Journal of Informetrics xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Book Review

The failure of a paradigm

Henk F. Moed. Applied Evaluative Informetrics, Springer International Publishing (2017)

1. Introduction

Applied Evaluative Informetrics is an ambitious book. The latest work of Henk Moed, a prize-winning bibliometrician who worked between 1986 and 2010 at the Centre for Science and Technology Studies, aims to give a sweeping overview of all aspects of the application of informetrics in the evaluation of scientific and scholarly research. The book is intended for non-specialist scholars from all domains of research as well as for all who are interested in research evaluation: senior researchers, students, research managers, funders and science policy actors.

The book consists of six parts. The first four parts cover past, present and future of evaluative informetrics, while the last two parts consist of lectures and articles by the author. In the first four parts the chapters present original arguments in combination with excerpts and presentations of results and insights from older articles. The book is more a synthesis of the state of the art in evaluative informetrics than a novel analysis. But it is more ambitious than just a review in book format. It is also framed as an intervention in the current debate about research evaluation and the role of indicators. Moed tries to make the case for a politically neutral form of informetrics, thereby hoping to clarify the current discourse which he apparently finds often confused and sometimes overly political.

2. Broad overview

Let me start with praising the author for his endeavor. Not many bibliometricians have both such a broad overview of the field and such a firm grounding in the practice of research evaluation. Most practitioners in our field limit themselves to their specific area and do not start such an ambitious project. So, the idea to write this book, which can also be seen as an extension of his earlier book *Citation Analysis in Research Evaluation* from 2005 (Moed, 2005), is commendable.

Readers who are not themselves scientometricians or informetricians will find valuable references they may wish to follow up. Policy officials and research managers pressed for time will find Part I especially useful. It gives a short introduction to the field of informetrics and bibliometrics whilst also presenting an extensive synopsis (in Chapter 2) of all following chapters. Indeed, for many users of evaluative informetrics these first 41 pages will be sufficient to understand the key arguments and references. It is also a good starting point for those in need of information about particular topics without wishing to cover the field as a whole.

Parts II, III and IV contain the key arguments of the book. Part II discusses indicators of research performance, Part III zooms in on the application context, and Part IV presents Moed's ideas about the way forward. Each part consists of several relatively short chapters (most are between 5 and 20 pages). The book has a wide scope; I could not think of topics that were not mentioned at least once in the book. Perhaps as a consequence of this sweeping excursion, the book does not dig deep into the issues. Readers with a taste for more in-depth analysis or discussion may be disappointed if they do not find more than what they see as obvious or common knowledge. This raises the question whether the author would not have served his audience better by focusing the book more on the most burning questions. Of course, this may also be a matter of taste.

3. Performance indicators

Part II about indicators of research performance has three chapters. The first (Chapter 3) introduces the different dimensions of research performance. It distinguishes components of the research process (input, output, process, and impact) and presents a table of 28 important informetric indicators and lists their definitions, their strong points as well as their limits. The indicators listed vary from very general indicators (such as "citations" or "measures of economic value") to very specific

https://doi.org/10.1016/j.joi.2018.03.002
1751-1577/

ones (such as the "Integrated Impact Indicator" or "Glanzel's (sic) negative-binomial model"). Chapter 4 then discusses a number of these indicators in more detail under the heading of informetric tools. The currently existing types of indicators are discussed in short sections: publication and citation indicators, journal metrics, patent and usage indicators, altmetrics and webometric indicators, economic indicators, reputation and esteem based indicators, collaboration indicators and indicators of research infrastructure. Each type of indicators has about one page, so readers who wish to work with these indicators or understand them in some depth will also have to read the later chapters in which these indicators come back or consult the literature references. The chapter concludes with a short discussion of the fact that scientometric data have become big data and of the application of science maps.

Chapter 5 discusses what Moed sees as three important problems in applied informetrics: are journal impact factors good predictors of citations of individual articles?; to what extent do errors cancel out in large datasets?; and what is the correct interpretation of rank correlation coefficients? The interpretation of the journal impact factor is currently a hot topic and an important problem for all users of informetric tools and indicators. Moed is one of the best informed experts on this matter (Moed, 2002). Hence, I had expected a rigorous update to guide users in the application of the Journal Impact Factor. Instead, the chapter is a short presentation of the distribution of citations over articles (which is known to be highly skewed) compared to the length of soccer players (which is a normal one). Moed shows that this means that if one compares two journals with different impact factors, the probability that a randomly selected paper from the journal with the lower impact factor has a citation rate at least as high as that of a randomly selected paper from the journal with the higher impact factor cannot be ignored. So Moed builds his argument that the journal impact factor is not a good predictor of individual articles' citation rates on the skewed distribution of citations over articles within journals. And yet, the section of this chapter ends inconclusively. Moed mentions in the concluding sentences that while preparing the final version of the book, an article by Waltman and Traag (2017) was published which engages with precisely the type of statistical arguments against the journal impact factor that Moed presents in this section. These authors agree with Moed's statistical critique of the journal impact factor but argue that this does not mean that the journal impact factor cannot be used at the level of individual publications. Based on computer simulations, they even conclude that the journal impact factor may be a more accurate indicator of the "value" of an article than the number of citations to that article. Moed summarizes their argument but does not engage with this counter-argument. The reader is left on her own.

Unfortunately, this is a recurring phenomenon in this book: the most interesting points of discussion are brushed over, ignored, or left for the reader to sort out. I must say, I am a bit puzzled about this noncommittal attitude of an author who has spent his professional life in the field of scientometrics and informetrics. The third section of Chapter 5 is one page about errors with the conclusion that random errors do cancel out in large datasets but systematic errors do not. That might be conspicuous even for readers who are not familiar with informetrics but do have some basic statistical knowledge. The last section of this chapter is a two-page discussion of correlation coefficients, basically showing that correlation coefficients should never be taken at face value and that the underlying form of the distribution as well as the data range should be taken into account. Well, indeed.

4. The context of application

Part III is about the application context and consists of three chapters on respectively research assessment as an evaluation science, non-informetric factors influencing indicators, and the policy context. The chapter on research assessment as an evaluation science (Chapter 6 of the book) aims to explore the relationship between the field of quantitative research assessment and evaluation science. I see this indeed as a very important topic, and it was the reason to invite a keynote speaker from evaluation science, Peter Dahler-Larsen, to the ENID/STI 2014 conference in Leiden (Moed also gave a keynote at this conference). However, this chapter is quite disappointing. It observes, correctly, that evaluation science is a multidisciplinary field with concepts and analytical distinctions that can be quite useful in research assessment, but then limits itself to a short presentation based on a very limited slice of the literature in evaluation science. Moed has missed the, in my view, most interesting areas in that field and seems to have thought that selecting one approach (Hunter & Nielsen, 2013) would be sufficient.

Perhaps more importantly, Moed shows in this chapter to be surprisingly uninformed about the research in which scholars from sociology of science and science & technology studies have already made the connection between research assessment and evaluation science (Fochler & De Rijcke, 2017; Gläser, Lange, Laudel & Schimank, 2010; Gläser & Laudel, 2016; Hammarfelt & De Rijcke, 2015; Hammarfelt & Rushforth, 2017; Kaltenbrunner & De Rijcke, 2016; Laudel & Gläser, 2006; Sauder & Espeland 2009; Whitley, 2010; Wouters et al., 2015). For example, Moed cites an article by the Danish evaluation researcher Peter Dahler-Larsen but does not mention his seminal book *The Evaluation Society* (Dahler-Larsen, 2012). As a result, he is regularly stating the obvious in this chapter but presents this as a new approach. Unfortunately, this decreases not only the extent to which the chapter can be called a presentation of the state of the art, but also the quality of the argument. The result is a mundane summing up of rather practical distinctions, such as the difference between a summative and a formative evaluation and between normative and criterion based reference frameworks, in addition to some hairsplitting about the supposed difference between the concepts of evaluation and assessment.

Chapter 7 discusses non-informetric factors that influence the development of indicators. This chapter shows how performance indicators inevitably are based on views on what constitutes research performance that cannot themselves be based on informetric data. This is clearly a crucial chapter for the argument of the book. Moed shows the role of assumptions

Please cite this article in press as: Wouters, P. The failure of a paradigm. *Journal of Informetrics* (2017), https://doi.org/10.1016/j.joi.2018.03.002

about performance in the use of size-dependent versus size-independent indicators, in the decision to focus on the top part of the performance distribution, in the way indicators are normalized, in the decision to focus on the short-term versus the long-term, and in the definition of a proper reference framework. The chapter argues convincingly that the choice for a particular type of indicator inevitably brings with it a particular definition of what should count as performance. For example, if one thinks that production of articles should be rewarded, rather than punished, one typically chooses a size-dependent indicator. And including particular "corrections" in a normalization procedure is, according to Moed, "not a theoretically neutral act, as the developer expresses that these factors *distort* the indicator, so that it does not properly measure what it is supposed to measure.(p. 109). In this sense Moed rediscovers the arguments that have been made in the sociology of science and science and technology studies regarding the political nature of measures and measuring.

The chapter also contains two shorter sections on the need to be aware of the wider context (for example whether one sees the research group or the individual researcher as the basic unit of analysis of performance) and on the relation between indicator development and business interests. I was curious to see whether in this section Moed would share with us his experiences at Elsevier (the information company to which he moved after his departure from CWTS) but this is not the topic of this section. It is a rather dry summary of the various relations in the field between companies and academic units. Moed concludes that indicators are increasingly becoming a tool in business strategies, but does not spell out what this means for the field of research evaluation, informetrics or for the practical application of evaluative informetrics.

Chapter 8 delves into the important topic of the policy context. In nine pages the chapter presents the multi-dimensional research matrix based on an earlier European expert group report of which Moed was one of the members. Indicators constructed to measure one particular aspect may not be suitable to measure other aspects. "Diverse institutional missions, and different policy environments and objectives require different assessment processes and indicators" (p. 119). Users of the matrix are supposed to make explicit choices with regards to the unit of analysis, the relevant dimension of research, the purpose of the assessment, and the relevant systemic characteristics of the unit of analysis. This then should lead to the choice of the proper indicators to be used. Readers who wish to see whether the matrix leads to indicators that are practical in their situation will have to go to the AUBR report since the chapter does not further specify this (European Commission Expert Group on Assessment of University-Based Research, 2010). The remainder of the chapter is devoted to "systemic characteristics" in an attempt to show how policy objectives may influence indicator choice. For example, according to Moed the use of the journal impact factor to measure international orientation is informetrically defensible in a situation where international publication is not the norm and policy wants to stimulate the local scientific community to become more international in its ambitions.

5. The way forward

Part IV sketches Moed's ideas about the way forward. It consists of four chapters dealing with the use of informetric indicators, research performance measurement, altmetrics, and indicator development. Chapter 11 discusses altmetrics, using a book by the consultant and essayist Michael Nielsen (Nielsen 2011). The chapter is a practical expose of what Moed calls "computerization of the research process". It treats the role of ICT in academic research at a very basic level, disregarding the exciting insights that have been generated in the fields of digital humanities, computer science, networked science and science & technology studies (Boonstra, Breure, & Doorn, 2004; Borgman, 2007; Bulger et al., 2011; Hey, Tansley, & Tolle, 2009; Kaltenbrunner 2014; Wouters, Beaulieu, Scharnhorst, & Wyatt, 2013). I felt sorry to see that it does not generate interesting new insights with respect to the potential of altmetrics for research evaluation (Bornmann & Haunschild, 2017; Costas, Zahedi, & Wouters., 2015; Konkiel, 2016; Thelwall, 2005; Thelwall, 2008; Thelwall, Haustein, Larivière, & Sugimoto, 2013; Wouters et al., 2015).

Chapter 9 about indicator use lists what Moed sees as the main problems that need to be tackled to make progress: assessing individual researchers, the limited time horizon of citation analysis, assessing societal impact, effects of indicator use on authors and editors, constitutive effects of indicators, and the need for what Moed calls "evaluative frameworks". With respect to assessment of individual scholars Moed supports the conclusion of the ACUMEN project that researcher-provided narratives should be the core of this assessment (Wouters et al., 2014). In the section on the limited time horizon of both altmetric and citation indicators, Moed defends the position that the value of contributions to the scientific record will in the end only be visible in the long term. "Indicators do *not* measure contribution to scientific-scholarly progress in this sense, but tend to indicate attention, visibility, or short term impact" (p. 134). Although Moed has always made the distinction between intellectual influence and citation impact (see eg. Moed, 2005:222), he seems to emphasize the limitations of citation analysis now even more than he did in the past (see eg. Moed, 2002:731).

The section on the effects of indicator use on authors and editors limits itself to studies in which Moed was involved: a study on mechanisms to raise the journal impact factor and an analysis of the effects of the UK research assessment exercises on publication patterns. Most studies in the fields of evaluation science, management studies or science & technology studies on this topic are ignored (Fochler & De Rijcke, 2017; Gläser et al., 2010; Gläser & Laudel, 2016; Hammarfelt & De Rijcke, 2015; Hammarfelt & Rushforth, 2017; Kaltenbrunner & De Rijcke, 2016; Laudel & Gläser, 2006; Sauder & Espeland 2009; Whitley, 2010; Wouters et al., 2015). Since the field of evaluation studies is recognized as an important field earlier in the book, this systematic omission is rather puzzling and does decrease the value of the book as an overview of the state of the art. A comprehensive inclusion of the relevant literature would have been particularly valuable since the book tries to defend a specific position with respect to effects of indicator use. According to Moed, the issue is not whether scholarly

Please cite this article in press as: Wouters, P. The failure of a paradigm. *Journal of Informetrics* (2017), https://doi.org/10.1016/j.joi.2018.03.002

practice is changed by indicators, but whether or not the application of indicators enhances research performance (p. 137). Of course, this raises the question how research performance is defined. How do we know whether or not performance has increased? Moed does not provide the answer but shows how difficult it is to say something meaningful about it in the case studies in which he was involved. A study of increases of the journal impact factor leads him to conclude that "it is difficult to maintain that the observed changes indicate an increase in journal performance" (p. 137). A case study of the effects of the UK Research Assessment Exercise was more complex to interpret. Moed does not want to exclude the possibility that it stimulated universities to increase the volume and the significance of their research. "On the other hand, one cannot claim a priori that this use indeed did have a positive effect upon performance. Without additional qualitative and quantitative evidence, it is impossible to draw any solid conclusions about this matter." (p. 137) Basically, Moed wrestles with the problem of constitutive effects of indicators because they make it more difficult to interpret the efforts to measure performance. This is not a minor issue but the very heart of the problem of performance management and research evaluation in academia.

But before going into this in more depth, the question must be addressed whether Moed is right in bracketing off possible effects on academic practice. I think he is wrong. Evaluation practices and their instruments, both quantitative metrics and indicators and qualitative procedures have effects. These effects do not only relate to "enhancement" of research performance but also include, among others: the research agenda of disciplines; the priorities that the most influential groups set in their practices; the diversity in terms of gender and ethnicity of the workforce in academia; the nature of the PhD training both at the group level (what kind of interactions are stimulated) and at the individual level (which psychological traits are most conducive to an academic career); the technical and social infrastructure of scientific and scholarly research; and last but not least the power relationships between different fields with respect to funding and career possibilities. In other words, indicators may directly and indirectly influence the very character of what doing research means in our current societies, both for better and for worse. In my view, these matters are actually far more important, also for informetricians, than the narrow-minded focus on performance enhancement.

This brings us to the problem of constitutive effects, the topic of the next two-page section in this chapter (Section 9.5). Moed defines constitutive effects as follows: "The notion of constitutive effects of indicators questions the validity of the distinction between intended and unintended effects, and claims that there is a tendency that the use of indicators of a theoretical concept such as research quality creates a reality in which the concept and the indicator tend to coincide". Hence, "research quality would be more and more defined as to what citations measure." (p. 137) Moed sees this as "a critique" that should be taken "very seriously" and proceeds to make three arguments responding to this critique. First, he labels the tendency "to replace reality with symbols" as "*magical* thinking" (italics in the original) which he clearly rejects. "If such a belief is dominant, a situation would have emerged that, while modern science surpassed a magical attitude *towards nature*, such an attitude has emerged again, but now *towards science*, namely in research assessment and science policy" (p. 138, italics in the original). Second, he disputes the empirical basis of the hypothesis of the constitutive role of research performance indicators. "This basis is in the perception of the current author still rather weak, and not seldom based on personal considerations and informal evidences." (p. 138). The third comment is that constitutive effects should be seen in an historical context. If there are constitutive effects of indicators, "one should not point the critique on current assessment practices merely towards informetric indicators as such, but rather towards any claim for an absolute status of a particular way to assess quality" (p. 138). In Moed's view, it does not follow that an intelligent combination of peer judgement and indicators is fundamentally flawed as long as one does not attribute an absolute status to either one of the main assessment methods.

Of course, Moed's last argument is correct, constitutive effects are not limited to informetric indicators. But who would dispute this? I do not meet many people who wish to lay the blame only on indicators and scientometricians. And most scientists as well as evaluators prefer to combine indicators with human judgement. The debate about abuse of indicators is essentially a debate about a research assessment system *as a whole* that has gone awry. But scientometricians including Moed did play an important role in the shaping of this assessment system and they should therefore take responsibility for that role (De Rijcke & Rushforth, 2015). It is difficult to read Moed's third argument as a way of being accountable, it reads more as an attempt to deflect the critique. His first argument is not much stronger. Clearly, Moed seems to think that it is possible to have a world without constitutive effects of evaluation mechanisms, a rather fundamental misunderstanding. Apparently he does not recognize that the phenomenon of constitutive effect is only a specific instantiation of the more general phenomenon that social indicators change the reality they are purported to measure. Every social scientist who has been engaged in somehow measuring or studying parts of social reality has to deal with this (Berger & Luckmann, 1967; Bijker, Hughes, & Pinch, 1989; Hacking, 1999; Porter, 1995; Schinkel, 2013). It has nothing whatsoever to do with magical thinking.

A distorted and narrow view of social science is also built into Moed's second comment on the concept of constitutive effects. He doubts the empirical evidence for the occurrence of constitutive effects. That is a rather bold claim to make given the fact that Moed does not give any indication that he attempted to process the relevant literature. The suggestion that the evidence is "informal" and "based on personal considerations" is itself made without any empirical evidence. Of course, the study of the effects of, for example, citation indicators, cannot itself be purely informetric and a host of qualitative methodologies and theories are needed, ranging from ethnographic observations, surveys, focus groups, to interviews, often in combination with formalized qualitative and quantitative approaches (Fochler & De Rijcke, 2017; Gläser et al., 2010; Gläser & Laudel, 2016; Hammarfelt & De Rijcke, 2015; Hammarfelt & Rushforth, 2017; Kaltenbrunner & De Rijcke, 2016; Laudel & Gläser, 2006; Sauder & Espeland, 2009; Whitley, 2010; Wouters et al., 2015). Informetricians and bibliometricians

Please cite this article in press as: Wouters, P. The failure of a paradigm. *Journal of Informetrics* (2017), https://doi.org/10.1016/j.joi.2018.03.002

could learn a lot from this type of studies. For example, the journal impact factor plays different symbolic roles in different contexts which explains a good part of its popularity (Rushforth & De Rijcke, 2015). The suggestion that these approaches can be summarized under the heading of "informal evidence" is a poor presentation indeed. I will come back to this in the conclusion of this review.

The last section of Chapter 9 states that ad-hoc arguments of evaluators or unreflective assumptions about indicators can play a role in research assessment because of the lack of an evaluative framework and an assessment model. The solution to these problems in research assessments is therefore the systematic development and application of such frameworks. In the following chapters of Part IV Moed sets out a number of proposals for moving forward which should clarify how such evaluative frameworks should look like. In a recent blogpost, Moed has tried to further clarify the concept (Moed, 2018). He does not argue in favor of one or other ready-made frameworks. Rather, he proposes that for each new evaluation exercise, the organizers of the assessment create such a framework by deciding on, in this order: the unit of assessment, the dimension of research that should be assessed, the objective of the assessment, and the relevant characteristics of the unit of assessment. Using the multi-dimensional indicator matrix, one can then decide which indicators should be used. Hence, the conclusion that an evaluative framework should be central is a methodological guideline, not a substantive one. The strength of this proposal is that the participants in a research assessment are forced to think hard about what they actually want to reach. This is a good recipe against the unreflective use of standardized indicators. At the same time, Moed's proposal for evaluative frameworks suffers from a major flaw. It disregards the way processes of attributing value, of which research evaluation is an exemplar, are deeply embedded in social life and assumes it is possible to neatly separate research evaluation from the other work processes in academic research. For example, research evaluation does not only take place when an expert panel visits a research group in an academic medical hospital in the Netherlands in the framework of the Dutch Standard Evaluation Protocol requirements. Evaluation takes place *all the time*: in the process of hiring a PhD candidate, in the decision to submit to journal A instead of to journal B, in the decision to submit a proposal to one call rather than another, and in career choices of senior researchers. And these evaluative moments influence each other. Knowledge creation and research evaluation have become so entangled that the idea to neatly separate them is almost laughable. In the end, such an endeavor remains a technocratic solution to a deeply social problem.

In Chapter 10, Moed launches a number of proposals that exemplify how his idea of evaluative frameworks can be applied. Among them are indicators of communication effectiveness, indicators for the various functions of journals, and indicators of research training and the contribution institutions make to research in developing countries. In addition, Moed suggests to define minimum performance standards as a criterion to decide which staff members can be considered research active. According to Moed, this would entail a shift from measuring performance to measuring the preconditions of performance. The chapter includes a section on university rankings with proposals for universities in which, to be honest, I did not find much new. And last, the chapter provides an alternative to performance based funding in which emerging groups are put central, rather than established research groups which is currently commonly the case. These are all interesting proposals that may fit in research assessments depending on the context and goals of the evaluation. I do not see, however, in what ways they would contribute to a systemic change in the way we evaluate research and researchers.

I had the same feeling about the proposals in the last chapter of Part IV of the book, which presents four ideas that have also been published as separate articles by the author: indicators of the peer review process of journals and books; an ontology based data management system; informetric self-assessment tools; and informetric models of scientific development. All of them are sensible ideas, some of them are already in development at publishers and universities in a variety of projects (e.g. the ACUMEN portfolio is a suite of self-assessment tools in the making) and some of them are more research projects rather than tools for assessment (such as informetric models of scientific development). In other words, Chapters 10 and 12 of the book present valuable ideas but it is not clear what binds them together and why these proposals, rather than others, are selected for this overview. I guess what I miss here is an overarching vision. Instead we get a set of technical ideas that may work, or may not work.

6. Can informetrics be value free?

Let me be clear, I do not think this lack of overarching vision is coincidental. Rather it is the consequence of the perspective that Moed defends in his book. As I indicated, the book is not only a state of the art perspective as perceived by the author, but also an intervention in the current debate about research assessments. It tries to make the case for a foundational paradigm in informetrics which has guided many bibliometricians and scientometricians in the past (Van Raan, 1988). Hence, Moed does not only defend his own position, his book is essentially the expression of this paradigm. Moed states it succinctly in the introduction to the book: "The current author conceives informetrics as a value free, empirical science. Being value free is conceived as a *methodological* requirement. This book shows how statistical definitions of indicators of research performance are based on theoretical assumptions on what constitutes performance. The informetric component and the domain of evaluative or political values in an assessment are disentangled by distinguishing between quantitative-empirical, informetric evidences on the one hand, and an evaluative framework based on normative views on what constitutes research performance and which policy objectives should be achieved on the other.¨(p. 12) This means that informetricians should maintain a neutral position in their informetric work.

For Moed, values play a role in three different ways. First, they are important in the choice of an evaluative framework. Informetricians should not be involved in this since these choices cannot be based on informetric evidence. Second, infor-

Please cite this article in press as: Wouters, P. The failure of a paradigm. *Journal of Informetrics* (2017), https://doi.org/10.1016/j.joi.2018.03.002

metricians are also researchers subject to assessment and as such they will also have opinions but they should separate these from their role as informetric technician. Thirdly, every indicator has some normative views built-in and in this sense no indicator is truly value-free. But if indicators are constructed on values, how can one imagine a value-free informetrics? Moed answers this conundrum by pretending that, although all indicators are infested with values, it is still possible to analytically separate value from fact.

Moed's evaluative framework works like a cookbook and the informetrician is the neutral cook who prepares the food but abstains from judgement about its health. In a contrived world this might work. But such worlds do not exist nor will they ever come to exist. To frame it in more theoretical terms, the paradigm that Moed adheres to is based on a triple purification of the complex reality of the lifeworld in which we exist (Latour, 2005). First, it tries to separate the technical characteristics of indicators from their value-laden dimensions. Second, it tries to insulate the process of research assessment from all the other work processes in which these assessments are embedded. I already argued why this is fallacious. For example, suppose a hospital would follow Moed's advice and develop an ealuative framework for its formalized research assessment exercises. It would still be confronted with the *hundreds* of evaluation decisions that have to be taken *every day* in the course of the daily work of its researchers. Third, Moed's paradigm tries to ignore the formative effects of evaluations in general and informetric tools in particular on the character of the process of knowledge creation by narrowly focusing on "performance enhancement" and ignoring the huge variety of social effects (not at all necessarily detrimental by the way) on the research system and on individual research groups. As I already indicated, the processes of research assessment are intimately intertwined with the processes of agenda setting, funding, recruitment, project management, infrastructure development and maintenance, and human resource management. An informetrics that is oblivious to these connections and the feedback processes that result from them, and which does not digest the state of the art knowledge of the social sciences and humanities, will always be blind to its own role in society and the intended and unintended effects it will generate.

Moed's book is testimony to this. It is a book that shows nothing less than the failure of a paradigm. And this is, paradoxically, an important contribution to the field of informetrics.

Acknowledgements

I would like to thank Thomas Franssen, Thed van Leeuwen, Sarah de Rijcke, and Ludo Waltman for comments on an earlier version of this review.

References

Berger, P. L., & Luckmann, T. (1967). *The social construction of reality: A treatise in the sociology of knowledge*. Doubleday.
Bijker, W. E., Hughes, T. P., & Pinch, T. (1989). *The social construction of technological systems*. Cambridge, Mass: The MIT Press.
Boonstra, O., Breure, L., & Doorn, P. (2004). *Past, present and future of historical information science*.
Borgman, C. (2007). *Scholarship in the digital age*. Cambridge: MIT Press.
Bornmann, L., & Haunschild, R. (2017). Do bibliometrics and altmetrics correlate with the quality of papers–A large-scale empirical study based on F1000 prime, altmetrics, and citation data. In *STI 2017. open indicators: innovation, participation and actor-based STI indicators*. [Paris].
Bulger, M., Meyer, E. T., Flor, G. de la, Terras, M., Wyatt, S., Jirotka, M., et al. (2011). *Reinventing research? Information practices in the humanities*.. Retrieved (https://www.rin.ac.uk/our-work/using-and-accessing-information-resources/information-use-case-studies-humanities)
Costas, R., Zahedi, Z., & Wouters, P. (2015). Do altmetrics correlate with citations? extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, *66*(10), 2003–2019.
Dahler-Larsen, P. (2012). *The evaluation society*. Stanford University Press. Retrieved August 9, 2013. http://www.amazon.com/The-Evaluation-Society-Peter-Dahler-Larsen/dp/080477692X
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. http://dx.doi.org/10.1002/asi.23382. Retrieved February 20, 2018
European Commission Expert Group on Assessment of University-Based Research. (2010). *Assessing europe's university-based research – Expert group on assessment of university-based research*. European Commission.
Fochler, M., & De Rijcke, S. (2017). Implicated in the indicator game? An experimental debate. *Engaging Science, Technology, and Society*, *3*, 21–40.
Gläser, J., & Laudel, G. (2016). Governing science. *European Journal of Sociology*, *57*(1), 117–168. Retrieved (http://www.journals.cambridge.org/abstract_S0003975616000047)
Gläser, S., Lange, G., & Laudel, U. (2010). Informed authority? the limited use of research evaluation systems for managerial control in universities. In R. Whitley, J. Gläser, & L. Engwall (Eds.), *Reconfiguring knowledge production: Changing authority relationships in the sciences and their consequences for intellectual innovation* (pp. 149–183). Oxford: Oxford University Press. Retrieved (http://www.worldcat.org/title/reconfiguring-knowledge-production-changing-authority-relationships-in-the-sciences-and-their-consequences-for-intellectual-innovation/oclc/548626398&referer=brief_results)
Hacking, I. (1999). *The social construction of what*. Harvard University Press.
Hammarfelt, B., & De Rijcke, S. (2015). Accountability in context: effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the faculty of arts at uppsala university. *Research Evaluation*, *24*(1), 63–77.
Hammarfelt, B., & Rushforth, A. D. (2017). Indicators as judgment devices: Citizen bibliometrics in biomedicine and economics. *Research Evaluation*, 1–25. Res Eval rvx018.
Hey, T., Tansley, S., & Tolle, K. (2009). The fourth paradigm: Data-Intensive scientific discovery. In T. Hey, S. Tansley, & K. Tolle (Eds.), *Microsoft research*. Retrieved (http://www.amazon.com/dp/0982544200)
Hunter, D. E. K., & Nielsen, S. (2013). Performance management and evaluation: Exploring complementarities. *New Directions for Evaluation*, *2013*(137), 7–17. Retrieved February 8, 2018 (http://doi.wiley.com/10.1002/ev.20042)
Kaltenbrunner, W., & De Rijcke, S. (2016). *The micropolitics of quantifying research output for evaluation in dutch law schools. pp. 1–7 in STI 2016: peripheries, frontiers and beyond*. Valencia, Spain.
Kaltenbrunner, W. (2014). *Reflexive inertia. Reinventing scholarship through digital practices*. Leiden University.
Konkiel, S. (2016). Altmetrics: Diversifying the understanding of influential scholarship. *Palgrave Communications*, *2*, 16057. Retrieved http://www.palgrave-journals.com/articles/palcomms201657

Please cite this article in press as: Wouters, P. The failure of a paradigm. *Journal of Informetrics* (2017), https://doi.org/10.1016/j.joi.2018.03.002

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford: Oxford University Press.
Laudel, G., & Gläser, J. (2006). Tensions between evaluations and communication practices. *Journal of Higher Education Policy and Management*, *28*(3), 289–295.
Moed, H. (2002). The impact-factors debate: The ISI's uses and limits. *Nature*, *415*(6873), 731–732. http://dx.doi.org/10.1038/415731a. Retrieved
Moed, H. F. (2005). *Citation analysis in research evaluation*. Dordrecht: Springer.
Moed, H. (2018). *Responsible use of metrics: An evaluative framework is essential, but cannot be grounded in informetric research*. The Bibliomagician. Retrieved February 20, 2018 (https://thebibliomagician.wordpress.com/2018/02/06/in-responsible-use-of-metrics-an-evaluative-framework-is-essential-but-cannot-be-grounded-in-informetric-research/)
Nielsen, M. (2011). *Reinventing discovery: The new era of networked science [Hardcover]*. Princeton University Press. Retrieved January 10, 2012 (http://www.amazon.com/gp/product/0691148902/ref=as_li_tf_il?ie=UTF8&tag=michaniels-20&linkCode=as2&camp=217145&creative=399373&creativeASIN=0691148902)
Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press. Retrieved (http://books.google.nl/books?id=oK0QpgVflN0C)
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. Retrieved (http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4469321&tool=pmcentrez&rendertype=abstract)
Sauder, M., & Espeland, W. N. (2009). The discipline of rankings: Tight coupling and organizational change. *American Sociological Review*, *74*(1), 63–82.
Schinkel, W. (2013). The imagination of 'society' in measurements of immigrant integration. *Ethnic and Racial Studies*, *36*(7), 1142–1161. Retrieved (http://www.tandfonline.com/doi/abs/10.1080/01419870.2013.783709)
Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other candidates. *Public Library of Science*, *8*(5), e64841.
Thelwall, M. (2005). *Link analysis: An information science approach*. San Diego: Academic Press.
Thelwall, M. (2008). Bibliometrics to webometrics. *Journal of Information Science*, *34*, 605–621 [Retrieved October 24, 2011 (http://jis.sagepub.com/cgi/doi/10.1177/0165551507087238)
Van Raan, A. (Ed.). (1988). *Handbook of quantitative studies of science and technology*. Amsterdam: Elsevier Science Publishers.
Waltman, L., & Traag, V. A. (2017). *Use of the journal impact factor for assessing individual articles need not be wrong.*. Retrieved (http://arxiv.org/abs/1703.02334)
Whitley, R. (2010). Reconfiguring the public sciences the impact of governance changes on authority and innovation in public science systems. In R. Whitley, J. Gläser, & L. Engwall (Eds.), *Reconfiguring knowledge production: Changing authority relationships in the sciences and their consequences for intellectual innovation* (pp. 3–47). Oxford: Oxford University Press. Retrieved (http://www.worldcat.org/title/reconfiguring-knowledge-production-changing-authority-relationships-in-the-sciences-and-their-consequences-for-intellectual-innovation/oclc/548626398&referer=brief_results)
Wouters, P., Beaulieu, A., Scharnhorst, A., & Wyatt, S. (2013). Virtual knowledge. In P. Wouters, A. Beaulieu, A. Scharnhorst, & S. Wyatt (Eds.), *Experimenting in the humanities and the social sciences*. Cambridge USA: MIT Mass Press. Retrieved (http://mitpress.mit.edu/books/virtual-knowledge-0)
Wouters, P., Bar-Ilan, J., Thelwall, M., Aguillo, I. F., Must, Ü., Havemann, F., et al. (2014). *ACUMEN final report*.. Brussels. Retrieved (http://cordis.europa.eu/result/rcn/157423_en.html)
Wouters, P., Thelwall, M., Kousha, K., Waltman, L., De Rijcke, S., Rushforth, A., et al. (2015). *The metric tide: Literature review (Supplementary report I to the independent review of the role of metrics in research assessment and management)*. pp. 188. Retrieved (http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/)

Paul Wouters
Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands
E-mail address: p.f.wouters@cwts.leidenuniv.nl
Available online xxx

Please cite this article in press as: Wouters, P. The failure of a paradigm. *Journal of Informetrics* (2017), https://doi.org/10.1016/j.joi.2018.03.002