

Bibliometrics: use and abuse in the humanities

Milena Žic Fuchs¹

University of Zagreb, Croatian Academy of Sciences and Arts, Croatia

Quo vadis bibliometrics?

The symposium *Bibliometrics: Use and Abuse in the Review of Research Performance*, held in Stockholm, 23–25 May 2013, fell within the context of an initiative to put an end to the misuse of JIF (journal impact factors), embodied in the San Francisco DORA (Declaration on Research Assessment) [1] instigated in December 2012 at the meeting of the American Society for Cell Biology. Interestingly, just before the start of the Symposium, reactions to DORA appeared in leading journals and news blogs, such as *Nature News Blog* [2], an editorial in *Science* [3], an editorial on research assessment in *The EMBO Journal* [4], an editorial in *Molecular Biology of the Cell* [5] and an editorial in *eLIFE* [6], to name but a few.

In a sense, DORA can be seen as a kind of culmination of decades-long discussions on bibliometrics, their use and abuse. Debates on the JIF also have a history of discussions as evidenced in leading journals as mentioned above, discussions *pro* and *contra* the JIF often mirroring divergent views (cf. e.g. [7]). In principle, the reactions to DORA can be said to support the main intent of the declaration, that is to address the over-reliance on JIFs in research assessment, as well as providing the basis for a change in the current system of research assessment prevalent in the so-called hard sciences.

It is of course impossible to go into the different nuances that the above-mentioned reactions to DORA stress. However, it should be noted that up to a certain point, they present a range of opinions and that they focus on different points of the declaration. One example is the view of the *Nature* [2], which has not signed DORA, and in which Philip Campbell, Editor-in-Chief of *Nature*, stresses that the group's journals had in the past published many critical views on the excessive use of JIFs. He also states the opinion that some statements in DORA are too sweeping, and thus does not merit the full support of DORA.

Although DORA can be seen as a move in the right direction, for if nothing else it has in a strong way initiated once again the big question of how to assess research performance, in order to assess quality and not peripheral indicators, open questions still remain. Would reducing the 'impact' of JIFs in favour of article-level ones be a true improvement in assessing article quality? Or, should the ultimate goal be evaluating research papers on the basis of their scientific merit? The ongoing discussions will, if nothing else, hopefully lay the foundations for a

¹Email: mzikfuch@ffzg.hr

possible new culture of research assessment, one primarily based on the content of articles, or simply put, one based on reading the papers themselves, as is clearly stated in the background motto of the Stockholm symposium, a citation of the 1991 Nobel Laureate Richard R. Ernst [8]: “Start reading papers instead of merely rating them by counting citations!”

DORA from the point of view of the humanities

One could expect that from a humanities perspective, the San Francisco DORA [1] would be welcomed in a ‘holistic’ sense. More specifically, it would be welcomed as a turning point which could bring humanities research assessment a step closer to what humanities scholars have been trying to achieve in articulating assessment criteria for the last couple of decades. A number of points from the declaration can be singled out in this respect, but with necessary additional elaborations in order to pinpoint the specific nature of humanities research outputs and how to evaluate them.

Thus, for instance, DORA underlines the following: “the need to assess research on its own merit rather than on the basis of the journals in which it is published in” [1]. If this is intended to mean a return to assessing scientific content, a return to peer judgement, then humanities scholars will endorse such a view. On the other hand, if the trend and intention of DORA is to introduce a citation-based metrics for assessing individual articles, then there is little, if any, support to be expected, for a number of reasons. First and foremost, such a metrics could cover only articles that appear in journals in, for instance, Thomson Reuters’ WoS (Web of Science) or Scopus. These databases cover the lesser part of the overall scholarly productions of the humanities in Europe. The multitude of articles published in national languages, in so-called national journals, could not be assessed by an article-based metrics, at least not without setting up sophisticated databases for all European languages, and their respective journals covering all the humanities disciplines.

Another point from DORA, one that would without a doubt be endorsed by humanities researchers, is the JIF ‘time frame’ of (usually) 2 years. Again, from a holistic perspective, the lifespan of research outputs in the humanities is by far longer than in other domains, not rarely extending over decades, and in some disciplines it even encompasses centuries. The time restriction that follows out of the JIF-led metrics mirrors neither the true nature of humanities research nor the weight of research results.

DORA also points to the necessity of considering a wider range of research outputs in research assessment: “for the purpose of research assessment consider the value and impact of all research outputs (including databases and software)” [1]. For the humanities, this is of special importance for the diversity of research outputs, which in many disciplines outnumbers or reflects different priorities in comparison with other domains of research. In the various ‘models’ of assessing research performance found for the humanities, much of this diversity is unfortunately not taken into consideration, or if it is, then priorities not inherent in different disciplines are postulated by universities’ funding bodies or national-level assessment rules and procedures.

Challenges of research assessment in the humanities

The previous section should be seen as only a brief overview that touches upon the complexity and difficulties encountered in research assessment in the humanities based on some points that appear in DORA. During the last couple of decades, we have been witnessing lively discussions on two interlinked issues, namely how to enhance the *visibility* of research outputs in the humanities and how to set up *evaluation systems*, systems that would reflect the specific nature of humanities research with regard to the diversity of outputs and the diversity of languages [9]. These discussions have shown that attempts at evaluating research performance on the basis of databases such as WoS and Scopus are lacking, particularly in the sense that they do not mirror in many cases the true quality and innovation that this research brings. Briefly, the coverage of humanities publications remains limited in WoS [10,11] or Scopus. Figure 1 gives an overall, clear visual insight into the coverage of the humanities and the social sciences in comparison with other domains [12].

Imposing assessment criteria inherent or derived from the above-mentioned databases ignores and, what's more, distorts insights into research achievements and the way they are publicized through specific outputs characteristic of the humanities. Namely, the specific nature of humanities research is in quite a few disciplines primarily reflected through publication of books, monographs and articles in *national languages*. In some disciplines, a key role is played by revised and commentated editions, collections of data, as well as translations as found in, for instance, philosophy, philology, etc. Such publications are not 'covered' by the databases that may be relevant for other domains of research. Sole reliance on them has already, through various evaluation systems

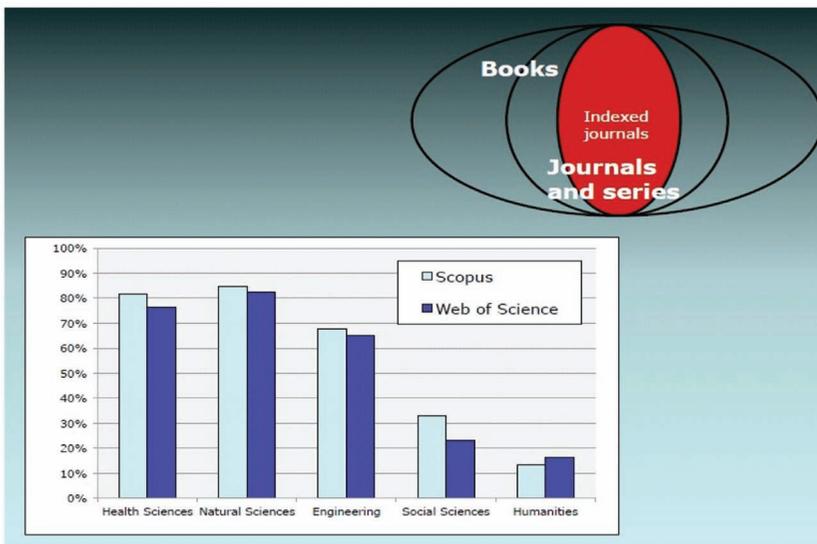


Figure 1

Coverage of Scopus and WoS in different scientific fields

(whether at university level or national level), produced changes in publication cultures in various national research environments throughout Europe. What has been evidenced, for instance, is a more pronounced tendency to publish in English, as shown for the region of Flanders, Belgium, as part of the Flemish performance-based funding system for university research. With respect to the analyses of humanities research outputs in the region of Flanders, Engels et al. [13] show that considerable differences have appeared across the disciplines of the humanities and social sciences in terms of coverage by the WoS. They also indicate a steady increase in the number and proportion of publications in English, with stress on the fact that they vary from discipline to discipline. The so-called English bias attributed to the influential WoS and Scopus is without doubt having impact on language usage in humanities research production, but changes in research agendas have also been noted. Thus within the domain of social sciences, more specifically for economics in Spain, a major shift to international journals indexed in WoS has been noted and analysed, showing that research topics more likely to be received by international journals have become a priority, often replacing local, regional and national topics [14].

Even in the bibliometric literature, one finds ‘simplifications’ pertaining to the above-mentioned issue of publishing in national languages as well as the audiences that they target. Nederhof [15], for example, identifies three audiences for the social sciences and humanities. First, as in the sciences, publications are directed to other scientists and scholars of the international research frontiers. Secondly, there is a separate audience of regional or national scientists or scholars. Here, use of a regional or national language, as well as regional or national publication media, and a focus on topics thought to be primarily of regional or national interest tend to curtail communication with colleagues elsewhere. Thirdly, there is the non-scholarly audience.

In the same vein, Diana Hicks [16] distinguishes four literatures for the humanities and the social sciences: first, international journals, secondly, national journals, thirdly, books and fourthly, non-scholarly press.

The problem with categorizations such as the above is that they view the humanities (and the social sciences) as in part adhering to the same publishing and outreach patterns most often found in other domains of research, as well as overlooking or glossing over in some cases major differences found in various humanities disciplines especially pertaining to their ‘international’ and ‘national’ audiences and subsequent relevance. Namely, the so-called national (or regional) publications directed at national (or regional) scholarly audiences, in disciplines such as literature and linguistics (philology), publications in Dutch, or any of the Slavic languages, to take just two examples of so-called ‘small’ European languages, are not only read by their respective national or regional scholarly audiences, but are also read by researchers and experts in these disciplines worldwide. Scholars in the philologies, apart from knowing the languages, and being able to read publications written in them, do not often perceive the topics covered as being of exclusively regional or national relevance. Talking to scholars in Dutch or Slavic studies in China, for example, one perceives that what might be ‘seen’ as a national publication, from their point of view, has a much wider relevance and an obviously wider reading public. The reverse also holds true: sinologists in Europe ‘read’, and

cite for their research, publications in Chinese, and this of course applies to all scholars doing research on ‘national philologies’.

It should also be noted that because of the so-called English bias, often impressive internationally first-ranking research traditions with outputs in French, German, Italian, Spanish, etc., are ‘overlooked’. In fact, we come across the phenomenon of English single-language authors who seem to be unaware of the huge body of knowledge within their own fields, for simple linguistic reasons.

The above indicates that a kind of ‘relativity of concepts’ exists as far as the distinction between what is ‘international’ and what is ‘national’ is concerned. It also shows that various disciplines in the humanities adhere to different publication practices, and different outreach, out of which subsequently follows that they may well have different types of ‘impact’, especially at the so-called international level. These practices do not mean that a bigger push for internationalization in the humanities is not welcome or should not be encouraged. In principle, this should be supported, but it should not be the sole criterion in research assessment, and what is ‘international’ should be more carefully and precisely defined.

Different disciplines also manifest different priorities in publication outputs: some lean more towards articles, whereas others consider monographs or books to be of the first priority. This brief overview of only some of the characteristics we find in various humanities disciplines confirm Henk Moed’s [17] claim that it does not make much sense to make statements about characteristics of the humanities and the social sciences as a whole. The present chapter also supports his view that much more research is needed in order to truly understand the complexities involved.

To make the whole ‘assessment landscape’ in the humanities even more complex, we are now faced with the advent of the so-called digital humanities, a new emerging area not only of research itself, or providing data for research, but also a new medium for ‘publishing’ research results, in formats that are a far cry from what is understood by traditional outputs. In June 2013, early-career researchers from the humanities met in Paris to discuss the issues involved in new ‘publication’ cultures emerging within the sphere of digital humanities. The result of this meeting is a manifesto entitled ‘Young Researchers in Digital Humanities: A Manifesto’ [18], a document which is an explicit appeal for new digital media, showcasing research and research results, to be included and properly evaluated in assessments for jobs, promotions, research funding, etc. It is hoped that new research outputs that come under this rubric will gain recognition, and they are very much in line with the already mentioned quote from DORA [1], “for the purpose of research assessment consider the value and impact of all research outputs (including databases and software).”

And last but not least, the issue of publishing research results which are outcomes of interdisciplinary research endeavours should at least be mentioned. With the advent of Horizon 2020, The Framework Programme for Research and Innovation (2014–2020), inter/multi/transdisciplinary research will become the focal point of all of the seven Grand Challenges (Societal Challenges as they are now called) envisaged in the documents that form the backbone of this new European funding programme. Inter/multi/transdisciplinarity is inherent in the Grand Challenges, as well as in the ‘focus areas’ they cover, whether it be person-

alized medicine, sustainable food security, energy efficiency, water innovation or climate change. Active participation of humanities and social sciences scholars is a must if answers to the Grand Challenges are to be found and the ‘focus areas’ challenges are to be met. The in-depth intermeshing of scholars from all domains of research should result in new ‘cultures of research’. *Mainstreaming* is the term most often used to indicate these new trends, and the fact that the role of the humanities and the social sciences should have their rightful place, and not simply be ‘icing on the cake’. Fully fledged engagement of the humanities and the social sciences should be implemented from start to finish.

In preparation for more integrated approaches to interdisciplinarity, the ESF (European Science Foundation) organized, in the summer of 2012, an event for early-career researchers entitled ‘Interdisciplinary Junior Summit, Water: Unite and Divide. Interdisciplinary approaches for a sustainable future’ [19]. The event was instigated by the Standing Committee for the Humanities of the ESF, and became a joint effort together with the Standing Committees for Life, Earth and Environmental Sciences, Physical and Engineering Sciences, and the Social Sciences. Without going into details of the event itself, this was an event that brought together 36 young researchers from different domains of science (from all over Europe), and a special issue portraying the Junior Summit on Water was published in JWARP (*Journal of Water Resource and Protection*), an international peer-reviewed open access journal [20]. This issue reflects the enthusiasm and synergy that resulted from the event itself.

However, the young researchers, although impressed by the potentials of interdisciplinary research (for most of them a completely new experience), expressed concerns as to where they could publish results from such research, and what is more, they expressed fears that interdisciplinarity implies publishing in journals with low impact factors. Needless to say, these concerns were directly related to possible negative effects on promotions and job applications. Thus an important ‘Grand Challenge’ for future models of research performance assessment will be how to evaluate interdisciplinary articles as well as other outputs, whose number can be expected to grow within the context of Horizon 2020. It would indeed be a huge drawback if the kind of metrics dominant at present were to block the possible innovative results of this kind of research, and this does not only pertain to the humanities and social sciences, but to all domains of research.

Research performance assessment in the humanities: how to deal with the challenges

The preceding sections have tried to outline, at least in broad strokes, the diversity we find in humanities research: diversity of outputs, languages and research traditions. Within such a context, it is truly difficult to set up research assessment models, and debates on how to do this have been going on for quite some time. These debates often mirror opposed views between, on the one hand, different humanities research communities throughout Europe, and on the other, funders and evaluation committees. Opposing perspectives can be summed up, at least in rough terms, into major approaches to research assessment. Thus we encounter

a strong voice that is against any kind of ‘bibliometric evaluation’, and strongly advocates reliance on peer judgements as the basis of research assessment. Conversely, equally strong arguments can be heard for taking humanities to a higher, more formalized, ‘bibliometric level’.

These opposed views are also especially visible when we view reactions to the ERIH (European Reference Index for the Humanities), launched in 2001 by the Standing Committee for the Humanities of the ESF. Although the main aim of ERIH was from its very beginnings to enhance *global visibility* of high-quality research in the humanities published in academic journals in various European languages across all of Europe, in some circles, it was initially perceived as a ranking mechanism, hence an evaluation tool. What is more, in some countries, it actually was used for assessing research performance after the so-called Initial Lists were published in 2007 and 2008. The use of ERIH as an ‘evaluation tool’ primarily lies in the categorization of journals into A, B and C according to the following ‘definitions’ in shorthand:

- Category A: high-ranking international journals with high visibility that are regularly cited all over the world.
- Category B: standard international journals with significant visibility and influence.
- Category C: high-ranking national or regional journals with recognized scholarly significance.

The ERIH Steering Committee, as well as the Standing Committee for the Humanities, considered feedback on the A, B and C category names, often perceived as a *ranking*, and in order to avoid further misuses introduced new category names as well as giving *national journals* major prominence.

- National journals (or NAT): European publications with a recognized scholarly significance among researchers in the respective research domains in a particular (mostly linguistically circumscribed) readership in Europe.
- International journals (or INT1 and INT2): both European and non-European publications with an internationally recognized scholarly significance among researchers in the respective research domains, and which are regularly cited worldwide.

The revised ERIH lists, with new category names, were published in 2011 and 2012 and were preceded by the following warning [21]:

“More specifically it is not intended as bibliometric information for use in assessment processes of individual candidates, be it for positions, promotions, research grant awards, etc.”

Nevertheless, misuses and the abuse of ERIH, especially in national evaluation systems, have persisted in the sense of the two international categories (INT1 and INT2) having been more highly valued than the national category. Thus articles published in national languages were deemed of lesser importance and value, and a strong English-language bias was enforced in direct opposition

to the main aim and intent of ERIH as a visibility tool for humanities research in Europe. With the winding down of the ESF (planned for the end of 2015) the main concern of the Science Review Group for the Humanities (the successor of the Standing Committee) is to ensure that ERIH remains a visibility mechanism for humanities research in the future.

During the last decade or so, a growing awareness of the importance of monographs and books, as well as chapters in volumes has even been recognized by the big commercial databases primarily used in research performance assessments. However, the main inclination, as is the case with journals, is to include books published by leading publishing houses worldwide and in principle this again means that we are faced with a strong English-language bias. As stated above, the inclusion of books, monographs, revised edition, collections of data, etc. must be included in the evaluation of research performance in the humanities. This issue should be addressed at both the national level as well as at the European level, and possibly in the future the importance of books may well reappear in the research assessment of other domains of research. In this respect, it is worth quoting the following from an editorial entitled ‘Back to Books’ published in *Nature* in 2010 [22]:

“The expansiveness of a book allows sophisticated arguments to be put forward and widely debated; new ideas that cross disciplinary boundaries can more readily be shared and worked through.

But if this exhortation is to have any traction, the effort and skill required to write a book needs to be rewarded in the career recognition of scientists who devote time to mastering the art to good effect, a recognition that is commonplace in the social sciences and humanities. It is time to bring the book back to the science mainstream.”

At present in Europe there are a number of endeavours aiming at solving the national language issue or challenge with respect to the diversity of research outputs in the humanities. Two stellar examples are, probably the best-known database of this kind, the Norwegian CRISStin (Current Research Information System in Norway) [23,24], as well as the VABB-SHW, a full-coverage database of peer-reviewed publication outputs in the humanities and social sciences developed for the region of Flanders as part of the Flemish performance-based funding system for university research [25,26]. Databases of these kinds can be seen as a way forward in encompassing the full diversity of humanities research outputs, thus providing a healthy foundation for humanities research performance assessment. They also provide, which is of the utmost importance, insights into the changing publication cultures evolving [13].

The above examples are very much in line with what humanities scholars have been insisting on for quite some time, namely the importance of peer judgements. If ‘metrics’ has to be included in evaluation, then assessment should be *metrics-informed* and not *metrics-led*. The creation of national databases of journals, monographs, and other research outputs is the necessary first step in achieving ‘healthy’ and quality-oriented research performance assessments in the humanities. The ideal at some point, possibly in the distant

future, would be a European-level database encompassing all existing national databases for the humanities and the social sciences found in European research environments. This would ensure that all the types of diversities discussed so far would be reflected both at national and European levels, and it would ensure that the humanities would finally have a research assessment system in line not only with different languages and different outputs, but also with different research traditions in the various disciplines. Such an approach could also ensure ways to account for academic excellence in the true sense of the word and would at the same time provide a basis for transparency in research assessment.

References

1. San Francisco Declaration on Research Assessment (DORA). http://www.ascb.org/SF_declaration.html
2. Van Noorden, R. (2013) Scientists join journal editors to fight impact-factor abuse. In *Nature News Blog*. 16 May 2013, <http://blogs.nature.com/news/2013/05/scientists-join-journal-editors-to-fight-impact-factor-abuse.html>
3. Alberts, B. (2013) Impact factor distortions. *Science* **340**, 787
4. Pulverer, B. (2013) Impact fact- or fiction? *The EMBO Journal* **32**, 1651–1652
5. Bertuzzi, S. and Drubin, D.G. (2013) No shortcuts for research assessment. *Molecular Biology of the Cell* **24**, 1505–1506
6. Schekman, R. and Patterson, M. (2013) Science policy: reforming research assessment. *eLIFE* **2**, e00855
7. How to improve the use of metrics (2010) *Nature* **465**, 870–872
8. Ernst, R.R. (2010) The follies of citation indices and academic ranking lists. *CHIMIA International Journal for Chemistry* **64**, 90
9. Žic Fuchs, M. (2011) Humanities and social sciences at a crossroads? The bibliographic/bibliometric divide. ECOOM Colloquium: assessing research performance in the social sciences and humanities. <https://www.ecoom.be/sites/ecoom.be/files/downloads/4%20Milena%20Zic%20Fuchs.pdf>
10. Adams, J. and Tesla, J. (2011) Thomson Reuters Book Citation Index. The 13th Conference of the International Society for Scientometrics and Informatic, Durban, South Africa (Noyons, E., Ngulube, P. and Leta, J., eds), pp. 13–18, ISSI, Leiden University and University of Zululand
11. Archambault, E., Vignola-Gagne, E., Cole, G., Larivière, V. and Gingras, Y. (2006) Benchmarking scientific output in the social sciences and humanities: the limits of existing databases. *Scientometrics* **68**, 329–342
12. Sivertsen, G. (2011) Achieving complete and comparable data and output indicators for scholarly publishing in the humanities and social sciences. http://bibliometrie.univie.ac.at/fileadmin/user_upload/ub_bibliometrie/sivertsen.pdf
13. Engels, T.C.E., Ossenblok, T.L.B. and Spruyt, E.H.J. (2012) Changing publication patterns in the social sciences and humanities. *Scientometrics* **93**, 373–390
14. Carmona, S., García-Ferrer, A. and Poncela, P. (2005) From zero to infinity: the use of impact factors in the evaluations of economic research in Spain. *IE Working Paper*, WP05-22, 13-05-2005
15. Nederhof, A.J. (2006) Bibliometric monitoring of research performance in the social sciences and the humanities: a review. *Scientometrics* **66**, 81–100
16. Hicks, D. (2004) The four literatures of social science. In *Handbook of Quantitative Science and Technology Research* (Moed, H.F., Glänzel, W. and Schmoch, U., eds), Kluwer Academic, Dordrecht
17. Moed, H.F. (2011) Research assessment in social sciences and humanities. ECOOM Colloquium: assessing research performance in the social sciences and humanities, <https://www.ecoom.be/sites/ecoom.be/files/downloads/1%20Lecture%20Moed%20Ecoom%20Antwerp%209%20Dec%202011%20SSH%20aangepast%20%282%29.pdf>
18. Young researchers in digital humanities: a manifesto. https://docs.google.com/document/d/1rk1nfSqG5qqOoNSs9Q-LC3Km8glfaBP_tC4Q79-hT6E/edit?usp=sharing

19. Interdisciplinary Junior Summit, Water: Unite and Divide. Interdisciplinary approaches for a sustainable future <http://www.esf.org/hosting-experts/scientific-review-groups/humanities/strategic-activities/esf-junior-summit.html>
20. *Journal of Water Resource and Protection* (JWARP) 5, No. 4A, <http://www.scirp.org/journal/jwarp> or <http://www.esf.org/water>
21. ESF Standing Committee for the Humanities (SCH) foreword. <https://www2.esf.org/asp/ERIH/Foreword/index.asp>
22. Back to books. (2010) *Nature* 463, 588
23. Schneider, P. (2009) An outline of the bibliometric indicator used for performance based funding of research institutions in Norway. *European Political Science* 8, 364–378
24. Sivertsen, G. (2010) A performance indicator based on complete data for the scientific publication output at research institutions. *ISSI Newsletter* 6, 22–28
25. Verleysen, F.T. and Engels, T. (2012) Historical publications at Flemish universities, 2000–2009. *Journal of Belgian History* XLII, 110–143
26. Verleysen, F., Ghesquière, P. and Engels, T. (2014) The objectives, design and selection process of the Flemish Academic Bibliographic Database for the Social Sciences and Humanities (VABB-SHW). In *Bibliometrics: Use and Abuse in the Review of Research Performance* (Blockmans, W., Engwall, L. and Weaire, D., eds), pp. 115–125, Portland Press, London