Journal of Informetrics xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of Informetrics

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

Towards a multi-paradigmatic, value free informetrics: A reply to Paul Wouters' book review "The failure of a paradigm"

1. Preamble

In this contribution I present my comments on the most important statements in Paul Wouters' review "The failure of a paradigm" (Wouters, 2018) of my book "Applied Evaluative Informetrics" (Moed, 2017), statements that seem often based on a mismatch between the book's intentions and Wouters' expectations, but, fortunately, some of which properly mark essential differences in view between his and my own thinking. I focus on the three key aspects expressed in the book's title, namely that its subject is informetrics; that it deals with its application in research assessment; and that it gives special attention to its evaluative dimension. This contribution ends with notes on the way forward and concluding remarks. But I start with a few comments about the book's intended audience

2. The book's intended audience

The book is directed towards a broad academic audience, not merely to the research community in the field of bibliometrics, informetrics, quantitative science studies, or social studies of science. I have deliberately chosen to provide broadness and background, and to leave out technical details. This is not so much a matter of taste or style, but the consequence of the notion that it is a matter of social responsibility that specialists in the field seek to explain to the "outside world" in an understandable manner what they are doing and talking about in their professional life. My monograph aims to give such an explanation.

My reference group consists of the users of indicators for which I have conducted bibliometric studies, including policy officials and peer committee members; academic librarians who have to reply to questions about bibliometric methods by their customers and superiors; employees from the information industry, especially from academic publishers (indeed, I worked for four years with the scientific information company Elsevier). Last but not least, I also took into account my experiences in discussions with numerous other persons, academic or otherwise, in universities, but also at parties and on the sports field, about technicalities, pros and cons, potential and limits of informetric indicators as tools in research assessment.

Paul Wouters is disappointed that the book does not present more details on a series of complex issues. Given the composition of my reference group and intended reading audience I have decided to be selective in my overview of the activities and achievements concerning such issues. The book is for the greater part based on about 10 full research articles published by myself and co-authors during 2014–2017, all cited in the book's introduction chapter and in its reference list. There is a huge literature on these issues. More detailed information can be found in the specialist journals in our field, and also in a new Handbook on Science and Technology Indicators (Glanzel et al., 2018) that I am currently co-editing with three distinguished colleagues, in which competent authors give detailed accounts of some 40 topics in the field, and that will be published in the course of 2018.

3. The informetric dimension

The book focuses on the role of informetrics in research assessment. The use of the term informetrics rather than bibliometrics in the book's title marks an important development not only in the field of quantitative science studies, but virtually in all domains of science and scholarship, and in society at large, namely the computerization or digitalization of information and communication. The book aims to adopt a broad perspective on this development as well. It does not only show how new developments in ICT and computational techniques facilitate the creation of new databases, indicators and assessment methodologies. It also aims to show how these developments fit into trends in society as a whole. After all, there is desktop bibliometrics, but, for instance, also desktop medicine, and desktop notary, etc. The broadness and pervasiveness of these developments ask for a more general view on their meaning and potential, a view that reaches far beyond that of quantitative

https://doi.org/10.1016/j.joi.2018.03.003
1751-1577/

research assessment. I believe that a presentation of the main lines of the creative thinking of Michael Nielsen is an excellent way to illustrate what such a view could look like.

The book is based on the firm belief that debates about the pros and cons, potential and consequences of informetric artefacts and methodologies, must be based on a sound knowledge of their informetric 'substance'. As further discussed below, a key objective of the book is to show how technical, informetric indicator concepts are influenced by non-informetric, possibly political, values. This can be properly and convincingly achieved only by technically de-composing an informetric construct into its main components, and pointing towards the 'part' – or 'connection' between parts – in which these values actually assert their influence, so that it becomes clear to an outsider that they really do matter. The chapter comparing five university ranking systems, for example, provides along these lines a clear illustration of how technical indicator normalizations affect the position of universities in a ranking.

4. The book's applied character

The book deals with the *application* of informetric techniques in research assessment. If it deserves the qualification scientific, it is *applied* science, but one that is fully aware of the need of a theoretical foundation of assessment methods and practices. The book does not give a comprehensive research agenda in the field of social studies of science. Nor does it aim to continue the technical debates on informetric issues taking place in the specialist journals in our field.

It aims to create a certain distance from – or reflection upon – current assessment practices, not only by discussing their limitations and pitfalls, but also by showing that assessment methods could be *different* from what they are now. The book's message is: *current* practices are *not* necessarily the most appropriate way to use informetric/bibliometric indicators. Alternative approaches are at least thinkable *and* technically feasible. In this way the book seeks to create openness towards the future, and to make a further step in disclosing the potential of the informetric method.

An author who claims that informetrics itself does not evaluate, and that actual assessments should be guided by an essentially extra-informetric, evaluative framework, should be cautious when expressing his view on what such a framework should look like, because there is a real danger that this would direct the attention too strongly towards his personal views rather than to the claim of the need of such a framework as such. This is why the various proposals that I make in the book are presented as 'possible applications'. I give as it were the desirability or social value of these applications a *hypothetical* status, show how they could be shaped and organized.

5. Facts and values

For many years I have been actively involved in debates about bibliometric indicators, during which I more and more realized that in seemingly technical debates, political or policy considerations influenced the functional form of indicators. I included an entire chapter (Chapter 7) to illustrate this by means of a series of concrete examples, and I want to invite all colleagues in the field – and of course all interested outsiders – to read it. It also shows how indicator concepts mirror the political context in which they are developed. This is an important observation, because it makes us aware and warns us that indicator development is not only influenced by the interests of the information industry – we experience this directly almost on a daily basis –, but also, at a distinct level, by the interests and preferences of the developers themselves: they are not merely producers, but at the same time also potential or actual objects of research assessment based on their indicators.

In the book I do *not* adopt a constructivist viewpoint, and, hence I do *not* argue that my observations prove that indicator concepts and, as a result, numerous assessment processes all over the world, are being *determined* by the personal experiences, and political and business strategies of a group of indicator developers in informetric academic institutes or companies. Yet, I believe my observations bear relevance not only to the wider research community, but also to practitioners in our field, as they represent an incentive to *reflect* upon our informetric activities. And this is precisely what I aim to do in my book.

My position is that evaluative informetrics, i.e., the use of informetric techniques in the assessment of research, does itself not evaluate, and I defend the notion of a value free informetrics – but also a value free social studies of science – in the following sense. First I state what it does *not* mean. Value free science does *not* mean that it cannot study values as object of research; Robert Merton's studies of the norms of science are a point in case. It does *not* mean that there are no values underlying theoretical concepts; social science using concepts the formation of which is fully independent of societal or cultural values is hardly thinkable; for instance, there would be no economical science. It does *not* mean that the application of research findings in technology and in society at large is value free; on the contrary, bibliometric indicators are used to achieve political or managerial goals. Value free does certainly *not* mean that the scientific practice is not guided by a series of methodological values or rules. It does *not* mean that investigators are as it were robots who can completely switch off their value perceptions and historicity. It acknowledges that in social sciences investigators are not the type of external observers they are in the natural sciences.

What then *do* I mean with a value free informetrics? The book states: "Evaluation criteria and policy objectives are not informetrically demonstrable values. Informetric research may study such values empirically, but cannot provide a theoretical foundation of the validity of the quality criteria or the appropriateness of policy objectives. Informetricians should in their informetric work maintain a neutral position towards these values (p. 20)". "The methodological requirement of a value free informetrics does not mean that informetricians are not 'allowed' to have these views, nor that they are not

Please cite this article in press as: Moed, H.F. Towards a multi-paradigmatic, value free informetrics: A reply to Paul Wouters' book review "The failure of a paradigm". *Journal of Informetrics* (2017), https://doi.org/10.1016/j.joi.2018.03.003

allowed to communicate their views to the outside world. But they should make such views *explicit*, and make at the same time clear that these are their own views that are *not* informetrically *demonstrable"* (p. 13).

The latter is proposed as a *methodological* requirement for practitioners in informetric research. It is a requirement defined at a "meta" level that constitutes or regulates the research practice at the "object" level. And this is precisely the requirement I comply with in my analysis of a series of indicator concepts. Rather than striving for consensus in a self-contained debate about the validity of indicators, technical solutions and the evaluative assumptions underlying it, informetricians should make their value assumptions as explicit as possible and communicate them to users, rather than tacitly making themselves decisions about their validity *for* them.

I am not claiming that researchers are fully transparent to themselves. Hence, they also have to keep *each other* sharp. I have in mind any category of values: not only political values, but especially assumptions as to what constitutes a notion as complex as research quality, which is, in itself, in my view, not a political issue. I want to underline that I focus in this book on informetrics – after all, this is my métier –, but the proposed methodological requirement applies to research in sociology of science and other social sciences as well. Providing insight to non-experts into the social values underlying his products is for an expert in any scientific-scholarly domain a meaningful way to exercise social responsibility, and *not* to deflect from it.

My position goes back to fundamental notions of Max Weber, and is based on later work of the German philosopher Hans Albert, who, inspired as he was by Karl Popper's concept of an open society, defended in a debate during the 1960s on the status of the social sciences the position that scientists can 'be free of any value judgement', even for research in the fields of values (Albert, 1965; "Hans Albert", n.d.; "Critical Rationalism", n.d.). I am not *separating* values from facts, but propose to move as it were the discussion about evaluative assumptions in informetrics to a distinct *analytical* level, with great implications for the way in which informetricians should behave towards – and communicate with – the outside user community.

6. Evaluation science

The book expresses the multi-disciplinarity and multi-paradigmatic nature of evaluation science, that asks for the development of an inter-disciplinary view, a task that cannot be easily achieved with practitioners who assign an absolute or preferred status to one particular approach or paradigm. The book definitely aims to raise the discussion beyond the framework of a single paradigm. Its message is: let us look broader than that, there are so many issues in business studies, educational science, and even technical science in which researchers struggle with the same type of problems as those informetricians encounter in their field.

I use the term evaluation science as a generic term, and not to indicate a research field in the sociological sense. Some practitioners use the term, others do not. There are some journals carrying the term in their title. Informetric, computational-linguistic analyses of full text literatures and advanced mapping techniques could provide a helpful tool to further depict the cognitive and social structures in this collection of activities, and in this way help to develop a kind of overarching view.

Reading Wouters' comments I come to the conclusion that the book should have mentioned and cited more studies dealing with what I term as evaluation science, including those that he cited in his review, especially those aiming to provide a social construction of science and scholarship. As a matter of fact, I am convinced that studies of the social construction of research quality can broaden our insight into assessment processes. But when a *constructive* approach turns into a *constructivist* one, claiming that social construction is the preferred if not the only approach towards valid knowledge on the functioning of science and scholarship, it is in my view deprived from most of its value.

7. Evaluative frameworks

The notion of an evaluative framework plays an important role in my book. It refers to a specification of the qualitative principles that provide guidance to a concrete assessment process. A core element in an evaluative framework is the specification of a performance criterion, in a set of propositions on what constitutes research quality or performance. In short, they state what is valuable. As argued above, the book defends the position that such values cannot be grounded in bibliometric/informetric research. From such propositions follow the indicators that should be used, and, in a next logical step, the data sources from which these are to be calculated.

Paul Wouters states that "the conclusion that an evaluative framework should be central is a methodological guideline, not a substantive one". I would say that guidelines that constitute an assessment process are essential to that process. He also argues that my proposal "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". I note that Wouters refers here to evaluators in general, not to informetricians in particular. My comments on this statement are as follows.

I do not claim that an evaluator can or should erase his own quality perceptions, experiences, and background. It is not true that the book, as Wouters states, "tries to insulate the process of research assessment from all the other work processes in which these assessments are embedded." What it does claim is that in a formal assessment, for instance, within the framework of a national research assessment exercise or individual assessment for promotion or hiring, a clear definition of the evaluation criteria logically precedes the choice of an indicator or methodology.

Please cite this article in press as: Moed, H.F. Towards a multi-paradigmatic, value free informetrics: A reply to Paul Wouters' book review "The failure of a paradigm". *Journal of Informetrics* (2017), https://doi.org/10.1016/j.joi.2018.03.003

In defining such criteria, the views based on the daily practices and social lives of the evaluators evidently do play a role. But at the same time evaluators have the task to maintain a certain distance or 'space' towards the object of assessment. It is not *merely* their *personal* view that provides a standard. If an assessment process claims to have any validity and legitimacy at all, the standard should be more objective and explicit. Especially in formal processes, I would hypothesize that the search for such a standard stimulates evaluators to use quantitative methods. I would argue that even though evaluative criteria and indicators do *not* coincide, the latter can help to explicate the former. This marks the potential of the indicators, and directs the focus towards the organization of assessment processes. If one would succeed in developing adequate tools to reach more objectivized assessment criteria, it would have strong implications for the way we evaluate research.

In a blog-post at the Bibliomagician I give the following example of an evaluative framework. "In an assessment process aiming to select from a set of early career scientists the best candidate for a tenured position, for me the important criteria would be: integrity, impartiality; creativity; open mindedness; and capability to reason at distinct analytical levels. These criteria form core elements of an evaluative framework to be used in this assessment. None of these can be assessed with bibliometric indicators, but require in-depth interviews, possibly informed by interview techniques. Evidently, when assessing professional competence, the ability to write and orally present would be important factors too. But citation-based indicators such as H index and journal impact factors are of little use to assess such aspects, because they were not developed as tools to assess early career scientists. Making a solid contribution to a paper in a good, specialist journal (to be established in an interview) would be more significant than a co-authorship in a multi-team paper published in a high impact factor outlet" (Moed, 2018, example 1).

Contrary to Paul Wouters, who evaluates my line of reasoning as a manifestation of a "*contrived*" worldview, I believe here lies a great common challenge for informetricians, qualitative students of science, policy officials and evaluators to develop such frameworks, in close interaction with the wider academic community. I can imagine that competing frameworks are being proposed. This is in itself not a problem, on the contrary, this would in my perception add value to the process. It would definitely *not* mark a failure as long as it is clear to all stakeholders that the discussion at stake is primarily one of the validity of the values and standards to be assessed, and that informetricians as informetricians, and social scientists as social scientists, do not have a preferred status in *this* discussion.

Paul Wouters seems to argue: one will never find such standards in this way. My reply is two-fold. Firstly, I believe to have participated in formal assessment processes in which a good first step was made towards the explication of an evaluative standard or framework – although the norm of objectivity and disinterestedness binds me to add that I am also aware of cases in which the formulation of such a standard seems to have failed. My second comment is that this type of reply may easily act as a self-fulfilling prophecy: the more people are made to believe it, the more probable it is that the realization of a more objective evaluative framework fails. I am sure Paul Wouters is aware of this.

8. About constitutive effects

In his concluding paragraph Wouters claims that my book "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 social effects (not at all necessarily detrimental by the way) on the research system and on individual research groups". Fortunately, Wouters gives several quotes showing that the book does not ignore these effects at all, and cites several of my own papers in which changes in researchers' behaviour under the influence of the use of particular indicators are studied in an informetric approach. But it is clear he does not agree with the way in which the book deals with them. The book states that solid research – sociological or otherwise – on these effects is important and urgently needed. But it makes a few critical comments.

In a very well written, stimulating essay by Peter Dahler-Larsen (Dahler-Larsen, 2013) cited in my book, an essay that deserved more attention than I have given to it, the author develops the view that the effects of the use of indicators are becoming so strong and pervasive, that they cannot be merely characterized in terms of "unintended consequences" that could perhaps be neutralized or repaired, as they increasingly change the way in which researchers and evaluators perceive *reality*. The latter type of effects is denoted as constitutive (Wouters uses the term formative). My own interpretation is that this view then suggests that the distinction between a theoretical concept and its operationalization is more and more erased from the minds of the people (the terminology is surely mine, not Dahler-Larsen's). This is for me the essence of the notion of constitutiveness of informetric indictors, and I use it in the book in this sense. I hope this not unfair towards Peter Dahler-Larsen, and if it is, I apologize for it.

My first comment is that the very notion of "gaming" with indicators often raised in debates on the use of informetric indicators, according to which researchers change certain behaviors in order to obtain higher ratings on particular quantitative measures, and so well described in studies of the social construction of research quality, implies that this process of ideology formation in the minds of researchers has not, or not yet, or at least not fully, occurred. In simple language: studies show that gaming researchers are very well aware of the fact that they are "just" gaming. Even managers may refer to bibliometric indicators not because they believe these are valid measures of research quality, but because of their publicity value. The fact that a university *manager* justifies in public the appointment of a particular professor selected in a peer-review based procedure by referring to the candidate's h-index, does in my view *not* necessarily mean that the value of this index was actually used as *primary* information, let alone as a decisive criterion in the selection process.

Please cite this article in press as: Moed, H.F. Towards a multi-paradigmatic, value free informetrics: A reply to Paul Wouters' book review "The failure of a paradigm". *Journal of Informetrics* (2017), https://doi.org/10.1016/j.joi.2018.03.003

A second comment, to which Wouters rightly dedicates a lot of attention, relates to the implications of the observation that the use of indicators *changes* research and evaluation practices. He quotes the statement from my book: "the issue is not whether scholarly practice is changed by indicators, but whether or not the application of indicators enhances research performance (p. 137)". I fully agree with Wouters that this raises the question as to how research performance is defined. I am very well aware that in an assessment of the validity of a set of indicators the use of the very same indicators necessarily leads to circular reasoning. My point is that my statement indeed marks the limits of the informetric-bibliometric approach, but *also* the limitations of a socially constructive approach seemingly so strongly advocated by Wouters.

A third comment holds that the issue of use and effects of indicators is more and more debated on a journalistic level in the daily and weekly press, including news sections of scientific journals, and in social media. I have no objections against this. But if a research article submitted to a scientific-scholarly journal claims the validity of the constitutiveness hypothesis using statements from journalists in the press as a primary source of evidence, I believe we are on the wrong track. This explains partly my reservations. But I am convinced that the papers cited by Wouters do not do such things.

Wouters quotes the following statement related to 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). I believe I should have expanded on this matter in the book. I have already explained why the fact that researchers change their practices is for me not a convincing argument pro the constitutiveness hypothesis as interpreted above. But I could have dedicated more time to more studies analyzing changes in behavior and reporting on researchers' motives to do so. I have respect for these studies and I believe they are useful.

Wouters claims that "Moed seems to think that it is possible to have a world without constitutive effects of evaluation mechanisms, a rather fundamental misunderstanding". Here he uses the term in a broader sense than the book does. Please allow me to refer to my proposal in Part IV for an alternative use of indicators, namely as tools in the formation of *minimum* quality standards. The book states: "Using indicators to define a baseline, researchers will most probably change their research practices as they are stimulated to meet the standards, but if the standards are appropriate and fair, this behavior will actually increase their performance and that of their institutions" (p. 120).

9. The way forward

In a chapter entitled "The Way Forward in Quantitative Research Assessment" the book questions what it identified as a base assumption of the application of bibliometric indicators in the 1980s, namely that "it is not the potential influence but the actual influence, not the importance but the impact that is most closely linked to the notion of scientific-scholarly progress, and that in an actual research assessment, it is not the importance but the impact that is of primary interest to policy makers" (p. 142). Next, it poses the question: "In *academic institutions*, especially research universities, it is generally considered appropriate to require academic staff to make contributions to scientific-scholarly progress. But is it defensible to require that they generate impact?" (p. 142). Finally, it proposes an *alternative* view, representing a shift from research quality or progress to communication effectiveness, and from performance as such to *preconditions* to performance.

Wouters argues against this proposal – and to virtually all other proposals in the book as well – that he does not see "in what ways they would contribute to a systemic change in the way we evaluate research and researchers". My reply is in three-fold. Firstly, Wouters apparently assumes that current assessment practices *must* change. This is clearly a *political* assumption. It is up to Wouters to explain what precisely must be changed and why. Next, a series of proposals in the book aim to illustrate the feasibility of an academic assessment model that does *not* aim to assess quality or performance as such, but to focus on whether the preconditions can be considered sufficiently favorable to expect research groups to perform well. In other words, in this model, viability would be a primary criterion, and production or impact *secondary*. The proposal to use informetric indicators as tools to define minimum standards fits into such a model. The book also proposes indicators that would facilitate the implementation of such a model better than currently used indicators such as journal impact factors do. Finally, how can Wouters be so certain that the application of such a model would not lead to "systemic" changes in the way academic research is evaluated? Or would its effect be in a politically 'wrong' direction?

10. Concluding remarks

I wish to thank Paul Wouters for his efforts to write his book review. Perhaps the most intriguing element of Wouters' review is its title. The failure of a paradigm! It is not clear to me which paradigm he is referring to. Is it perhaps the critical rationalist school of thought, providing an intellectual basis for the thesis that evaluative informetrics – and also the social construction of evaluation – do themselves not evaluate? In this case it is clear to me that Paul Wouters is a strong opponent to this paradigm, and I see why he believes it fails. He cites Bruno Latour supportively, who, as Harriet Zuckerman (2018) notes in her superb commemoration essay on Eugene Garfield, expressed in his book "Science in Action" the view that citations give citing authors the opportunity to "do whatever you need to the former literature to render it as helpful as possible" (Latour, 1987, p. 37), and who adds to this "where all deformations are fair" (p. 40).

Wouters also mentions, be it briefly, "a foundational paradigm in informetrics" that he believes my book is defending, and in relation to this he cites Van Raan's Handbook on Science and Technology Indicators, including a separate part on the study of "science in a social context". It is a great honor to be linked to Anthony van Raan, the founder of CWTS, of the first S&T conference, the first S&T Handbook, and my PhD supervisor! Is this the paradigm Wouters is referring to? Or is a

Please cite this article in press as: Moed, H.F. Towards a multi-paradigmatic, value free informetrics: A reply to Paul Wouters' book review "The failure of a paradigm". *Journal of Informetrics* (2017), https://doi.org/10.1016/j.joi.2018.03.003

multi-paradigmatic, value free informetrics for which the book pleas the failing paradigm? I would feel honored with the qualification "paradigm". It sounds so much more academic than "cookbook".

References

Albert, H. (1965). Wertfreiheit als methodische Prinzip. In E. Topitch (Ed.), *Logik der Sozialwissenschaften*. Koln, Berlin: Unknown.
Critical Rationalism, n.d. https://en.wikipedia.org/wiki/Critical_Rationalism.
Dahler-Larsen, P. (2013). Constitutive effects of performance indicators: Getting beyond unintended consequences. *Public Management Review*, *16*, 969–986.
Glanzel, W., Moed, H. F., Schmoch, U., & Thelwall, M. (2018). *Springer Handbook of Science and Technology Indicators*. Heidelberg (Germany): Springer. Forthcoming
Hans Albert, n.d. https://en.wikipedia.org/wiki/Hans_Albert.
Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.
Moed, H. F. (2017). *Applied evaluative informetrics*. pp. XXI+312. Springer. http://dx.doi.org/10.1007/978-3-319-60522-7 [ISBN 978-3-319-60521-0 (hard cover); 978-3-319-60522-7 (E-Book)]
Moed, H. F. (2018). *Responsible use of metrics: An evaluative framework is essential, but cannot be grounded in informetric research – guest post by Henk Moed*. https://thebibliomagician.wordpress.com/2018/02/06/
Wouters, P. F. (2018). The failure of a paradigm. *Journal of Informetrics* [to be published].
Zuckerman, H. (2018). The sociology of science and the Garfield effect: Happy accidents, unpredictable developments and unexploited potentials. *Frontiers Research Metrics and Analytics* [to be published].

Henk F. Moed
Sapienza University of Rome, Italy
E-mail address: henk.moed@uniroma1.it
Available online xxx

Please cite this article in press as: Moed, H.F. Towards a multi-paradigmatic, value free informetrics: A reply to Paul Wouters' book review "The failure of a paradigm". *Journal of Informetrics* (2017), https://doi.org/10.1016/j.joi.2018.03.003