of data quality and the construction of indicators. Yet, de Rijcke and Rushforth (2015) highlight that there is also an ‘implementation problem’ as the use of these data for evaluation purposes has already spread widely. Besides providing information for researchers, librarians, administrators, policy makers and funders, Elsevier also promotes Scopus as being used by influential ranking organizations such as Times Higher Education for their World University Rankings or the Shanghai Ranking Consultancy for the Best Chinese University Ranking Report (see Elsevier, n.d.-b). Clarivate Analytics supports the British Research Excellence Framework, Dimensions was recently used for the first time by the Nature Index 2019 annual tables (see Digital Science 2019), and GoogleScholar can already be used without any further restrictions by anybody interested in his or her personal metrics or the metrics of fellow researchers.

The bibliometrics community is therefore not the first reference any more for doing research evaluation as the providers themselves offer evaluations or evaluation tools (see also Jappe, Pithan, & Heinze, 2018; Petersohn & Heinze, 2018) that enable even ‘lay persons’ such as research managers and policy makers to do evaluations on their own. In particular, Clarivate Analytics, Elsevier, and Digital Science have built an encompassing digital infrastructure that produces and collects different sorts of data and metadata and processes and assesses them: Tools such as literature management systems like Menderley or EndNote facilitate the production of bibliometric data that do not only help to organize research literature but also provide information about the use of publications as well as formally correct and thus easily collectable citations. Furthermore, so called ‘research intelligence solutions’ are offered and promoted as enabling research managers from research institutions up to the policy level to gain, analyze, and also visualize information on the development of current research trends as well as on the performance of individual researchers, research groups, or research organizations. In addition, these tools are also designed to enable research managers to collect data about their own research organizations and to analyze and manage it.

Methodological problems in terms of data collection and assessment are still and have become even more a predominant issue as the use of such bibliometric infrastructures constantly spreads. Moreover, by collecting and managing data and facilitating the assessment of ‘research performance,’ these infrastructures play a performative role in creating what they seek to count and calculate, namely, a particular understanding of science and research practice. Hence, such digital infrastructures are never neutral but embody already in their design and the calculative models behind them a particular understanding about the world and about its users (see Krüger, Heßelmann, & Hartstein, in press; Mühlhoff, 2018). Already in 1999, Bowker and Star (1999, p. 230) have emphasized that “values, policies, and modes of practice become embedded in large information systems.” In a similar vein, Winner has argued that:

Machines, structures, and systems of modern material culture can be accurately judged not only for their contributions to efficiency and productivity, not merely for their positive and negative environmental side effects, but also for the ways in which they can embody specific forms of power and authority. (Winner, 1980, p. 121)

Building on these insights, Bowker, Elyachar, Mennicken, Miller, and Randa Nucho (2019, p. 1) highlight three important aspects of what they call “thinking infrastructures” defined as social or material infrastructures that “structure attention, shape decision-making and guide cognition.” They are “valuation regimes that constitute orders of worth” (Bowker et al., 2019, p. 4) through def-

initions of success and failure; they make objects and practices “visible and available...for possible interventions” (Bowker et al., 2019, p. 4) thereby “establishing a distinct conception of the objects and objectives” of governance (p. 5). Digital infrastructures thus influence the practices they are supposed to support or reflect. They are performative because “they change the very nature of what it is to do work, and what work will count as legitimate” (Bowker & Star, 1999, p. 239).

In their research agenda on the role of sociomateriality in organization research, Orlikowski and Scott (2008) furthermore argue to go beyond the dichotomy of infrastructures and its users. Instead, they highlight that “people and things only exist in relation to each other” (Orlikowski & Scott, 2008, p. 455). Following Callon (1986) and Latour (1987) with their idea on actor-networks and the notion of ‘performativity,’ they claim that “entities (whether humans or technologies) have no inherent properties, but acquire form, attributes, and capabilities through their interpenetration” (Orlikowski & Scott, 2008, p. 455). The authors therefore think of the social and the material as “inherently inseparable” (p. 456) because they conjointly enact what this sociomaterial assemblage is about. Pollock, Williams, and Procter (2003) have applied this theoretical lens to a study on the construction and implementation of enterprise resource planning systems at universities. They demonstrate how this infrastructure is developed through its providers according to the perceived needs of the users, while the universities simultaneously adapt to this infrastructure, its integrated standardized processes and inscribed ideas about working practices. In this regard, infrastructure and universities mutually constitute each other.

Applied to the case of bibliometric infrastructures, this theoretical perspective allows us to see, first, that the assemblage of infrastructures, their providers and users does not produce stable constructs. Instead, bibliometrics are performed either as research information, research on research, or as research evaluation. Second, focusing on the performativity of these assemblages enables us to analyze how an understanding of science and research practice as “research performance” is constructed while attempting to measure and describe it.

In their comparative study of Web of Science and Scopus, Stahlschmidt et al. (2019) were able to demonstrate how different bibliometric databases construct a particular understanding of science and research practice leading to differences in their results. Building on a sample of German publications indexed in both databases, they found a “database-specific valuation of these publications” (Stahlschmidt et al., 2019, p. 64). Publications from the economic sector or from research institutes with a focus on applied sciences got better citation scores in Scopus than in Web of Science, while, conversely, in Web of Science organizations scored better that had a focus on basic research. Stahlschmidt et al. (2019) explain these differences through the impact that the respective content of each database has on the

valuation of a specific publication. In valuating a publication, Web of Science and Scopus draw on this content for “relating a publication to a specific environment of similar publications. Due to differences in coverage, Web of Science and Scopus apply different environments to appraise the same publication” (Stahlschmidt et al., 2019, p. 64). They therefore find that:

Any differences in the valuation of the same content result from differences in the respective environment, i.e., the exclusive content. Hence a comparison of the diverging valuation of the same content does not inform on the content itself, but on the exclusive content causing any differences and therefore the databases themselves. (Stahlschmidt et al., 2019, p. 65)

Evaluation results thus depend on the assumptions about research that are inscribed in the databases as well as how users apply which database for which kind of evaluation purpose. Bibliometric infrastructures therefore have a particular impact on the production of knowledge about science as well as on scientific knowledge production as such. The assemblage of bibliometric infrastructures, their providers, and their users performs a particular understanding about the way research should be practiced and therefore effects which kind of research is consequently regarded as valuable.

5. Quantification 2.0

The growing abilities of bibliometric infrastructures due to digital publishing, however, do not only pose a problem in terms of performativity that influences data production and assessment. The digitalization of academic publishing has changed bibliometric data production as such. Data does not have to be produced manually anymore, but can be collected, processed, and furthermore assessed automatically. This has led to an increase in quantitative data production as such.

So far, research on evaluation has highlighted quantification in terms of “the production and communication of numbers” (Espeland & Stevens, 2008, p. 402; see also Heintz, 2010) that turn qualitative characteristics into quantitative metrics by making different things commensurable (see Espeland & Stevens, 1998). However, focusing on data production in bibliometrics shows that the production of data does not only follow from operationalizing qualitative differences in research performance in terms of quantitative output. Moreover, the competition for expertise highlights that these digital infrastructures contribute to what I suggest to call quantification 2.0: the decoupling of data production and data application.

In the literature on critical data studies phenomena such as ‘big data’ and ‘datafication’ and the automated use of data termed ‘algorithmization’ are already widely discussed. boyd and Crawford (2012, p. 663) define ‘big data’ as “a cultural, technological, and schol-

arly phenomenon” that results from “maximizing computation power and algorithmic accuracy to gather, analyze, link, and compare large data sets” and “to identify patterns in order to make economic, social, technical, and legal claims.” Similarly, Amoore and Piotukh (2015, p. 345) highlight that “the rise of big data witnesses a transformation in what can be collected or sampled as data, and how it can be rendered analyzable.” Yet, most importantly, boyd and Crawford (2012) claim that ‘big data’ rests on some kind of “mythology,” i.e., “the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy.” The accumulation and automated analysis of large amounts of data has thus attained high credibility in the provision and objectivation of “digitally recorded, machine processable, easily agglomerated, and highly mobile” (Sadowski, 2019, p. 4) information about the social world. Fourcade and Healy (2017, p. 9) therefore claim that “modern organizations follow an institutional data imperative to collect as much data as possible.” They argue that “it does not matter that the amounts collected may vastly exceed a firm’s imaginative reach or analytic grasp” (Fourcade & Healy, 2017). Data is thus, as Sadowski puts it, “very often collected without specific uses in mind” (Sadowski, 2019, p. 4). Instead, “the assumption is that it will eventually be useful, i.e., valuable” (Fourcade & Healy, 2017, p. 13). Thus, metaphors such as ‘data mining’ or data as the ‘new oil’ have already spread widely understanding data as a new form of capital. Sadowski highlights that “the imperative...is to constantly collect and circulate data by producing commodities that create more data and building infrastructure to manage data” (Sadowski, 2019, p. 4). Producing, collecting, and processing data has become an “intrinsic motivation” (p. 4).

In the case of bibliometric databases, such developments display in the massive growth of digital publishing and the technological advancements in bibliometric infrastructures allow for collecting, processing, and assessing large amounts of data through automated processes (see de Rijcke & Rushforth, 2015; Taubert, 2013). Citations can now be collected and analyzed much easier through specific tools. Processes of standardization, e.g., of author names or organizations through the introduction of ‘persistent identifiers,’ simplifies data collection because they allow for the exact attribution of publications to authors and research organizations. Usage-based metrics such as, e.g., article views and downloads from journal websites, can be counted and analyzed as giving insights into the reception of publications beyond citations (see Haustein, Bowman, & Costas, 2016). Bibliometric tools, furthermore, produce data themselves that can be collected and analyzed for evaluation purposes. Taubert refers, for instance, to online reference managers that produce metadata about the usage of publications from storing and bookmarking to annotations by the reader (see Taubert, 2013, p. 25).

Providers of bibliometric infrastructures such as Clarivate Analytics, Elsevier, or Digital Science thus constantly seek to collect bigger amounts of data and meta-data about publications with better data quality and to propose new ways of using and applying it. Companies such as Altmetric that is related to Digital Science or Plum Analytics, which belongs to Elsevier, have even started to collect data about the usage of scientific publications extending to mentions on Twitter and Facebook, among others (see Franzen, 2015). The valuation of bibliometric data furthermore displays in monetary transactions. When, in 1992, Thomson Reuters bought Web of Science from the Institute of Scientific Information, they paid \$210 million (see Jayapradeep & Jose, 2017). When they resold the product to Clarivate Analytics in 2016, they received \$3.55 billion for it (see Thomson Reuters, 2016). In addition, as already shown, the rise of new players such as Scopus, GoogleScholar, or recently Dimensions demonstrates that data on academic publications have become a valuable commodity. They capitalize on these data as they turn freely accessible and massively available data into edited databases. The problem of digital infrastructures is thus not only a question of data quality and of predefined assumptions inscribed in calculative models for data assessment. Moreover, the production of data has become a self-serving end as a profitable, privately owned and secretly kept business model while their providers are constantly seeking for new tools to make use of it.

6. Conclusion and Outlook

The possibilities of data production and assessment through bibliometric infrastructures have massively increased through digital publishing and automated data processing. In this article, I have argued that this does not only contribute to a quantification and evaluation of science and research practice as ‘research performance.’ Moreover, the rise of new providers and constantly more databases and related tools suggests what I have called quantification 2.0: the decoupling of data production and data application. Data production has become a self-serving end that is generating the development of new tools in search for a purpose. Political attitudes that understand evaluations as solving the problem of societal distrust in the performance of public organizations supported by increasingly complex digital infrastructures for data production and assessment have not only triggered a need for quantifying qualitative performances. Moreover, triggered through these developments, increasingly more quantifiable data are produced and processed which allow new ways to propose research evaluation to potential users.

From this, however, follows that even though or right because tools for bibliometric data production and assessment have become a business model for their providers, they have to address the needs of potential customers belonging to the science system. Their

providers thereby profit from what boyd and Crawford have called “mythology” in terms of the perceived objectivity of big data and automated data analysis (boyd & Crawford, 2012, p. 663). Yet, the functionalities of newly designed bibliometric infrastructures and their characteristics still need to be explained and justified in a way to attract potential customers who are situated in the science system either as researchers and research managers or as policy makers, publishers, librarians or funding bodies. The providers of bibliometric infrastructures therefore have to propagate an understanding of science and research practice that resonates with the perception of their users.

In their research on “the extended practice of global software development,” Campagnolo, Pollock, and Williams (2015) address the question how software development works when providers seek to enter new contexts with their products. They describe how software providers try to make sense of matters and customers that they do not yet know much about. They refer to the phenomenological concept of ‘appresentation’ as a way of taking “account of the needs of future customers and also of their current users of whom they have no direct knowledge” (Campagnolo et al., 2015, p. 150). The software providers thus design their products according to the imagined needs of potential customers they seek to address with it. In their study, Campagnolo et al. (2015) demonstrate how these appresentations affect the interactions between providers and customers in their search for a mutually shared understanding of the functionalities that the software needs to provide. In addition, they are able to show how the appresentation of the providers heavily influences how customers finally understand their own needs.

Regarding bibliometric infrastructures, the providers also have to work with an appresentation of anticipated needs of their customers when bringing new tools for data assessment on the market. These appresentations are already inscribed in the functionalities of bibliometric tools. Bibliometric infrastructures thus already display a particular understanding of science and research practice by enabling specific modes of observation and the presentation of respective results. Yet, this means simultaneously that the functionalities of bibliometric infrastructures also have to account for the ways in that their users understand ‘research performance.’ The appresentation of customer needs and how customers are supposed to be attracted therefore yields insights into the understandings of science and of research practice that are already inscribed in these bibliometric infrastructures and have a performative effect on the perception of their users.

It therefore seems promising for future research not only to keep track with recent technological developments, but, furthermore, to ask for the understandings of science and research practice that are displayed in the sociomaterial assemblage of infrastructures, their providers, and their users. We therefore need more re-

search, on the one hand, on the understandings of science and research practice that are inscribed in these infrastructures. On the other hand, we need more research about the application of these tools to understand how the mythology of ‘better data, better decisions’ affects how research is practiced and valued.

Acknowledgments

The author owes special thanks to Stephan Stahlschmidt, Sabrina Petersohn, Judith Hartstein, and Felicitas Heßelmann from the German Centre for Higher Education Research and Science Studies (DZHW) for critical debates and fruitful comments. The author also wants to thank the Academic Editors and the anonymous reviewers for their helpful comments on earlier versions of this article.

Conflict of Interests

The author declares no conflict of interests.

References

- Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *SAGE Open*, 9(1). <https://doi.org/10.1177/2158244019829575>
- Amoore, L., & Piotukh, V. (2015). Life beyond big data: Governing with little analytics. *Economy and Society*, 44(3), 341–366.
- Bode, C., Herzog, C., Hook, D., & McGrath, R. (2018). *Dimensions report: A guide to the dimensions data approach*. London: Digital Science and Research Solutions Inc. Retrieved from https://figshare.com/articles/A_Guide_to_the_Dimensions_Data_Approach/5783094
- Bornmann, L., Thor, A., Marx, W., & Schier, H. (2016). The application of bibliometrics to research evaluation in the humanities and social sciences: An exploratory study using normalized Google Scholar data for the publications of a research institute. *Journal of the Association for Information Science and Technology*, 67(11), 2778–2789.
- Bowker, G. C., Elyachar, J., Mennicken, A., Miller, P., & Randa Nucho, J. (2019). Introduction to thinking infrastructures. In M. Kornberger, G. Bowker, J. Elyachar, A. Mennicken, P. Miller, J. Randa Nucho, & N. Pollock (Eds.), *Thinking infrastructures* (pp. 1–13). Bingley: Emerald Group Publishing.
- Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.
- boyd, d., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5), 662–679.
- Brankovic, J., Ringel, L., & Werron, T. (2018). How rankings produce competition: The case of global university rankings. *Zeitschrift für Soziologie*, 47(4), 270–288.
- Callon, M. (1986). Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St. Brieuc Bay. In J. Law (Ed.), *Power, action, and belief: A new sociology of knowledge?* (pp. 196–233). London and Boston, MA: Routledge & Kegan Paul.
- Campagnolo, G. M., Pollock, N., & Williams, R. (2015). Technology as we do not know it: The extended practice of global software development. *Information and Organization*, 25(3), 150–159.
- Clarivate Analytics. (2017a). *Web of Science Core Collection: Emerging Sources Citation Index*. Philadelphia, PA: Clarivate Analytics. Retrieved from http://wokinfo.com/media/pdf/ESCI_Fact_Sheet.pdf
- Clarivate Analytics. (2017b). Evaluation criteria for Web of Science Core Collection journals. *Clarivate Analytics*. Retrieved from <https://clarivate.com/essays/evaluation-criteria-web-science-core-collection-journals>
- Dahler-Larsen, P. (2012). *The evaluation society*. Stanford, CA: Stanford University Press.
- de Rijcke, S., & Rushforth, A. (2015). To intervene or not to intervene; is that the question? On the role of scientometrics in research evaluation. *Journal of the Association for Information Science and Technology*, 66(9), 1954–1958.
- de Rijcke, S., Wouters, P., Rushforth, A., Franssen, T., & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use: A literature review. *Research Evaluation*, 25(2), 161–169.
- de Solla Price, D. (1963). *Little science, big science*. New York, NY: Columbia University Press.
- Digital Science. (2019, June 19). Dimensions tracks normalized ranking for the first time in the Nature Index annual tables. *Digital Science*. Retrieved from <https://www.digital-science.com/press-releases/dimensions-tracks-normalized-ranking-for-the-first-time-in-the-nature-index-annual-tables>
- Elsevier. (n.d.-a). Content policy and selection. *Elsevier*. Retrieved from <https://www.elsevier.com/solutions/scopus/how-scopus-works/content/content-policy-and-selection>
- Elsevier. (n.d.-b). Ranking organizations. *Elsevier*. Retrieved from <https://www.elsevier.com/solutions/scopus/who-uses-scopus/ranking-organizations>
- Elsevier. (2020). *Scopus content coverage guide*. Amsterdam: Elsevier. Retrieved from https://www.elsevier.com/__data/assets/pdf_file/0007/69451/Scopus_ContentCoverage_Guide_WEB.pdf
- Espeland, W., & Sauder, M. (2016). *Engines of anxiety: Academic rankings, reputation, and accountability*. New York, NY: Russell Sage Foundation.
- Espeland, W., & Stevens, M. (1998). Commensuration as a social process. *Annual Review of Sociology*, 24(1), 313–343.
- Espeland, W., & Stevens, M. (2008). A sociology of

- quantification. *Archives Europeennes de Sociologie*, 3, 401–436.
- Fourcade, M., & Healy, K. (2017). Seeing like a market. *Socio-Economic Review*, 15(1), 9–29.
- Franzen, M. (2015). Der Impact Faktor war gestern: Altmetrics und die Zukunft der Wissenschaft [The impact factor is old-fashioned: Altmetrics and the future of science]. *Soziale Welt*, 66(2), 225–242.
- Garfield, E. (1979). *Citation indexing: Its theory and application in science, technology and humanities*. New York, NY: John Wiley & Sons.
- Garfield, E. (2007). The evolution of the Science Citation Index. *International Microbiology*, 10, 65–69.
- Hammarfelt, B., & Rushforth, A. (2017). Indicators as judgment devices: An empirical study of citizen bibliometrics in research evaluation. *Research Evaluation*, 26(3), 169–180.
- Harzing, A.-W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2), 787–804.
- Haustein, S., Bowman, T., & Costas, R. (2016). Interpreting “altmetrics”: Viewing acts on social media through the lens of citation and social theories. In C. R. Sugimoto (Ed.), *Theories of informetrics and scholarly communication: A festschrift in honor of Blaise Cronin* (pp. 372–406). Berlin and Boston, MA: De Gruyter Mouton.
- Haustein, S., Sugimoto, C., & Larivière, V. (2015). Guest editorial: Social media in scholarly communication. *Aslib Journal of Information Management*, 67(3). <https://doi.org/10.1108/AJIM-03-2015-0047>
- Hazelkorn, E. (2011). *Rankings and the reshaping of higher education: The battle for world-class excellence*. Houndmills: Palgrave Macmillan.
- Heintz, B. (2010). Numerische Differenz: Überlegungen zu einer Soziologie des (quantitativen) Vergleichs [Numerical difference: Ideas on a sociology of (quantitative) comparison]. *Zeitschrift für Soziologie*, 39(3), 162–181.
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429–431.
- Hillebrandt, M. (2020). Keeping one’s shiny Mercedes in the garage: Why higher education quantification never really took off in Germany. *Politics and Governance*, 8(2), 48–57.
- Huber, M. (2020). Steered by numbers: How quantification differentiates the reform of a German university. *Politics and Governance*, 8(2), 26–35.
- Jappe, A., Pithan, D., & Heinze, T. (2018). Does bibliometric research confer legitimacy to research assessment practice? A sociological study of reputational control, 1972–2016. *PLoS ONE*, 13(6). <https://doi.org/10.1371/journal.pone.0199031>
- Jayapradeep, M., & Jose, B. (2017). Eugene Garfield: The ultimate impact factor. *Informatics Studies*, 4(2), 7–24.
- Karpik, L. (2010). *Valuing the unique: The economics of singularities*. Princeton, NJ: Princeton University Press.
- Krüger, A. K., Heßelmann, F., & Hartstein, J. (in press). Bewertung in und durch digitale Infrastrukturen [Evaluation in and through digital infrastructures]. In T. Peetz & F. Meier (Ed.), *Bewertung und Organisation* [Evaluation and organization]. Wiesbaden: VS Springer.
- Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.
- Leydesdorff, L., Wouters, P., & Bornmann, L. (2016). Professional and citizen bibliometrics: Complementarities and ambivalences in the development and use of indicators—a state-of-the-art report. *Scientometrics*, 109(3), 2129–2150.
- Martín-Martín, A., Orduña-Malea, E., Ayllón, J. M., & López-Cózar, E. (2015). *Does Google Scholar contain all highly cited documents (1950–2013)?* Retrieved from <https://arxiv.org/abs/1410.8464>
- Matthies, H., & Simon, D. (Ed.). (2008). *Wissenschaft unter Beobachtung* [Science under observation]. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Mau, S. (2019). *The metric society: On the quantification of the social*. Cambridge: Polity Press.
- Mennicken, A., & Espeland, W. (2019). What’s new with numbers? Sociological approaches to the study of quantification. *Annual Review of Sociology*, 45(1), 223–245.
- Mingers, J., & Meyer, M. (2017). Normalizing Google Scholar data for use in research evaluation. *Scientometrics*, 112(2), 1111–1121.
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106(1), 213–228.
- Mühlhoff, R. (2018). Digitale Entmündigung und User Experience Design: Wie digitale Geräte uns nudgen, tracken und zur Unwissenheit erziehen [Digital disenfranchisement: How digital devices nudge and track us and how they teach us to become ignorant]. *Leviathan*, 46(4), 551–574.
- Orlikowski, W., & Scott, S. (2008). Sociomateriality: Challenging the separation of technology, work and organization. *The Academy of Management Annals*, 2(1), 433–474.
- Petersohn, S., & Heinze, T. (2018). Professionalization of bibliometric research assessment: Insights from the history of the Leiden Centre for Science and Technology Studies (CWTS). *Science and Public Policy*, 45(4), 565–578.
- Pollock, N., Williams, R., & Procter, R. (2003). Fitting standard software packages to non-standard organizations: The ‘biography’ of an enterprise-wide system. *Technology Analysis & Strategic Management*, 15(3), 317–332.
- Power, M. (1999). *The audit society: Rituals of verification*. Oxford and New York, NY: Oxford University Press.

- Prins, A., Costas, R., van Leeuwen, T., & Wouters, P. (2016). Using Google Scholar in research evaluation of humanities and social science programs: A comparison with Web of Science data. *Research Evaluation*, 25(3), 264–270.
- Ràfols, I. (2019). S&T indicators in the wild: Contextualization and participation for responsible metrics. *Research Evaluation*, 28(1), 7–22.
- Sadowski, J. (2019). When data is capital. Datafication, accumulation, and extraction. *Big Data & Society*, 6(1). <https://doi.org/10.1177/2053951718820549>
- San Francisco Declaration on Research Assessment. (2013). *San Francisco Declaration on Research Assessment*. Retrieved from <http://www.ascb.org/dora>
- Schimank, U. (2005). ‘New Public Management’ and the academic profession: Reflections on the German situation. *Minerva*, 43(4), 361–376.
- Stahlschmidt, S., Stephen, D., & Hinze, S. (2019). Performance and structures of the German science system. In Expertenkommission Forschung und Innovation (EFI) (Eds.), *Studien zum deutschen Innovationssystem* [Studies on the German Innovation System]. Berlin: Expertenkommission Forschung und Innovation (EFI).
- Taşkin, Z., Doğan, G., Akça, S., Şencan, I., & Akbulut, M. (2015). *Does Scopus put its own journal selection criteria into practice?* Paper presented at the 15th International Conference of the International Society for Scientometrics and Informetrics, Istanbul, Turkey.
- Taubert, N. (2013). Bibliometrie als Forschungsevaluation: Zur Konstitution und Funktionslogik wechselseitiger Beobachtung zwischen Wissenschaft und Politik [Bibliometrics as research evaluation: The constitution and logic of mutual observation between science and politics]. In J.-H. Passoth & J. Wehner (Eds.), *Quoten, Kurven und Profile: Zur Vermessung der sozialen Welt* [Quotas, curves and profiles: Measuring the social world] (pp. 179–204). Wiesbaden: Springer.
- Testa, J. (n.d.). Journal selection process. *Clarivate Analytics*. Retrieved from <https://clarivate.com/essays/journal-selection-process>
- Thomson Reuters. (2016). *Thomson Reuters announces definitive agreement to sell its intellectual property and science business to Onex and Baring Asia for \$3.55 billion* [Press release]. Retrieved from <https://www.thomsonreuters.com/en/press-releases/2016/july/thomson-reuters-announces-definitive-agreement-to-sell-its-intellectual-property-science-business.html>
- Times Higher Education. (2018). Europe teaching rankings. *Times Higher Education*. Retrieved from <https://www.timeshighereducation.com/world-university-rankings/europe-teaching-rankings-2018-methodology>
- Waltman, L., van Eck, N., Visser, M., & Wouters, P. (2016). The elephant in the room: The problem of quantifying productivity in evaluative scientometrics. *Journal of Informetrics*, 10(2), 671–674.
- Wilsdon, J. (2015). *The metric tide: Independent review of the role of metrics in research assessment and management*. London: Sage Publications.
- Wilsdon, J., Bar-Ilan, J., Frodeman, R., Lex, E., Peters, I., & Wouters, P. (2017). *Next-generation metrics: Responsible metrics and evaluation for open science* (Report of the European Commission Expert Group on Altmetrics). Brussels: European Commission.
- Winner, L. (1980). Do artifacts have politics? *Daedalus*, 109(1), 121–136.
- Wouters, P. (1999). *The citation culture* (Doctoral dissertation). University of Amsterdam, Amsterdam, The Netherlands.

About the Author

Anne K. Krüger is Scientific Coordinator of the Interdisciplinary Research Group ‘Changing Universities and Their Social Environment: Consequences for Academic Freedom?’ at the Berlin Brandenburg Academy of Sciences and Humanities. Prior to this, she was Researcher at the Department of Science Studies at Humboldt University Berlin and leader of the research group ‘Evaluation Practices in Science and Higher Education’ at the German Centre for Higher Education Research and Science Studies (DZHW). In her current research, she focuses on evaluative information infrastructures and the sociology of (e)valuation.