POLITECNICO DI TORINO
Repository ISTITUZIONALE

How bibliometrics is affecting SSH

Original
How bibliometrics is affecting SSH / Morriello, Rossana. - ELETTRONICO. - (2021), pp. 1-10. (Intervento presentato al convegno International Conference on Information-communication technologies enhanced Social Sciences and Humanities (ICTeSSH) 2021) [10.21428/7a45813f.16c632d9].

Availability:
This version is available at: 11583/2930232 since: 2021-10-22T12:39:13Z

Publisher:
Science 2.0 Alliance

Published
DOI:10.21428/7a45813f.16c632d9

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

12 October 2024

International Conference on Information-communication Technologies enhanced Social Sciences and Humanities - ICTeSSH • Proceedings of the ICTeSSH 2021 conference

How bibliometrics is affecting SSH

Rossana Morriello[1]

[1]**Programming, Development and Quality Area, Polytechnic University of Turin, Turin, Italy**

Published on: Jul 20, 2021

License: Creative Commons Attribution 4.0 International License (CC-BY 4.0)

Abstract. Our society requires an increasing collaboration between SSH and STEM disciplines so to face the growing complexity effectively. However, current methods for research evaluation in Italy separate STM and SSH disciplines and make collaboration difficult. Academic libraries use citation counts, together with usage statistics, to support collection development policies, that is to decide cancellations and renewals of journals and books. A survey in Italian academic libraries I resume in this contribution shows how most of libraries use some kind of quantitative indicator in collection development. This might reinforce the trends and problems of bibliometrics as well as the oligopolistic concentration of the scientific publishing market in the hands of few big players reinforcing the role of the producers of citation databases. Academic libraries allocate most of their funding for big deals acquisitions of journals and for citation databases to support evaluation. Therefore, there is not much budget remaining for other kind of materials, for supporting non-bibliometric disciplines. The bibliometric world is prevailing and increasingly SSH habits are changing to become more similar to STM.

Corresponding author: rossana.morriello@polito.it

Much has been written about current massive use of bibliometrics for research assessment, and risks in the short term and especially in the long run. Bibliometrics are increasingly being used as a proxy for quality of publications in national evaluation exercises, and increasingly inside universities to assess researchers for promotion or even for them to enter universities and beginning their career, to allocate resources coming from the government and inside universities for distribution of resources among departments. In some countries, like Italy, the use of bibliometrics is reaching such levels as to create distortions in the production of science, and tendencies to game the metrics. [1] Many are the risks in the short term, like the increase of self-citations or citation clubs, as well as the proliferation of predatory journals that promise fake impact factors. [2]

However, what is more problematic is the standardization of research, with levelling to a medium level and the discard of highly innovative research. Mainstream topics, or topics already well known by the scientific community, are more likely to get citations, while new topics or niche topics receive fewer citations, particularly in the short term. Novel topics need more time to have impact. Paula Stephan, Reinhilde Veugelers and

Jian Wang, published the results of a study according to which “fifteen years after publication, highly novel papers are almost 60% more likely to be in the top 1% of highly cited papers. Highly novel papers also tend to be published in journals with lower impact factors.”[3] As stated by the authors, this study proves that “the more we bind ourselves to quantitative short-term measures, the less likely we are to reward research with a high potential to shift the frontier — and those who do it.” The Covid-19 pandemic showed us very clearly how a very specialized research conducted by few scientists could become essential at a certain point for society. Basic research as well as applied research, mainstream as well as niche research are equally important and any choice that rewards them differently is a bad choice.

Besides all intrinsic limits of bibliometrics, this kind of indicators refer only to a part of the research process and only to one output, the publication. However, the research lifecycle is a long, articulated and complex process of which the publication is only one step, not necessarily the most important and not necessarily the final one. The publication is the tip of the iceberg, which is visible, but most of the work lies under water. [4] This is the reason why there are new and different kinds of output to disseminate research at different steps of the lifecycle, like preprints, data papers [5], micropublications and nanopublications. [6]

The output of research can be at a level even smaller than the article, that is the level of micropublications or dataset. This makes even more ineffective the use of metrics at journal level, like the impact factor, that are frequently and inappropriately used, on the contrary, to assess articles. Impact factor, quartiles, and similar indicators refer to the “container” not the contents. The Science Citation Index, which led to the impact factor index, was created for other purposes. Eugene Garfield conceived the index to help librarians define the core collection, based on the assumption provided by Bradford’s Law that 20 percent of journals in a library collection are sufficient to address 80% of user’s needs. Therefore, it was a metric conceived mainly for libraries, and in fact many libraries still use this and other kind of bibliometrics in collection development, and for researchers, in order to facilitate knowledge of what is published in a field and to improve scientific communication. As known, in the origin bibliometrics was called “statistical bibliography”, a definition later changed after the challenge by Alan Pritchard, who was a librarian, and proposed to change it into the new word “bibliometrics”, with the meaning of “application of mathematics and statistical methods to books and other media of communication.”[7]

The meaning of the word “bibliometrics” was originally much wider than the unidimensional aspect we refer to today, the count of citations. About ten years later, Dieter Schmidmaier, librarian in a technical university library, explains clearly the complexity of the application of mathematics and statistical methods to publications defined as bibliometrics. [8] In his words, besides citation counts, bibliometrics includes other indicators like published papers and number of authors of published papers, time budgets involved in the research, number of staff as weIl as size and composition of collectives, funds granted or spent, buildings, devices, plants and machines used, visible tendencies and trends important for the future development of research. A complex set of indicators that could not be inferred from publications only and that needed many sources such as published works, mainly in the form of original papers, with data about citations, but also literature compilations and reference books. In addition to publications, indicators are needed about records of the use made of published works (usage statistics), people and the result of their work, mainly in the form of time studies, questioning of individuals and of groups of people, other sources, as records, compilations and lists of contents of published works.

With time, and the commercialization of the Science Citation Index and later other Citation Indexed by the Institute of Scientific Information, the previous multifaceted meaning of bibliometrics was lost, and the value of the indices of citations was reinforced. Later on the adoption of citation indices for evaluation of research and the creation in 2004 of a concurrent index database, Scopus, by a pushing publisher like Elsevier, reinforced the perception of bibliometrics as a monodimensional measure referred to citations only. In Italy, a law stated that for research evaluation the databases used are Web of Science and Scopus (and marginally MathSciNet for mathematics). Academic libraries, therefore, have to buy these two databases because they are needed for all kind of evaluation of research and researchers.

Considering the influence that bibliometric indices have on research trends, I decided to verify if these trends are in some way furthered by libraries choices. I prepared a survey that in March 2019 I sent through Google Forms to directors of library services and heads of digital services in Italian universities. [9] I got the list of universities from the register of the Ministry of the University and Research published on the website of Cineca, the information technology consortium for Italian universities. I excluded private universities and telematic universities, and universities where I could not find information on the existence of a library. I then sent the survey to 67 public universities and 19 legally recognized universities (which are not of the state, but the state recognize them a status of public university and the degree they give is equated

to the state degree). After the first sending, I received 40% of responses, so I sent the survey a second time in April to those who had not answered before. After the second sending, I received a response from 70% of academic library systems. In the accompanying letter, I explained that I intended the meaning of bibliometrics in the original sense, so referring also to another indicator, usage statistics. The questions in the survey covered many aspects of collection development, but I will show here only some of the results useful to the purpose of this paper.

The results of the survey generally confirm the use of bibliometrics, since only 28,6% of library systems declared they did not use metrics at all. A vast majority (91,7%) declared that they use usage statistics for collection development, and about half of the respondents use counts about loans. As for citation indices, 36,1% use impact factor, 19,4% citation counts, 5,6% quartiles. When asked what database they used, 71% declared Scopus, 67,7% Web of Science, 58,1% Journal Citation Reports, 29% MathSciNet, 12,9% Google Scholar, 9,7% Scimago Journal & Country Rank.

The main reason for using bibliometrics in libraries is deciding about cancellation of journals, books, databases. However, for print journals 82,1% refer to usage statistics and only 7,1% to citation indices. While for e-journals 100% refer to usage statistics.

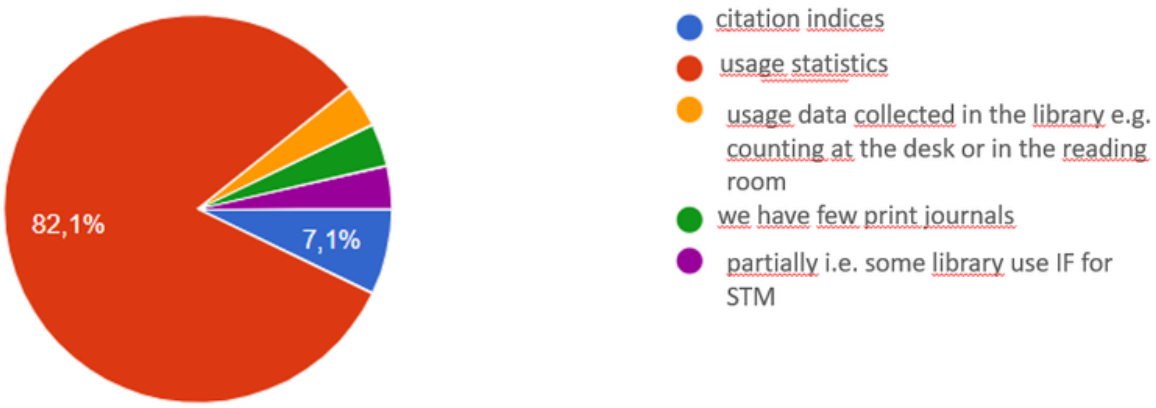

Fig. 1. Cancellation of print journals

Also for books, the cancellation of the electronic format is based 100% on usage statistics, while the print format is mainly based on download statistics, and partially on faculty decisions. As one may expect, citation indices are not used for books because citation databases do not include books in a substantial quantity. Books are a

kind of research output more typical of SSH disciplines that citation databases include only in small quantities.

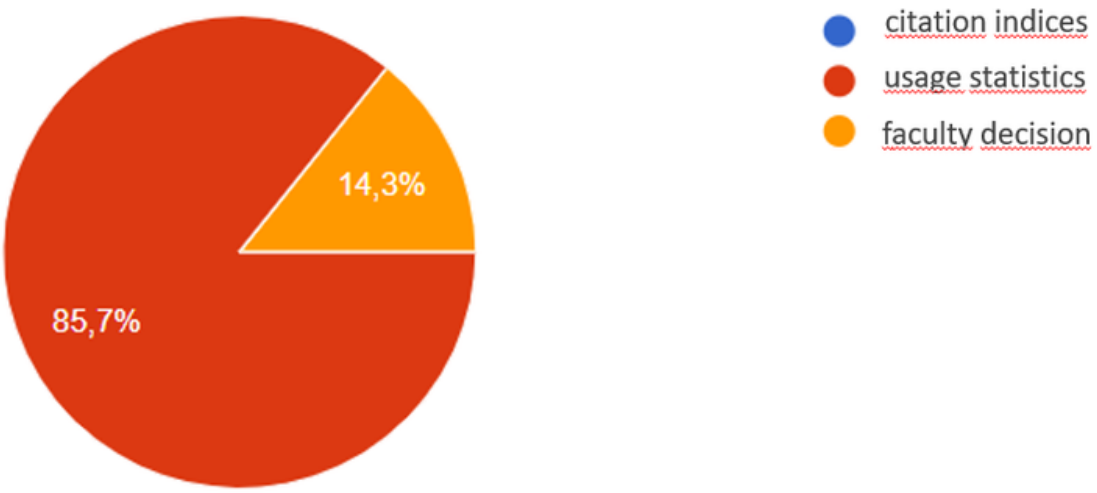

Fig. 2. Cancellation of print books

Other uses of bibliometrics in Italian academic libraries for journals, books, and databases can be put together since they have a similar proportion, about 70-80% for usage statistics and up to 30% for citation indices. They are defining the core collection, deciding about new titles to acquire, analyzing the use of the collections in big deals, and, with slightly different proportions, compiling guides and reference materials.

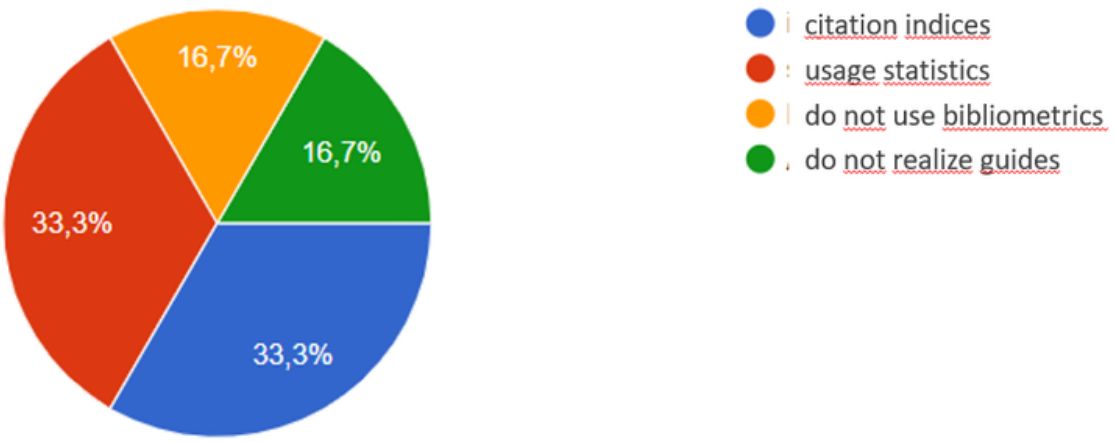

Fig.3 Compilation of guides, bibliographies and reference materials

Librarians interviewed in the survey were asked why they think the use of bibliometrics is helpful. Despite some of them answered that they believe it is an objective way to support decisions to cut subscription or to ask for funding of new resources, many stated that they are useful only if supported by other indicators, like comparison with collections of other libraries, particularly in cooperation agreements, and qualitative analysis. Librarians were also aware of bibliometrics being partially useful because they can be used only for STM resources and not for SSH resources. This is, in fact, the main evidence showing how also for collection development in libraries there are two different ways of assessing collections.

Considering the result of this survey, we can catch some evidence. Firstly, libraries might consolidate trends in research evaluation by strengthening the role and use of citation indices and should be aware of this and try to avoid it. The risk is that only journals with high impact factors will survive and this is very much against plurality and bibliodiversity, which libraries must guarantee. In the scientific publishing market there is a concentration and a very well know oligopolistic asset that massive use of citation indices makes stronger. Oligopolistic players are the same that also offer other kinds of products for the whole lifecycle of academic research, like tools for organizing bibliographies, for organizing data, and citation databases for evaluation of research. Libraries devolve most of budgets to these products and little remains for other disciplines, which are outside the citation game, like social science and humanities. SSH research outputs are very much linked to bibliodiversity, because they have greater diversity of products formats, there is not a prevalence of the journal article

but rather of the monographs, and include book chapters, scripts, working papers, critical editions, and oral literature, for example. In addition, they have more diversity in publishing places and habits since SSH researchers in non-English speaking countries, like Italy, publish less in English and more in national languages, their research is often local, and consequently they publish with countless small and medium publishers. For researchers in SSH publishing in English in international journals is less important, because in many publications the intended audience is local, maybe not even national but linked to a very specific territory (let's think of urban architecture, for example). In SSH, there is more variety in the habits of making research, it is more individual, effects and impact are not viewable immediately, not even maybe after five years, sometimes it takes a whole life.

Nevertheless, the weight of bibliometrics measurement and the illusion that it represents an objective way to evaluate research impact (but confusing impact with popularity), are producing a separation between what looks like to be measured and evaluated objectively, STM publications, and what is not, SSH publications. Unfortunately, this trend seems to bring a change of SSH publication habits and evaluation procedures rather than the opposite. In Italy, the national evaluation agency, ANVUR, created for each academic discipline some lists of journals considered "scientific" and "class A". There are lists of scientific journals and list of class A scientific journals in which researchers and professors must publish for career advancement. If they do not publish in those journals, their publications are not evaluated. Therefore, it is not important anymore to choose the right national or international journal, the most suitable for publishing your research, but you have to choose one of those in the lists. This means again assessing the container instead of the content, very much like the impact factor. If academic libraries would ever start using these lists for collection development, the possible negative consequences would be the same as for bibliometrics. Another indicator used for career promotion in SSH is the number of publications and this means that many researchers prefer to split what could be a monograph in many articles or book chapters in miscellaneous publications. The length of monographs is decreasing, and mini-monographs are a new kind of publication spreading. [10] The quantitative kind of measurement inducted by bibliometrics for STM, in the current meaning of bibliometrics, quite different from the original, is affecting also SSH and changing the nature of research in humanities and social sciences. Not considering the implication as regards open access that traditionally show a great divide of perception between STM and SHH.[11] One of the fears and restraints of going open access for researchers in the humanities and social

science is poor quality. This could even worsen because of the idea of being considered different and less important since their outputs are less valuable in an objective way. It is clearly an illusion that quantitative bibliometrics methods can be an objective measure, of course, but this perception is rooted in many cases, and publishers and producers of databases are fortifying it for commercial purposes and for affirming their position in the market. Librarians, together with researchers, should work to avoid favouring these trends and all consequent distortions in the making of research and the advancement of science. As for libraries, much can be done in rethink collection development strategies, relying less on citation indices, devolving much reflections and actions to supporting open science, and trying to move budgets to more sustainable models of dissemination of research results, like library publishing and open platforms created on cooperative agreements, even with publishers. Research must be brought back to freedom and diversity and only by a collective action able to offer an alternative to the present system, with all actors playing their role towards this aim, also research evaluation could change.

References

1. Baccini, A., De Nicolao, G. & Petrovich, E. (2019). Citation gaming induced by bibliometric evaluation: A country-level comparative analysis. *PLoS ONE* Sept 11 https://doi.org/10.1371/journal.pone.0221212
2. Bagues, M., Sylos-Labini, M., Zinovyeva, N., (2019). A walk on the wild side: 'Predatory' journals and information asymmetries in scientific evaluations," *Research Policy*, **48(2),** 462-477.
3. Stephan, P., Veugelers, R., Wang, J. Reviewers are blinkered by bibliometrics (2017). *Nature News*, Apr 26, 2017, https://www.nature.com/news/reviewers-are-blinkered-by-bibliometrics-1.21877#/novelty
4. Roskams-Edris, D., Poline, JB, Stikov, N. (2019). The Canadian Open Neuroscience Platform: Catching Up to Plan S and Going Further, *PLOS Blog* Aug 15, https://theplosblog.plos.org/2019/08/the-canadian-open-neuroscience-platform-catching-up-to-plan-s-and-going-further/
5. Chavan, V., Penev, L. (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science, *BMC Bioinformatics* **12, S2**, https://doi.org/10.1186/1471-2105-12-S15-S2
6. Raciti D., Yook, K., Harris, TW., Schedl, T., Sternberg, PW. (2018). Micropublication: incentivizing community curation and placing unpublished data into the public domain, *Database. The Journal of Biological Databases and Curation* Mar 2, doi: 10.1093/database/bay013

7. Pritchard, A., Statistical bibliography or bibliometrics (1969). *Journal of Documentation* **25(4)**, 348–349
8. Dieter Schmidmaier, Application of bibliometrics in technical university libraries (1977). *Proceedings of the IATUL Conferences*, Paper 11.
9. The results of the survey are partially published in Morriello, R. Le raccolte bibliotecarie digitali nella società dei dati. [2020] Milano, Editrice Bibliografica.
10. Palgrave Macmillan launches new format with Palgrave Policy Essentials (2018). 15 Oct, https://www.palgrave.com/gp/media-centre/press/palgrave-macmillan-launches-new-format/16227372 ; Stanford Briefs (2021), Stanford University Press https://www.sup.org/books/imprints/?imprint=STANFORD%20BRIEFS ; Cambridge Elements (2019). *Research Information*, Jan 17, https://www.researchinformation.info/product/cambridge-elements
11. Nature Publishing Group (NPG), Palgrave Macmillan, Author Insights Survey (2015). Version 7 Sept 23, retrieved from Figshare, https://figshare.com/articles/dataset/Author_Insights_2015_survey/1425362