



an open access  journal



Citation: Rushforth, A., & Hammarfelt, B. (2023). The rise of responsible metrics as a professional reform movement: A collective action frames account. *Quantitative Science Studies*, 4(4), 879–897. [https://doi.org/10.1162/qss\\_a\\_00280](https://doi.org/10.1162/qss_a_00280)

DOI:  
[https://doi.org/10.1162/qss\\_a\\_00280](https://doi.org/10.1162/qss_a_00280)

Peer Review:  
[https://www.webofscience.com/api/gateway/wos/peer-review/10.1162/qss\\_a\\_00280](https://www.webofscience.com/api/gateway/wos/peer-review/10.1162/qss_a_00280)

Supporting Information:  
[https://doi.org/10.1162/qss\\_a\\_00280](https://doi.org/10.1162/qss_a_00280)

Received: 9 December 2022  
Accepted: 26 September 2023

Corresponding Author:  
Alexander Rushforth  
[a.d.rushforth@cwts.leidenuniv.nl](mailto:a.d.rushforth@cwts.leidenuniv.nl)

Handling Editor:  
Vincent Larivière

Copyright: © 2023 Alexander Rushforth and Björn Hammarfelt. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



## REVIEW ARTICLE

# The rise of responsible metrics as a professional reform movement: A collective action frames account

Alexander Rushforth<sup>1</sup>  and Björn Hammarfelt<sup>2</sup> 

<sup>1</sup>Centre for Science and Technology Studies (CWTS), Leiden University, Leiden, The Netherlands

<sup>2</sup>Swedish School of Library and Information Science, University of Borås, Borås, Sweden

**Keywords:** evaluative bibliometrics, research assessment reforms, responsible metrics, responsible research assessment, social movements

## ABSTRACT

Recent years have seen a rise in awareness around “responsible metrics” and calls for research assessment reforms internationally. Yet within the field of quantitative science studies and in research policy contexts, concerns about the limitations of evaluative bibliometrics are almost as old as the tools themselves. Given that many of the concerns articulated in recent reform movements go back decades, why has momentum for change grown only in the past 10 years? In this paper, we draw on analytical insights from the sociology of social movements on *collective action frames* to chart the emergence, development, and expansion of “responsible metrics” as a *professional reform movement*. Through reviewing important texts that have shaped reform efforts, we argue that hitherto, three framings have underpinned the responsible metrics reform agenda: the *metrics skepticism* framing, the *professional-expert* framing, and the *reflexivity* framing. We suggest that although these three framings have coexisted within the responsible metrics movement to date, cohabitation between these framings may not last indefinitely, especially as the responsible metrics movement extends into wider research assessment reform movements.

## 1. INTRODUCTION

Recent years have seen a rise in discussions and campaigns around “responsible metrics” and wider calls for research assessment reforms. Since the 2010s, calls for change have accelerated, with reform champions calling on multiple research actors (researchers, universities, funders, policymakers, publishers, bibliometric producers, and so on) to alter their ways. Interventions, such as the San Francisco Declaration on Research Assessment (DORA) (DORA, 2013), the Leiden Manifesto (Hicks, Wouters et al., 2015), and the Metric Tide (Wilsdon, 2016), are considered important statements that have sought to steer (or “nudge”) research actors into reconsidering their relationship with quantitative performance indicators in research assessment.

More recently, umbrella labels such as “research assessment reform” and “responsible research assessment” have emerged, which aligns responsible metrics with other reform agendas, including open science, research integrity, and diversity and equity (Curry, de Rijcke et al., 2020; Wilsdon, 2021). Although the emerging assessment reform movement is transnational in its intended scope, it has only partially resonated among research actors internationally (Tijssen, 2020), and no doubt will not be on the radars of all research actors. Nonetheless,

this “responsibility landscape” (Davies & Horst, 2015) has expanded considerably over the past decade, with reform champions able to cite multiple initiatives and statements as signs of growing cross-national momentum. For example, the European Agreement on Reforming Research Assessment was published in 2022, setting out principles of “good” assessment practices that research-performing organizations can sign up to and pledge to implement (CoARA, 2022). In the Netherlands, a national program, Recognition and Rewards, was launched 2019, involving universities, the Netherlands Academy and funding agencies (VSNU, NFO et al., 2019). Comparable initiatives have also been launched in Finland (TJNK, 2020) and Norway (UiR, 2021), and multiple cross-national professional bodies, organizational networks, and (non-)governmental organizations have put out statements and principles supporting reform of academic research assessment cultures, supported by “responsible uses” of research metrics (CoARA, 2022; EC, 2017; EU, 2022; LERU, 2022; UNESCO, 2021).

The growing profile of international reform movements for responsible metrics and responsible research assessment more broadly has—with a small number of exceptions (Chen & Lin, 2017; Leckert, 2021; Petersohn, Biesenbender, & Thiedig, 2020; Tijssen, 2020; Wilsdon, 2017)—received little scholarly attention within science studies to date. With this article we aim to present the first (to our knowledge) empirical, sociologically informed account of responsible metrics as a *reform movement*, which has more recently extended into movements for wider assessment reforms. This is a contribution to what we hope will become a wider interest on contemporary assessment reforms in science studies, research on research, and research evaluation. We are interested in addressing the puzzle of how long-standing debates and controversies around evaluative bibliometrics have evolved from issues discussed in rather cloistered technical communities, or criticized or resisted by isolated evaluators and researchers, into a more widely recognized cause for collective international action.

We argue that this puzzle can be addressed with the help of sociological theorizing on *social movements*, particularly *professional reform movements*. Specifically, the concept of *collective action frames* from social movement theory will be used to describe how “responsible metrics” emerged and evolved as a reform movement cause. We will review and analyze what have become core statements and interventions within this movement (DORA, Leiden Manifesto, Metric Tide, and assessment reform statements and calls to action that they have inspired), unpacking how they sought to construct persuasive accounts of existing problems and suggest alternative ways of organizing academic assessments, with a view to mobilizing potential adherents and constituents, garnering bystander support, demobilizing antagonists, and inspiring and legitimating their campaigns (Benford & Snow, 2000).

Our first argument is that the relative absence of conspicuous, resonant, and well-packaged collective action frames prior to the 2010s is an important explanation for why long-standing concerns and calls for action around evaluative bibliometrics took so long to find a wider audience accepting of the issue’s importance and urgency. Second, we argue that the responsible metrics agenda has managed to grow and evolve while encompassing—and sometimes artfully combining—three collective action frames, each of which has tried to diagnose and offer solutions for problems associated with evaluative bibliometrics: *metrics skepticism framing* (critiquing the validity and the politics of numbers), *professional-expert framing* (offering technical solutions to reinstate the authority of evaluative bibliometric knowledge), and *reflexive framing* (producing “responsible citizens” willing and able to handle indicators with care and responsiveness). Each of these three framings is, we argue, responding to a crosscutting “master-framing,” which we label the *managerial-realist framing* of bibliometrics as rational, efficient, and objective tools for evaluating research performance, and plausible alternatives to peer review.

The paper is structured as follows. In the next section we will introduce further the core concepts and arguments from the sociology of social movements on collective action frames. We will then briefly outline our approach to reviewing assessment reform texts, before outlining in greater depth the master framing and the three framings that we argue have most informed and shaped the responsible metrics collective action frame. We then analyze how these framings have variously featured and interacted with one another in important texts associated with the responsible metrics movement (DORA, the Leiden Manifesto, Metric Tide Report), their receptions, and how, more recently, the responsible metrics movement has been extended and translated into broader reform agendas—“responsible research assessment”—that encompass values and concerns from parallel reform movements such as research integrity and open science. We end by reminding readers that reform movement framings are rarely settled permanently but are subject to ongoing events, negotiations, and contestations.

## 2. HOW SOCIAL MOVEMENTS PURSUE CHANGES

In sociological research, a social movement is “a persistent and organized effort involving the mobilization of large numbers of people to work together to either bring about what they believe to be beneficial social change or resist or reverse what they believe to be harmful social change” (DeFronzo & Gill, 2019, p. 27). Social movements are a broad label incorporating many different forms of collective action with varying aims, from religious movements and cultural movements to revolutionary or resistance movements. One of the most common forms of social movement is *reform movements*, which do not aim to overthrow an existing system or government, but nonetheless are dissatisfied enough to seek to achieve some significant economic, social, or political change to an existing order (DeFronzo & Gill, 2019).

The responsible metrics movement is not the only professional scientific reform movement to have emerged in recent times: Responsible research assessment, open science and research integrity are other notable examples (Brundage & Guston, 2019; Penders, 2022). Penders (2022, p. 3) suggests that recent scientific reform movements tend to “foreground values that never were strangers to science such as honesty, transparency and accountability ... yet seek[s] to embed them into changed or improved scientific processes and instruments.” These movements call on researchers to restore or reconfigure the “etiquette” of their practices and will often appeal to core professional values (e.g., fairness, quality, accountability) and a sense that the current science system has gone awry, is “severely disturbed,” or is perhaps even “broken” (Derksen & Field, 2022; Penders, 2022). Though promising wider societal benefits (by generating more socially responsive and sustainable academic research systems), the targeted agents of responsible metrics reforms are fellow professionals and research system actors, with little interest, let alone heated controversy, likely to be stirred in wider society—hence the name *professional reform movement* (Frickel, 2004b, p. 459).

Our analysis draws on *collective action frames* (Benford & Snow, 2000; Frickel, 2004a), a concept from social movements research, to articulate how the responsible metrics reform agenda has evolved as it has gathered momentum internationally. Collective action frames are a means of analyzing the course and character of social movements, and are our core focus here, rather than (say) resource mobilization or political opportunity processes. Note also that our aim is *not* to determine if the responsible metrics reform movement has helped to reconfigure research assessment practices on the ground in universities or funding agencies, a contentious issue for which there is mixed evidence (Chen & Lin, 2017; Pontika, Klebel et al., 2022; Rice, Raffoul et al., 2020).

Collective action frame accounts conceptualize social movement actors as “signifying agents actively engaged in the production and maintenance of meaning for constituents, antagonists, and bystanders or observers” (Benford & Snow, 2000, p. 613). This signifying work is approached through the term “framing,” a contentious process whereby a movement generates and promotes shared meanings (“interpretive frames”) about social conditions and practices “that not only differ from existing ones but that may also challenge them” (Benford & Snow, 2000, p. 614). The products of this framing activity are known as *collective action frames*.

Collective action frames are constructed through the core tasks of *diagnosis* (problem identification and attribution), and *prognosis* (proposed solutions, correctives, or ways of negotiating the problems). These socially constructed vocabularies appeal to varying motives for others to follow the movement, including urgency, severity, propriety, or some combination thereof. There is of course a moral dimension to such framing elements, which often draw distinctions between good and bad, and in- versus outgroups. As social movements often straddle complex multiorganizational, sometimes multi-institutional fields, where other social movements also come and go, intramovement and intermovement conflicts can occur (Benford & Snow, 2000). Processes of attributing causality, blame, or responsibility often generate tensions and conflict within a movement, and even where there is agreement on the source of a problem, consensus does not necessarily follow on the appropriate correctives.

Framing processes evolve over time, particularly as a social movement expands across organizational networks and encounters new actors and discourses. Forms of frame mutation include *frame amplification* (some issues come to be highlighted or accented as more important than others), *frame extension* (a social movement’s interests and frames extend beyond their original focus to incorporate issues and concerns that animate potential adherents and recruits) (Benford & Snow, 2000), and *frame translation* (whereby the issues and concerns of one social movement are presented as essential for other movements and interested actors to pass through in order to realize their own goals and interests) (Frickel, 2004a).

Together, this package of concepts provides a valuable means of accounting for how efforts to “responsibilize” bibliometrics have emerged and evolved over time. However, we do not seek to exaggerate the reach or successes of the responsible metrics movement’s accomplishments: On the contrary, collective action frames literature helps highlight the limitations of the responsible metrics movement’s framings in *resonating* with certain actors and mobilizing them into action. For example, collective action frame theorists also point to *counterframings*, which resist, challenge, or simply ignore framings projected by a social movement. As such, collective action frame concepts help to make visible some of the ongoing struggles and limitations that the responsible metrics reform movement seems likely to face going forward, within the movement and externally.

### 3. ANALYTICAL REVIEW APPROACH

This analytical review is based on desk-based research of publications, announcements, statements, reports, and websites pertaining to evaluative bibliometrics and research assessment reforms from the 1970s until today. Various source materials were gathered as evidence of efforts by various actors to intervene in debates and discussions around appropriate uses of bibliometrics for evaluation purposes, and in some instances to set out an explicit agenda for change. The materials were identified through our own prior knowledge of evaluative bibliometrics, further reading, and web searches, and discussions with colleagues and peers. This was an interpretive, rather than a rigid and systematic, approach to identifying texts,

guided by whether they added meaning to our understanding of this phenomenon. A list of the position papers, statements, reviews, and resources promoting assessment reforms that were covered in this review is provided in the [Appendix](#).

As researchers, we believe that data do not speak for themselves but need to be given meaning by analysts. Our analysis involved separate close readings of the selected texts, note taking, and shared online conversations and review sessions. We annotated texts with high-order concepts from collective action frames, such as diagnosis, prognosis, frame amplification, frame extension, frame translation, master-frame, and so on. This allowed us to map and chart the changing contours of the movement across texts over time. This was an iterative process that followed the constant comparative approach, involving moving backward and forward between the materials and our own readings, interpretations, discussions, and experiences as researchers to critically identify and sharpen the collective action frames. Our findings have been written up as a composite narrative across three sections: on the three collective actions frames underpinning the responsible metrics movement; how these feature in key texts that have been central to the growth of the movement; and how the responsible metrics movement and its frames have extended and translated into a broader assessment reform movement.

Although we attempt to analyze these texts as sociologists, we cannot claim to be completely “detached” or “outsider” observers. As individuals we have worked more-or-less closely with some of the authors of the research assessment reform texts reviewed in this article, while working at CWTS, Leiden University (see Competing interests statement). We, like all researchers, are products of our surroundings, meaning that our work comes out of particular times and circumstances. If we were not located where we were when some of these responsible metrics texts were produced, we perhaps would never have gained the curiosity or motivation to conduct this piece of research or formed some of the impressions that we did. Ultimately, we believe prior experiences, relations and knowledge are not weaknesses to be played down but an inbuilt feature of interpretive social science research. Our primary objective in this analytical review is to apply critically the sociology of collective action frames to prominent statements and texts that support and symbolize this reform movement. In presenting this rich and complex history, we aim to privilege originality and meaningfulness over completeness and replicability (Boell & Cecez-Kecmanovic, 2014).

#### **4. THE THREE FRAMES OF THE RESPONSIBLE METRICS MOVEMENT**

We present three ideal types of framing, which we claim have shaped key texts and interventions in the responsible metrics movement (and the later, expanded research assessment reform movement). The three framings set out distinct problems and attributions (“diagnosis” in the collective action frames lexicon) and solutions (“prognosis”) in relation to evaluative bibliometrics. Each prognosis proposes alternative values and practices as solutions to the evaluative “bibliometrics problem” (variously conceived). Each framing consists of clusters of ideas and values recurrently used together to help construct problems and solutions. The proposed solutions are rarely presented as being qualitatively new, but instead, as values already present in practice (at least in part) but marginalized or threatened by the current regime (c.f. Penders (2022) on the “rediscovery” narratives of contemporary science reform movements). In the texts we analyzed, two or more of these ideal types often coexisted.

Before outlining substantial differences between the three framings that cohabit under the “responsible metrics” umbrella, we argue that each has emerged over time in response to a master-framing: the managerial-realist framing.

#### 4.1. The Managerial-Realist Framing of Evaluative Bibliometrics

In what we have termed the managerial-realist master frame, bibliometrics are projected as suitable and effective proxies for measuring research quality and impact—be they of individuals, groups, departments, institutes, research fields, or countries and regions. For accounts of the emergence of this way of thinking in 1970s science policy in select countries, see Chubin and Hackett (1990), Csiszar (2023), and Wouters (1999). We recognize this master-framing as part of a more general rational organizing myth in modern societies that evaluating and comparing quality and impact can be rationally and efficiently administered across disciplines and organizations, through standardized, objective measurement systems (Dahler-Larsen, 2012). Numbers carry an aura of precision and rationality in modern societies (Desrosières, 1998; Porter, 1996), and have been depicted by social theorists as a defining form of knowledge within Modernity (e.g., Poovey, 1998).

Evaluative bibliometrics have promised more transparent, equitable, efficient, and truthful windows into scientific quality and impact than peer review. We label this rational myth a *master-framing* because it is not movement-specific but rather a *generic frame* (Benford & Snow, 2000) that features widely across multiple policy contexts, more so following the rise of “New Public Management” styles of governing in many countries since the 1980s (Dahler-Larsen, 2012). The managerial-realist framing of evaluative bibliometrics thus belongs to a story that some managers, policymakers, and scientists have projected about how (or how they wish) evaluation processes and systems to proceed (Wouters, 1999). Although this framing has no doubt been advanced at times by zealous policymakers, bureaucrats, and academics, we see this framing as also being evoked strategically by its critics as a major, urgent threat to good research evaluation and healthy research systems.

We will now set out the three action frames—metric skepticism, professional-expert, and reflexive—that have shaped the responsible metrics as a reform movement, and consider how each of them seeks to construct and counter the managerial-realist frame.

#### 4.2. Metrics Skepticism Framing

The metrics skepticism framing encompasses many broad and specific concerns, which are united by being largely incredulous of the value of quantitative measurement of research performance, and through being critical of the politics and effects of bibliometrics on research systems. The specter of the managerial-realist framing has long troubled actors in research systems since the early introductions of evaluative bibliometrics in the 1970s and 1980s, even when the presence of bibliometrics was relatively small scale compared to today (see for example Collins, 1985). Weingart (2005) notes there have long been backlashes from within scientific communities to efforts to introduce evaluative bibliometrics, pointing to methodological objections and sociological conflicts about power and control of research evaluations, which he claims have been perennial features of how bibliometrics tools have been received by some academic researchers. Although multiple reasons have been advanced as to why quantitative performance indicators should not lead (or inform, even) research evaluations, such criticisms were not able to find (or latch onto) a concerted reform movement until the responsible metrics movement emerged in the 2010s. By this time, a number of critical social science research-informed critiques of bibliometrics and their effects on evaluation practices had emerged in response to the rise of international rankings and “audit explosion” of the 1990s and 2000s in many national research systems (Power, 1997).

Skepticism towards bibliometrics has often been informed by humanistic and social constructivist research philosophies, which privilege the language and methods of “social,”

“theory,” “political,” and “context” above “performance,” “measure,” and “results” (Kang & Evans, 2020). Research on audit cultures (Strathern, 2003), has been influential in arguing that in neoliberally oriented systems of new public management, quantitative indicators are used to exert forms of control over academic research and researchers’ working lives. Commonly articulated fears of bibliometrics’ adverse impacts on research systems include Goodhart’s Law (when a measure becomes a goal, it ceases to be a good measure), the McNamara fallacy (making decisions on quantitative information at the expense of other inputs) and the streetlight effect (bibliometrics measure what can be counted, not what counts). Critically oriented social science literature has claimed the expansion of performance measures has led to growing pressures on individuals, leading to burnout and unhealthy work cultures; the narrowing of research results and outputs at the expense of research diversity; and marginalization of certain research topics, groups, and career paths that do not satisfy dominant performance models (de Rijcke, Wouters et al., 2016). Critical readings of the streetlight effect depict bibliometrics as narrowing the policy and managerial view of research down to visible performance indicators that uphold the existing political economy, shining light only on topics already hegemonic in the science system and the economy (Råfols, 2019).

Though some within arts, humanities and social sciences have on occasion mobilized technical criticisms of bibliometrics to justify resistance, not least unsatisfactory database coverage of their research outputs (Franssen & Wouters, 2019), for the most part the object of critique tends to be quantification as a “logic” (e.g., Burrows, 2012; Shore & Wright, 2015). In such fields, quantitative performance indicators are often considered reductive and contrary to the richness, diversity, and multidimensionality of research quality (Franssen, 2022; Nästesjö, 2021), and providing flawed representations of what is said to be measured (Aksnes, Langfeldt, & Wouters, 2019).

Independently of critical traditions in social sciences and arts and humanities, prominent natural science spokespersons also later emerged to criticize and warn of the corrosive effects of quantitative performance indicators in their field (Alberts, Kirschner et al., 2014; Sample, 2013). The backlash against bibliometrics has thus come in many forms, from disparate professional communities. All have in common a preference for “qualitative” peer review and “expert judgement” over numbers, and position bibliometrics as risky objects that threaten the health and legitimacy of research systems. Skepticism towards the notion of bibliometrics having any value in assessing research has been an important framing input into debates and calls for responsible metrics.

#### **4.3. Professional-Expert Framing**

In response to the growing presence of—and backlashes towards—bibliometric tools in academic research systems, more concerted, focused debates emerged in the field of scientometrics in the 1990s (van Raan, 1996). As part of a “regulatory science” that both studies and produces quantitative indicators (Wouters, 1999), scientometricians saw it as their professional duty to educate wider science communities about these tools: to “set the record straight” against inflated managerial-realist claims of evaluative bibliometrics, while pushing back against full-blooded metric skepticism. The professional-expert framing that has resulted is a largely technical diagnosis and prognosis, with rigorous, scientific methods believed to hold the key for addressing the social problem of evaluative bibliometrics (for a critique of this framing, see de Rijcke and Rushforth [2015]). Expertise in the professional-expert framing is binary: There is the expert professional knowledge of scientometricians about the design and properties of bibliometric tools and their limits, and then there is expert epistemic knowledge

of evaluators about a given domain of scientific research under assessment (Moed, 2007; van Raan, 1998). It is the responsibility of evaluators, as users of bibliometric tools, to educate themselves about the uses and limitations of the tools. Terminology such as *mis-use* and *abuse* feature noticeably in these accounts—suggesting that a duty of care is being flaunted and denigrated by some evaluators when using bibliometric tools (Hammarfelt & Rushforth, 2017).

Arguably the most notable recent example of professional-expert responses to criticisms of bibliometrics is the notion of “metric wiseness.” According to this perspective “all scientists should become knowledgeable about indicators used to evaluate them. They should know the publication-citation context, the correct mathematical formulas of indicators being used by evaluating committees, their consequences and how such indicators can be misused” (Rousseau, Egghe, & Guns, 2018, p. 1). Here the problem of bibliometric evaluation is not attributed to the measures and indicators in themselves, but to a general lack of knowledge on how to use them. Therefore, what was called for was not a general questioning of the methods employed, but a call for researchers “becoming metric-wise” (Rousseau et al., 2018). Contrary to the metric skepticism framing, which suggests weakening the influence of bibliometrics, this counterframing emphasizes that bibliometrics can be legitimate tools in research assessments, as long as they are of the “correct” kind (for a recent defense of this position, see Bornmann, Botz, and Haunschild (2023)).

Criticisms of bibliometrics are not ignored in such accounts—it is acknowledged that indicators used in citation analysis “... are not perfect; but neither is peer review” (Rousseau et al., 2018, p. 7). So-called “informed peer review,” whereby quantitative indicators inform, but do not supplant, expert qualitative judgement, is presented as the optimal solution (Butler, 2007). Given this context, the terms “caution,” “proficiency,” and “competence” have become the watchwords of the professional-expert framing, more than the recently fashionable “responsibility” (Petersohn, 2021). Importantly, many of the problems of evaluative bibliometrics can be avoided if we trust the experts. Such views are expressed, for example, in relation to citation databases, where direct use of these is discouraged: The “own databases” built by “professional experts” should be consulted for evaluation purposes rather than the “raw data” provided by databases such as Web of Science or Scopus (Rousseau et al., 2018, p. 284). It should be acknowledged that the need for collegial influence and contextual understanding is emphasized in this account, and critical literature on the use of quantitative indicators is frequently cited. Yet, the main conclusion is that “bibliometric expertise is needed, and counting is a necessity.” The idea of “metric wiseness” appeals to the notion of specialists sharing their wisdom to the uninformed. In this account, the nonexperts, using for example “raw data,” are to blame for misuse of indicators—a problem that increasing academics’ knowledge and skills in bibliometrics will address. Hence, expertise, knowledge, and techniques are the answers to the current problems of bibliometric evaluation: No strong reform is needed, just an intensification of the current ambitions of the bibliometric community. Better metrics, combined with more educated users, hold the solutions to reform, not altogether abandoning or dismissing bibliometrics (Bornmann et al., 2023).

#### 4.4. Reflexive Framing

Recently, the *reflexive framing* has gained in momentum, influenced by parallel “responsibilization” movements—such as Responsible Research and Innovation (Owen, Macnaghten, & Stilgoe, 2020). Both the Responsible Research and Innovation and the responsible metrics movements share the common influence of science and technology studies, particularly



research on technology assessment and public engagement. Like Responsible Research and Innovation, the reflexive framing towards bibliometrics emphasizes the need to guide evaluators and other research system actors and their practices from a distance, by setting out broad parameters of “responsible” action (Davies & Horst, 2015).

In the reflexive framing, bibliometric tools can in principle be used as supporting tools in evaluations, *if* good practices are kept constantly in mind. Codified standards and principles of good practice are important tools within the reflexive frame, and are part of a counterframing to what we have labeled the metrics skepticism framing, by seeking to shift the terms of debate “from a defensive one (e.g., ‘one cannot use these indicators in the humanities’) to a specification of the conditions under which assessments can be accepted as valid, and the purposes for which indicators might legitimately be used” (Leydesdorff, Wouters, & Bornmann, 2016, p. 2132). Unlike the metrics skepticism framing, the reflexive framing is less about legislating whether a given bibliometrics tool is “good” or “bad,” but tries to impress upon evaluators a moral obligation to practice their evaluations with *care* (the professional duty to act as a custodian of the apparatus of evaluation) and *responsiveness* (willingness and motivation from within to acknowledge ambiguity of research quality and enter into dialog with other actors and listen to others’ beliefs and practices) (Dorbeck-Jung & Shelley-Egan, 2013; Pellizzoni, 2004). *Democratization* is another dimension of responsible research and innovation (Owen et al., 2020) which has been translated into the reflexive framing of evaluative bibliometrics, via calls for diversifying the range of actors to be included in the construction and use of bibliometric tools (Ràfols, 2019). Last, *inclusion* (or *diversity*) is an important principle of the reflexive framing, particularly the imperative to widen what counts as valuable research contributions beyond publication and citation indicators.

In the reflexive framing, the appropriateness of bibliometrics can only be worked out in specific contexts by autonomous social agents committed to self-examining whether their practices are proceeding carefully and responsively. Bibliometrics are not one-size-fits-all tools that diffuse into evaluation settings but traveling standards to be translated by responsible evaluators to “fit the local context, resulting in heterogeneity and context-dependent spread” (Reymert, 2021, 53).

## 5. THE 2010S—RESPONSIBLE METRICS GATHERS MOMENTUM AS A PROFESSIONAL REFORM MOVEMENT

### 5.1. DORA

The San Francisco Declaration on Research Assessment (DORA) was one of the first organized attempts to raise awareness of problems associated with uses of research indicators for evaluative purposes and call research system actors to reform what its authors considered bad practices. Following a special session at the American Society of Cell Biology (ASCB) Conference in December 2012, the DORA statement was written by a committee of attendees and released in May 2013. At the time of writing this review, DORA’s text remains unchanged, calling on organizations and individuals to sign the declaration and commit to its principles for good evaluation. The statement includes 18 principles, some of which are general, and others aimed towards specific actors within the research system charged as accountable for change—including individuals, universities, funders, and publishers. Important general recommendations that signatories are expected to adhere to are not to use “... journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist’s contributions, or in hiring, promotion, or funding decisions” and instead “assess research on its own merits rather than on the basis of the journal

in which the research is published” (DORA, 2013). Today DORA is viewed as a broad initiative for changing how research is evaluated, yet it was launched in the discipline-specific setting of cell biology. This context is a likely explanation for why the Journal Impact Factor (JIF)—which is used across many fields, but is often thought to have an elevated position in biosciences (Rushforth & de Rijcke, 2015)—is the main metric targeted in the text. DORA is not presented as an expert account but rather as an account from “concerned citizens” in the field. Nonetheless, some of the arguments that the DORA statement presents against the use of the JIF align with professional-expert framings of the indicator’s problems, by commenting on technical and statistical deficiencies, not simply the consequences of misuse. Thus, one prognosis to counter the influence of the JIF is to “make available a range of article-level metrics.”

Yet, simultaneously, DORA’s solution to read articles qualitatively on their own merits, rather than rely on quantitative journal indicators mirrors the *reflexive framing* insofar as it calls on evaluators’ sense of moral responsibility to be *responsive* towards the content of each research contribution, rather than assuming that what constitutes “good” can be known independently of reading it. The statement calls for more *inclusive* definitions of what constitutes a worthwhile academic contribution (albeit the examples provided underscore the disciplinary origins of the statement and its imagined audience): “The outputs from scientific research are many and varied, including: research articles reporting new knowledge, data, reagents, and software; intellectual property; and highly trained young scientists.” This statement reads now as though it is an early precursor to later developments in the responsible research assessment movement to consider more holistic criteria of academic contributions (see Section 6).

Simultaneously, DORA can also be interpreted as a statement that is skeptical of metrics, insofar as it calls for abandonment of the JIF. The statement draws on many resources to criticize this measure, including technical arguments that we might associate with the professional-expert framing, though it is not clear that DORA is trying to recover the authority of bibliometric knowledge and expertise from “mis-uses.” In our reading, the original DORA statement is difficult to characterize entirely within a skeptical framing, as it does not address whether bibliometrics per se are inappropriate, only the shortcomings of the JIF and “other” journal-based indicators (which it does not name) used to judge the merits of research articles.

The Declaration has taken on something of a life of its own since its publication as a symbol for various arguments and calls to action. Since its publication, DORA has been interpreted as a specific critique of the JIF and its influence on science, and/or a more general critique of the use of metrics, which takes the JIF as its primary example, but may also take aim at other contentious indicators such as the *h*-index (which the original DORA statement does not mention). Reception towards DORA has varied: There are clearly many enthusiastic followers who follow its official Twitter (X) account, cite the statement positively, attend community engagement events, or have signed it as individuals (or encouraged their organizations to do so). There has also been criticism and counterframings of the Declaration by some within the world of biomedicine, including former *Nature* editor Philip Campbell (Anderson, 2013). A post by the founder of the influential Scholarly Kitchen blog countered DORA’s diagnosis of JIF’s impacts upon the research system thus:

“There’s a deeper problem with the DORA declaration, which is an unexpressed and untested inference in their writing about how the impact factor may be relevant to academic assessment groups. They assert repeatedly, and the editorials expand on these assertions, that tenure committees and the like connect the impact factor to each individual article, as if the article had attained the impact stated. I don’t believe this is true. I believe that the impact factor for a journal provides a modest signal of the selectivity of

the journal—given all the journal titles out there, a tenure committee has to look for other ways to evaluate how impressive a publication event might be.” (Anderson, 2013)

Broadly speaking, though, DORA has become an important reference point and symbol of an emerging global reform movement, and signing it has become something many institutions have been keen to communicate to their stakeholders as evidence of their commitment to being (or becoming) responsible actors.

## 5.2. Leiden Manifesto

Compared to DORA, the Leiden Manifesto offers a more focused defense of evaluative bibliometrics in general. The Leiden Manifesto was published in *Nature* in 2015, written by a small group of scientometricians and science studies scholars following a meeting at the 2014 Science and Technology Indicators conference held in Leiden, the Netherlands. It laid out 10 good practice principles for the appropriate uses of quantitative indicators—addressing audiences of evaluators and those being evaluated. As the authors themselves noted, most of the recommendations codified within the Manifesto were based on forms of technical and practical wisdom already well articulated within the field of scientometrics, but which had been, they argued, less well articulated beyond it.

The text’s introductory section provides a diagnosis of evaluative bibliometrics’ influence upon academic research systems. This shares much of the narrative features of the skeptical framing, including concerns about the ubiquity of performance indicators, a graph on “impact factor obsession,” and a strong assertion that: “The problem is that evaluation is now led by the data rather than by judgement” (Hicks et al., 2015). Words such as *obsession* imply that quantitative measures can pose an irrational threat to reasoned, expert judgment (Leckert, 2021).

Despite setting out contemporary research systems’ quantitative metrics problem in the introduction, the remainder of the text *does not* align with the skeptical framing prognosis (whereby full-scale delegitimation and abandonment of metrics is warranted). The authors instead state that uses of bibliometrics for evaluation purposes are “usually well intentioned, not always well informed, often ill applied”—a statement that introduces the *professional-expert* framing of users’ knowledge deficits as an important source of current bibliometric problems. Elements of the *professional-expert* framing put forward to address this diagnosis include recommendations that evaluators avoid “misplaced concreteness and false precision” of indicators, citing a well-known criticism of using three decimal points in the JIF, and for evaluators to “account for variation by field in publication and citation practices.” Calls for evaluations to ensure “robust statistics,” “data quality,” and “normalized indicators” also feature, as do terms such as “abuse” and “mis-application”—vocabulary that is common within the professional-expert frame.

The Manifesto combines information on technical limitations of bibliometrics with calls for evaluators to be more mindful of their limitations and effects within the research system, thereby eclectically combining the professional-expert and reflexive framings. As an intervention in a debate, the Leiden Manifesto can be read partly as professional-experts communicating pre-packaged technical knowledge to lay-persons, while also supporting “self-scrutinization” (Leckert, 2021, p. 9869), with a logic of coproducing more responsible academic citizens (or citizen bibliometricians). Consistent with the reflexive frame, the 10 principles are presented as modest, simplified rules that can help actors go on in the complex world of research evaluation. A feature of the responsible research and innovation movement, which the authors of the Leiden Manifesto also draw on, is equating responsibility with accountability: “We offer this

distillation of best practice in metrics-based research assessment so that researchers can hold evaluators to account, and evaluators can hold their indicators to account” (Hicks et al., 2015). The text calls not only for appropriate use of bibliometric indicators, but asks evaluators to be more *inclusive* of diverse indicators of research quality that go beyond bibliometrics, a feature that would be later amplified in assessment reform statements (see Section 6). Further instances of inclusiveness are calls to develop new indicators that better capture diverse forms of research, particularly outputs published in non-English-language outlets less well covered by existing bibliometric databases: “Metrics built on high-quality non-English literature would serve to identify and reward excellence in locally relevant research” (Hicks et al., 2015). There is a circumscribed form of democratization called for, in stating that those being evaluated via bibliometric methods and data should be able to check the data being used. However, calls for democratization are not extended to the peer review process (e.g., the Manifesto does not challenge what constitutes an expert, nor does it call for widening participation of nontraditional groups as expert reviewers).

The reflexive framing dimensions of the Manifesto have not resonated with all in the scientometrics field. David and Frangopol (2015) provided a counterframing to the Manifesto by doubling down on the message that users are the ones responsible for problems surrounding evaluative bibliometrics, not the tools or, by extension, the field of scientometrics (David & Frangopol, 2015). Another counterframing which emerged was the aforementioned “Metric-Wise” argument (Rousseau et al., 2018), whereby the expert scientometrician provides the lay evaluator with clear, solid knowledge to learn and to implement (with *metric wiseness* offered to reformers as an alternative umbrella label to organize under, instead of responsible metrics). So far, these counterframings have not restrained the global circulation of the Leiden Manifesto, which has been a relatively effective intervention compared with earlier attempts to address problems associated with evaluative bibliometrics that relied mostly on the professional-expert frame. This is likely in part because the genre of the manifesto of best practice principles is a lighter and more resonant intervention for mainstream academic and policy audiences than interventions such as technical textbooks, arguments, and courses through which professional-expert framings have been predominantly mobilized.

### 5.3. The Metric Tide

The Metric Tide Report initially was commissioned to address a policy question posed by the Higher Education Funding Council for England (the body that was responsible for distributing the United Kingdom’s higher education funds) in 2014: Should the next installment of the Research Excellence Framework (REF; the United Kingdom’s periodic national research evaluation exercise) become metrics driven rather than peer review driven? Unlike the Leiden Manifesto, which assumes bibliometrics are already a general presence in manifold evaluation settings, the Metric Tide is written not only for evaluators but also for a policy and management audience about what is to happen to a particular evaluation exercise that was scheduled for a particular time. In addition, the stakeholder consultation process led to an expansion from this scope, to review the impacts of metrics upon research systems more generally (Wilsdon, 2017).

The report contained an executive summary with recommendations and conclusions, which cited the report’s commissioned independent literature review (see our Competing Interests statement) and correlation analysis supplementary sections as evidence. The report’s executive summary concluded that introduction of bibliometrics into the REF was not appropriate for now. In doing so, the executive summary draws on all three framings—the skeptical framing, professional-expert framing, and reflexive framing.

In delivering its conclusion, the report positioned itself as spokesperson on behalf of the research community and its concerns about bibliometrics: “Across the research community, the description, production and consumption of ‘metrics’ remains contested and open to misunderstandings” (p. viii). In supporting this finding, the report cites DORA as an authority on the destructive effects of “narrow, poorly-defined indicators—such as journal impact factors” (p. viii). Like the Leiden Manifesto, the beginning of the executive summary starts with a diagnosis of problems that is compatible with the metric skepticism framing in its tone: “Too often, poorly designed evaluation criteria are ‘dominating minds, distorting behaviour and determining careers.’ At their worst, metrics can contribute to what Rowan Williams, the former Archbishop of Canterbury, calls a “new barbarity” in our universities” (p. iii).

Though a different genre of writing to the Leiden Manifesto, with a different history and target audience, the Metric Tide also did not advocate for wholesale removal of bibliometrics but, in theory, leaves the door open for their use in future assessment exercises (after 2021)—subject to certain conditions being met. The executive summary lays out five principles of Responsible Metrics (thereby coining this umbrella label that many have used to name the reform movement). These are mobilized as a set of standards that bibliometric indicators would need to meet in order to play a useful role in the REF (but never replace its peer review)—which, the report argued, they did not meet at that time. The executive summary explicitly links the five principles to the Responsible Research and Innovation agenda (Owen et al., 2020). However, we argue that some of the Metric Tide’s principles chime with the professional-expert framing more than the reflexive, Responsible Research and Innovation-style framing. The principle of robustness (“basing metrics on the best possible data in terms of accuracy and scope” (p. x)), for instance, aligns more closely with the professional-expert framing, and elsewhere the executive summary draws on a technical lens when citing evidence from its independently commissioned correlation analysis as to why metrics should not replace peer review. The report cites the Leiden Manifesto and endorses its principles, and shares the Leiden Manifesto’s call for democratization: Data and indicators should be subjectable to scrutiny by those being evaluated. Like the Leiden Manifesto, though, it does not challenge the authority of expert judgment and who should count as a peer in peer review. Indeed, the report adopts a spokesperson position on behalf of the academic community in arguing that peer review should be retained in the REF and in research assessments at large: “despite its flaws and limitations, [peer review] continues to command widespread support across disciplines” (p. viii).

The fifth principle of reflexivity (unsurprisingly) resonates most strongly with the reflexive framing, with its emphasis on recognizing and anticipating systemic and performative effects of indicators, challenging the limitations of current tools, updating them, and being alert to development of new indicators.

Since being published in 2015, the five principles of responsible metrics are perhaps the elements of the report that have circulated most widely (alongside the umbrella label Responsible Metrics). Universities in the United Kingdom were also encouraged to publish responsible metrics statements on their websites committing to these principles publicly, and the Metric Tide subsequently led to the establishment of an independent sector-wide U.K. Forum for Responsible Metrics.

#### **5.4. DORA, the Leiden Manifesto, and the Metric Tide Become Fellow Travelers**

In the years following the publication of these three texts, champions and supporters of research assessment reforms have begun to refer to responsible metrics as a *movement*

(e.g., Curry, Gadd, & Wilsdon, 2022). Although we have shown that DORA, the Leiden Manifesto, and the Metric Tide contain differences in aims, audiences, and arguments, we would argue that in the years following their publication, their similarities have tended to be emphasized much more than their differences. Let us consider how the three texts have often come to be cited together in statements and policies found online. In the United Kingdom, the Wellcome Trust, for example, introduced its Open Access policy with a request (though not requirement) that Wellcome-funded organizations publicly commit to assessing research outputs and other contributions based on their intrinsic merit and discourage inappropriate metrics or proxies such as the title or impact factor of the journal in which work is published. The Wellcome Trust cites DORA's principles as a key text in informing this policy, referencing two of its principles directly, before referring to "other equivalent declarations" such as the Leiden Manifesto (Wellcome Trust, 2018). As requested of universities by the Metric Tide, the University of Bristol published a responsible metrics statement which is exemplary of how one or more of DORA, the Leiden Manifesto, and the Metric Tide have come to be cited together:

"This Policy Statement builds on a number of prominent external initiatives on the same task, including the San Francisco Declaration on Research Assessment (DORA), the Leiden Manifesto for Research Metrics and the Metric Tide report." (University of Bristol, n.d.)

Rhetorical practices of grouping together these three texts resemble the use of "concept symbols" in scientific articles (Small, 1978), whereby certain references become shorthand symbols for authoritative sources that lend credibility to an author's statement and over time cease being expanded upon in detail. In citing such texts together under headings such as *Responsible Metrics Statement*, universities and funders present themselves as having signed up to a burgeoning movement, symbolized by these three texts. Such groupings tend to present responsible metrics as a unified, established agenda with a set of shared values and principles, with frictions between metrics skepticism, professional-expert, and reflexive framings rendered invisible.

## 6. FRAME EXTENSION: FROM RESPONSIBLE METRICS TO RESPONSIBLE RESEARCH ASSESSMENT

Since the late 2010s, a notable *frame extension* (Benford & Snow, 2000) of the responsible metrics reform movement has occurred, from the more specific focus on appropriate uses of bibliometrics into a widened framing of "responsible research assessment." A 2020 report by members of the Research on Research Institute (including authors of DORA, the Leiden Manifesto and the Metric Tide texts) defined responsible research assessment as "an umbrella term for approaches to assessment which incentivize, reflect and reward the plural characteristics of high-quality research, in support of diverse and inclusive research cultures" (Curry et al., 2020, p. 7). A large number of texts have emerged supporting this widened agenda (CoARA, 2022; EC, 2017; EU, 2022; LERU, 2022; UNESCO, 2021). Published in 2022, *Harnessing the Metric Tide* served as a follow-up to the 2015 Metric Tide Report, in preparation for the United Kingdom's 2028 Research Excellence Framework, presenting a fine-tuning of the five original responsible metrics principles within the original report, and stating that any use of bibliometrics in research evaluations should be "judicious." *Harnessing the Metric Tide*, though, explicitly endorses the expansion of research assessment reform agendas from the narrower focus on responsible uses of bibliometrics towards ensuring research assessment aligns more widely with "with intersecting movements to support more fruitful, inclusive and positive research cultures" (Curry et al., 2022, p. 13).

A prominent feature of how responsible research assessment texts perform *frame extension* is through referring to and incorporating goals and agendas from parallel scientific reform movements, such as open science, research integrity, and societal relevance of research, as well as drives for equity, diversity, and inclusion, and to improving academic working environments. For example, the European Agreement on research assessment reforms calls on assessments to “reward research behaviour underpinning open science practices such as early knowledge and data sharing as well as open collaboration within science and collaboration with societal actors where appropriate” (CoARA, 2022, p. 4). A recent position statement by the League of European Research Universities (LERU) draws assessment reform and open science causes together as related: “This paper complements other recent papers from LERU on inclusion, scientific integrity, societal impact and on the implementation of Open Science” (LERU, 2022, p. 6). *Harnessing the Metric Tide* calls for research assessment to better support, incentivize, and recognize shifts in academic culture around “issues of equality, diversity, bullying and harassment” (Curry et al., 2022, p. 13), while also reducing assessment burden. In responsible research assessment texts, bibliometric indicators and wider valuations of productivity and impact among scientific peers are widely cast as “traditional” approaches, in need of modernizing—rhetoric that is common across various contemporary science reform movements (Penders, 2022).

Interactions between the responsible metrics movement and other reform agendas have also resulted in *frame translation* (Frickel, 2004a). Texts associated with research integrity and open science movements have, for example, mostly accepted the responsible metrics’ movement’s diagnosis that perverse effects of bibliometrics are an obstacle to them realizing their own reform goals. The UNESCO Recommendation on Open Science states: “Assessment of scientific contribution and career progression rewarding good open science practices is needed for operationalization of open science” (UNESCO, 2021, p. 27). Similarly, when a group of research integrity reform champions proposed the Hong Kong Principles in 2020, they accepted evaluative bibliometrics as an urgent problem blocking realization of their agenda:

“We acknowledge ... the global leadership of those working on the San Francisco Declaration on Research Assessment (DORA), the Leiden Manifesto, and other initiatives to promote the responsible use of metrics, which have laid the foundations for much of our work. The HKPs [Hong Kong Principles] are formulated from the perspective of the research integrity community. We, like the DORA signatories, strongly believe that current metrics may act as perverse incentives in the assessment of researchers” (Moher, Bouter et al., 2020, p. 2).

In responsible research assessment texts, DORA, the Leiden Manifesto, and the Metric Tide are frequently cited as part of the evidence base for why assessment practices need to be reformed. Yet when responsible research assessment statements cite these three texts, certain elements get *amplified* more than others. The European Agreement on Research Assessment Reform amplifies, for instance, the need to recognize a diverse range of outputs and contributions and to respect the diversity of disciplines (CoARA, 2022, p. 4). The Agreement, though, omits the Leiden Manifesto’s call to reduce the dominance of the English language as the default language of academic publication.

Importantly, responsible research assessment texts amplify the message of the Leiden Manifesto and Metric Tide that bibliometrics ought to be given “license to continue,” as long as they are used appropriately. Like DORA, the Leiden Manifesto, and the Metric Tide,

responsible research assessment texts also make strategic use of the three action framings that our analysis identifies in order to persuade. Part of the argumentative structure of the Leiden Manifesto and the Metric Tide taken forward by responsible research assessment texts is to set the scene (diagnosis) by being heavily critical of evaluative bibliometrics. Ultimately, though, responsible research assessment texts have tended to align with the Leiden Manifesto's and Metric Tide's reflexive prognosis for addressing bibliometric-related problems: Appropriate uses of metrics by self-aware agents is the way forward, not the abandonment of metrics. The League of European University's (LERU, 2022) position statement, for example, initially diagnoses problems through the metrics skepticism framing, for instance, arguing that bibliometrics are used in assessments because they are "so much easier than focusing on what really counts" (LERU, 2022, p. 5) and linking bibliometrics to injustices: "They [candidates with different profiles or career choices] may feel misjudged and wronged, because their strengths did not get the same weight in the assessment than the publication ratio" (p. 7). However, the LERU position statement text then aligns with the Leiden Manifesto's reflexive framing of how to address problems with bibliometrics: "Although bibliometric data have played *and still play* an important role, LERU universities have always adopted a multidimensional perspective, where different dimensions of research performance and a variety of duties and responsibilities are taken into account for assessment" (p. 7, emphasis added). The primacy of expert judgement and "qualitative" peer review, with input of quantitative measures only where appropriate (rather than no bibliometrics)—is a commonly repeated mantra throughout many position statements, agreements, and guidelines advocating assessment reforms (CoARA, 2022; EC, 2017; EU, 2022; LERU, 2022; UNESCO, 2021).

## 7. CONCLUDING REMARKS: HOW SECURE IS THE RELEGITIMATION OF EVALUATIVE BIBLIOMETRICS?

In recent years, calls for research assessment reforms have begun to gain greater attention in research policy in *some* contexts—especially in parts of the Global North (Tijssen, 2020). Many assessment reform statements cite and credit DORA, the Leiden Manifesto, and the Metric Tide as evidence of growing momentum for change. These texts have become symbols for a "responsible metrics" agenda, and have helped long-standing concerns about evaluative bibliometrics, long discussed in quantitative science studies research, gain wider attention. Communicating practical wisdom accumulated within quantitative science studies communities via manifestos and statements has surely brought such concerns to a wider audience compared with more traditional academic communication channels such as specialized monographs, journals, and conference meetings. One of the major influences of the Leiden Manifesto and the Metric Tide on widening coalitions supporting assessment reforms has been to persuade them of the legitimacy of bibliometrics in assessments *if used appropriately* (DORA has also come to be associated with this prognosis, even though its original text is rather ambiguous on what the general role of bibliometrics should be in assessments). Currently the metric skepticism, professional-expert, and reflexive framings cohabit the responsible metrics movement. This combination of framings no doubt enables the responsible metrics movement to appeal to a broader audience than would be the case if only one of these framings was projected. Given that evaluative bibliometrics are often diagnosed as the reason for many ills in academic research systems, which frustrate the realization of certain other reform agendas, for how long can the metrics skepticism framing remain neutralized within expanding research assessment reform coalitions? This is an important matter of concern for the quantitative science studies community, as assessment reform movements continue to grow, evolve, and perhaps splinter (as expanding social movements often do). We therefore ask readers to consider: Can



this cohabitation of framings hold out peacefully within assessment reform movements, and what might happen to evaluative bibliometrics if it does not?

#### ACKNOWLEDGMENTS

Thanks to Ludo Waltman, Ismael Rafols, and Sven Ulpts for comments on an earlier version of this paper. Thanks to reviewers from *Quantitative Science Studies*, Bart Penders, and Maarten Derksen for insightful peer review comments and challenges.

#### AUTHOR CONTRIBUTIONS

Alexander Rushforth: Conceptualization, Data curation, Formal analysis, Writing—original draft, Writing—review & editing. Björn Hammarfelt: Conceptualization, Data curation, Formal analysis, Writing—review & editing.

#### COMPETING INTERESTS

AR contributed to an independent literature review that formed an appendix to one of the documents analyzed in this review (the Metric Tide), although he did not have any input or involvement in writing or reviewing the main executive summary of the report and the position it took (which is the section we review in this paper). AR also contributed to another of the documents analyzed (the IAP-GYA-ISC 2023 Report on the Future of Research Evaluation), providing written input to the section (3.2.1) about Research Perspectives and Developments in Europe. AR also currently works on a funded project called TARA, funded by Arcadia, on which some members of the DORA steering committee (whose original statement on bibliometric reforms is analyzed closely in this review) are project partners. AR also works on a project AgoRRA with the lead author of the Metric Tide Report, James Wilsdon. AR and BH are colleagues, present and past respectively, of some of the authors of the Leiden Manifesto text.

#### FUNDING INFORMATION

No funding has been received for this research.

#### DATA AVAILABILITY

Not applicable.

#### 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>
- Alberts, B., Kirschner, M. W., Tilghman, S., & Varmus, H. (2014). Rescuing US biomedical research from its systemic flaws. *Proceedings of the National Academy of Sciences*, 111(16), 5773–5777. <https://doi.org/10.1073/pnas.1404402111>, PubMed: 24733905
- Anderson, K. (2013). Does DORA need to attack the impact factor to reform how it is used in academia? *Scholarly Kitchen*. Retrieved from <https://scholarlykitchen.sspnet.org/2013/05/21/impact-crater-does-dora-need-to-attack-the-impact-factor-to-reform-how-the-it-is-used-in-academia/>.
- Benford, R. D., & Snow, D. A. (2000). Framing processes and social movements: An overview and assessment. *Annual Review of Sociology*, 26, 611–639. <https://doi.org/10.1146/annurev.soc.26.1.611>
- Boell, S. K., & Cecez-Kecmanovic, D. (2014). A hermeneutic approach for conducting literature reviews and literature searches. *Communications of the Association for Information Systems*, 34(1), 12. <https://doi.org/10.17705/1CAIS.03412>
- Bornmann, L., Botz, G., & Haunschild, R. (2023). Metrics have their merits. *Research Professional*. Retrieved from <https://www.researchprofessionalnews.com/rr-news-europe-views-of-europe-2023-2-metrics-have-their-merits/>.
- Brundage, M., & Guston, D. H. (2019). Understanding the movement(s) for responsible innovation. In *International handbook on responsible innovation* (pp. 102–121). Chichester: Edward Elgar. <https://doi.org/10.4337/9781784718862.00014>
- Burrows, R. (2012). Living with the *h*-index? Metric assemblages in the contemporary academy. *Sociological Review*, 60(2), 355–372. <https://doi.org/10.1111/j.1467-954X.2012.02077.x>
- Butler, L. (2007). Assessing university research: A plea for a balanced approach. *Science and Public Policy*, 34(8), 565–574. <https://doi.org/10.3152/030234207X254404>

- Chen, C. M.-L., & Lin, W.-Y. C. (2017). What have we learned from San Francisco Declaration on Research Assessment and Leiden Manifesto? *Journal of Educational Media and Library Sciences*, 54, 111–129. <https://joemls.dils.tku.edu.tw/fulltext/54/54-1/111-129.pdf>. <https://doi.org/10.6120/JoEMLS.2017.541/0015.OR.AM>
- Chubin, D., & Hackett, E. (1990). *Peerless science: Peer review and US science policy*. Albany, NY: SUNY Press.
- CoARA. (2022). *Agreement on reforming research assessment*. [https://coara.eu/app/uploads/2022/09/2022\\_07\\_19\\_rra\\_agreement\\_final.pdf](https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf)
- Collins, H. M. (1985). The possibilities of science policy. *Social Studies of Science*, 15(3), 554–558. <https://doi.org/10.1177/030631285015003009>
- Csiszar, A. (2023). Provincializing impact: From imperial anxiety to algorithmic universalism. *Osiris*, 38(1), 103–126. <https://doi.org/10.1086/725131>
- Curry, S., de Rijcke, S., Hatch, A., Pillay, D. G., van der Weijden, I., & Wilsdon, J. (2020). The changing role of funders in responsible research assessment: Progress, obstacles and the way ahead. *RoRI Working Paper No. 3*. <https://doi.org/10.6084/m9.figshare.13227914.v2>
- Curry, S., Gadd, E., & Wilsdon, J. (2022). *Harnessing the Metric Tide: Indicators, infrastructures & priorities for UK responsible research assessment. Report of The Metric Tide Revisited panel*. [https://rori.figshare.com/articles/report/Harnessing\\_the\\_Metric\\_Tide/21701624](https://rori.figshare.com/articles/report/Harnessing_the_Metric_Tide/21701624)
- Dahler-Larsen, P. (2012). *The evaluation society*. Stanford University Press. <https://doi.org/10.1515/9780804778121>
- David, D., & Frangopol, P. (2015). The lost paradise, the original sin, and the Dodo bird: A scientometrics *Sapere Aude* manifesto as a reply to the Leiden manifesto on scientometrics. *Scientometrics*, 105(3), 2255–2257. <https://doi.org/10.1007/s11192-015-1634-2>
- Davies, S. R., & Horst, M. (2015). Responsible innovation in the US, UK and Denmark: Governance landscapes. In *Responsible innovation 2* (pp. 37–56). Cham: Springer. [https://doi.org/10.1007/978-3-319-17308-5\\_3](https://doi.org/10.1007/978-3-319-17308-5_3)
- DeFronzo, J., & Gill, J. (2019). *Social problems and social movements*. Rowman & Littlefield.
- 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. <https://doi.org/10.1002/asi.23382>
- 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. <https://doi.org/10.1093/reseval/rvv038>
- Derksen, M., & Field, S. (2022). The tone debate: Knowledge, self, and social order. *Review of General Psychology*, 26(2), 172–183. <https://doi.org/10.1177/10892680211015636>
- Desrosières, A. (1998). *The politics of large numbers: A history of statistical reasoning*. Harvard University Press.
- DORA. (2013). *The Declaration*. <https://sfidora.org/read/>
- Dorbeck-Jung, B., & Shelley-Egan, C. (2013). Meta-regulation and nanotechnologies: The challenge of responsabilisation within the European Commission's code of conduct for responsible nanosciences and nanotechnologies research. *Nanoethics*, 7(1), 55–68. <https://doi.org/10.1007/s11569-013-0172-8>
- EC. (2017). *Evaluation of research careers fully acknowledging Open Science practices*. <https://op.europa.eu/en/publication-detail/-/publication/47a3a330-c9cb-11e7-8e69-01aa75ed71a1/language-en>. <https://doi.org/10.2777/75255>
- EU. (2022). *Research assessment and implementation of Open Science—Council conclusions*. <https://www.consilium.europa.eu/media/56958/st10126-en22.pdf>
- Franssen, T. (2022). *Cultivation devices: Sustainability as a quality*. Unpublished manuscript.
- Franssen, T., & Wouters, P. (2019). Science and its significant other: Representing the humanities in bibliometric scholarship. *Journal of the Association for Information Science and Technology*, 70(10), 1124–1137. <https://doi.org/10.1002/asi.24206>
- Frickel, S. (2004a). Building an interdiscipline: Collective action framing and the rise of genetic toxicology. *Social Problems*, 51(2), 269–287. <https://doi.org/10.1525/sp.2004.51.2.269>
- Frickel, S. (2004b). Just science? Organizing scientist activism in the US environmental justice movement. *Science as Culture*, 13(4), 449–469. <https://doi.org/10.1080/0950543042000311814>
- Hammarfelt, B., & Rushforth, A. D. (2017). Indicators as judgment devices: An empirical study of citizen bibliometrics in research evaluation. *Research Evaluation*, 26(3), 169–180. <https://doi.org/10.1093/reseval/rvx018>
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429–431. <https://doi.org/10.1038/520429a>, PubMed: 25903611
- Kang, D., & Evans, J. (2020). Against method: Exploding the boundary between qualitative and quantitative studies of science. *Quantitative Science Studies*, 1(3), 930–944. [https://doi.org/10.1162/qss\\_a\\_00056](https://doi.org/10.1162/qss_a_00056)
- Leckert, M. (2021). (E-)valuative metrics as a contested field: A comparative analysis of the altmetrics—and the Leiden Manifesto. *Scientometrics*, 126(12), 9869–9903. <https://doi.org/10.1007/s11192-021-04039-1>
- LERU. (2022). *A pathway towards multidimensional academic careers: A LERU framework for the assessment of researchers*. [https://www.leru.org/files/Publications/LERU\\_PositionPaper\\_Framework-for-the-Assessment-of-Researchers.pdf](https://www.leru.org/files/Publications/LERU_PositionPaper_Framework-for-the-Assessment-of-Researchers.pdf)
- 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. <https://doi.org/10.1007/s11192-016-2150-8>, PubMed: 27942086
- Moed, H. F. (2007). The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review. *Science and Public Policy*, 34(8), 575–583. <https://doi.org/10.3152/030234207X255179>
- Moher, D., Bouter, L., Kleinert, S., Glasziou, P., Sham, M. H., ... Dirnagl, U. (2020). The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLOS Biology*, 18(7), e3000737. <https://doi.org/10.1371/journal.pbio.3000737>, PubMed: 32673304
- Nåstesjö, J. (2021). Navigating uncertainty: Early career academics and practices of appraisal devices. *Minerva*, 59(2), 237–259. <https://doi.org/10.1007/s11024-020-09425-2>, PubMed: 33343041
- Owen, R., Macnaghten, P., & Stilgoe, J. (2020). Responsible research and innovation: From science in society to science for society, with society. In *Emerging technologies: Ethics, law and governance* (pp. 117–126). London: Routledge. <https://doi.org/10.4324/9781003074960-11>
- Pellizzoni, L. (2004). Responsibility and environmental governance. *Environmental Politics*, 13(3), 541–565. <https://doi.org/10.1080/0964401042000229034>
- Penders, B. (2022). Process and bureaucracy: Scientific reform as civilisation. *Bulletin of Science, Technology & Society*, 42(4), 107–116. <https://doi.org/10.1177/02704676221126388>

- Petersohn, S. (2021). The competent bibliometrician—A guided tour through the scholarly and practitioner literature. In R. Ball (Ed.), *Handbook of bibliometrics* (pp. 485–495). Berlin: De Gruyter. <https://doi.org/10.1515/9783110646610-044>
- Petersohn, S., Biesenbender, S., & Thiedig, C. (2020). Investigating assessment standards in the Netherlands, Italy, and the United Kingdom: Challenges for responsible research evaluation. In *Shaping the future through standardization* (pp. 54–94). IGI Global. <https://doi.org/10.4018/978-1-7998-2181-6.ch003>
- Pontika, N., Klebel, T., Correia, A., Metzler, H., Knoth, P., & Ross-Hellauer, T. (2022). Indicators of research quality, quantity, openness, and responsibility in institutional review, promotion, and tenure policies across seven countries. *Quantitative Science Studies*, 3(4), 888–911. [https://doi.org/10.1162/qss\\_a\\_00224](https://doi.org/10.1162/qss_a_00224)
- Poovey, M. (1998). *A history of the modern fact: Problems of knowledge in the sciences of wealth and society*. Chicago, IL: University of Chicago Press. <https://doi.org/10.7208/chicago/9780226675183.001.0001>
- Porter, T. M. (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press. <https://doi.org/10.1515/9781400821617>
- Power, M. (1997). *The audit society: Rituals of verification*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198296034.001.0001>
- Ràfols, I. (2019). S&T indicators in the wild: Contextualization and participation for responsible metrics. *Research Evaluation*, 28(1), 7–22. <https://doi.org/10.1093/reseval/rvy030>
- Reymert, I. (2021). Bibliometrics in academic recruitment: A screening tool rather than a game changer. *Minerva*, 59(1), 53–78. <https://doi.org/10.1007/s11024-020-09419-0>
- Rice, D. B., Raffoul, H., Ioannidis, J. P. A., & Moher, D. (2020). Academic criteria for promotion and tenure in biomedical sciences faculties: Cross sectional analysis of international sample of universities. *British Medical Journal*, 369, m2081. <https://doi.org/10.1136/bmj.m2081>, PubMed: 32586791
- Rousseau, R., Egghe, L., & Guns, R. (2018). *Becoming metric-wise: A bibliometric guide for researchers*. Chandos Publishing. <https://doi.org/10.1016/C2017-0-01828-1>
- 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. <https://doi.org/10.1007/s11024-015-9274-5>, PubMed: 26097258
- Sample, I. (2013). Nobel winner declares boycott of top science journals. *Guardian*. <https://www.theguardian.com/science/2013/dec/09/nobel-winner-boycott-science-journals>
- Shore, C., & Wright, S. (2015). Governing by numbers: Audit culture, rankings and the new world order. *Social Anthropology/Anthropologie Sociale*, 23(1), 22–28. <https://doi.org/10.1111/1469-8676.12098>
- Small, H. G. (1978). Cited documents as concept symbols. *Social Studies of Science*, 8(3), 327–340. <https://doi.org/10.1177/030631277800800305>
- Strathern, M. (2003). Introduction: New accountabilities: Anthropological studies in audit, ethics and the academy. In M. Strathern (Ed.), *Audit cultures* (pp. 13–30). London: Routledge. <https://doi.org/10.4324/9780203449721-3>
- Tijssen, R. (2020). Re-valuing research excellence: From excellentism to responsible assessment. In E. Kraemer-Mbula, R. Tijssen, M. Wallace, & R. McLean (Eds.), *Transforming research excellence: New ideas from the Global South* (pp. 59–78). Cape Town: African Minds. <https://doi.org/10.5281/zenodo.3607319>
- TJNK. (2020). *Good practice in researcher evaluation. Recommendation for the responsible evaluation of a researcher in Finland*. <https://www.aka.fi/en/research-funding/responsible-science/responsible-researcher-evaluation/>
- UiR. (2021). *NOR-CAM: A toolbox for recognition and rewards in academic careers*. [https://www.uhr.no/en/\\_f/p3/i86e9ec84-3b3d-48ce-8167-bbae0f507ce8/nor-cam-a-tool-box-for-assessment-and-rewards.pdf](https://www.uhr.no/en/_f/p3/i86e9ec84-3b3d-48ce-8167-bbae0f507ce8/nor-cam-a-tool-box-for-assessment-and-rewards.pdf)
- UNESCO. (2021). *Recommendation on Open Science*. <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>
- University of Bristol. (n.d.). *Statement on responsible research assessment*. Retrieved from <https://www.bristol.ac.uk/research/environment/responsible-research-assessment/>
- van Raan, A. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics*, 36(3), 397–420. <https://doi.org/10.1007/BF02129602>
- van Raan, A. (1998). In matters of quantitative studies of science the fault of theorists is offering too little and asking too much. *Scientometrics*, 43(1), 129–139. <https://doi.org/10.1007/BF02458401>
- VSNU, NFU, KNAW, NWO, & ZonMw. (2019). *Position paper 'Room for everyone's talent'*. <https://www.nwo.nl/en/position-paper-room-everyones-talent>
- Weingart, P. (2005). Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics*, 62(1), 117–131. <https://doi.org/10.1007/s11192-005-0007-7>
- Wellcome Trust. (2018). *Guidance for research organisations on how to implement responsible and fair approaches for research assessment*. <https://wellcome.org/grant-funding/guidance/open-access-guidance/research-organisations-how-implement-responsible-and-fair-approaches-research>
- Wilsdon, J. (2016). *The Metric Tide: Independent review of the role of metrics in research assessment and management*. Research England. <https://doi.org/10.4135/9781473978782>
- Wilsdon, J. (2017). Responsible metrics. In T. Strike (Ed.), *Strategy and planning in higher education* (pp. 247–253). Abingdon: Routledge. <https://doi.org/10.4324/9781315206455-15>
- Wilsdon, J. (2021). *From responsible metrics to responsible research assessment (RRA)*. Paper presented at the Tendencias en Medicion de CTI: Propuestas Internacionales.
- Wouters, P. F. (1999). *The citation culture*. Universiteit van Amsterdam.