

Supplementary Information for “Metrics for Openness”

David M. Nichols

*Department of Computer Science, University of
Waikato, Hamilton 3240, New Zealand.*

E-mail: dmn@cs.waikato.ac.nz

Michael B. Twidale

*Graduate School of Library and Information
Science, University of Illinois at Urbana-
Champaign, Champaign, IL 61820-6211, USA.*

E-mail: twidale@illinois.edu

Calculating a Personal Openness Index

For simplicity, we had chosen to consider just journals and conference papers, but that raised a further question of what exactly belongs in those two categories. We uncovered a number of edge cases:

- Annual Review of Information Science and Technology (ARIST). We decided that this should be included; in other words to treat serials as journals.
- *interactions*. This is a magazine published by the ACM. Articles go through several rounds of editing but not rigorous full peer review. The magazine is aimed at a broader audience of usability professionals. However interactions articles are often cited in the scholarly literature, frequently as good introductory overviews or examples of the application of a theory or method. We decided that this should be included.
- Encyclopedia entries (such as articles in the Encyclopedia of Library of Information Science (Bates & Maack, 2009)). We decided to treat these as book chapters and so excluded them.
- ACM SIGCHI Bulletin. This publication was not peer reviewed (but was editorially reviewed), but was in a journal-like format and was frequently cited. Later it became a supplement to interactions. This felt like a particularly awkward edge case, with the lack of peer review arguing for exclusion, but a

relatively large number of citations arguing for inclusion. We decided for exclusion, but feel that this particular case was especially difficult.

- Position papers in conference workshops. Peer reviewed, but by the workshop organizers. Included because they have the format of papers, and are often made available and cited - frequently as the first public airing of a piece of work.
- Descriptions of and calls for participation at conference workshops. These appear as documents in the ACM Digital Library, with assigned DOIs and can be cited in the same manner as normal papers. We chose to exclude these, as they do not seem to have the idea-conveying power of papers. Panels and other conference events were likewise excluded.

These and other edge cases require decisions, and we do not claim that ours are ideal. We mainly want to show the degree of fuzziness in the process of deciding what to include, and that it is highly desirable to make these decisions public. For example, from the perspective of maximising an openness measure, there is a temptation to include publications that are easier to make open (or are typically open) and exclude those with more restrictive copyright agreements. In our three-way classification then, only reading access matters and other licence terms (such as non-commercial restrictions) are not relevant.

Automated calculation of a POI (further emulating the rise of the h-index) will need a substantial degree of clarity—no algorithm is going to be able to make these kind of judgement calls. It is likely that presence in one or more bibliometric collections (Web of Knowledge, Scopus, ACM Digital Library or even Google Scholar) could be used as a convenient selection criterion.

One simplification of the basic POI calculation was to be somewhat version-agnostic (submitted, accepted or published versions count; pre-submitted versions and rejected papers do not). Looking at our own publications by downloading them if they were available, it is sometimes difficult to tell exactly which version they were. In theory, the paid-for published version could be downloaded and compared to see if it matches word for word, or even if the layout matches. To do this manually would be excessively laborious. However, to do it algorithmically over a large body of work for many authors may violate the publisher's subscription

agreement.

The whole OI assessment activity got more difficult as we worked back in time. Older papers may have been stored in different formats (such as PostScript). Papers may have been uploaded to a webserver that no longer exists. Given the authors' research interests, many of our papers had been harvested and indexed by CiteSeer. This certainly increased our POI scores in cases where a paper had been uploaded to a website, harvested and stored by CiteSeer and then that original website ceased to exist.

Deciding if older papers meet the inclusion criteria may be harder. If they are not freely available, making them so may also prove more difficult. The legal status may be unknown. A personal electronic version may be obsolete or non-existent. The author may even lack a paper copy to digitize. Consequently retrospective open archiving will be a problem. As a result we can expect older researchers to have lower POI scores and also find it hard to raise their scores to the levels easily attainable by younger researchers. This will need to be acknowledged in OI considerations. It is also a challenge for the introduction of the concept. Older researchers are often powerful gatekeepers and it can be problematic to propose a metric that shows people with power in a less than flattering light.

The process of hand computation also revealed a number of errors, even in carefully updated CVs. Citation information may turn out to be incorrect, including the exact title. The author may have evolved his conventions for citing his work over the years, perhaps forgetting that some early conferences did not have a published proceedings, or a published set of abstracts only. Our experience is a reminder that any such process is dependent on the quality of the input data.

We believe this learning experience is useful to share as a way of informing the metric design process. We hope that bibliometric purists will not be too horrified by the messiness and indeed sloppiness revealed. At least for the early iterations of this process, we suspect that such messiness is inevitable.

References

Bates, M.J. & Maack, M.N. (Ed.) (2009). *Encyclopedia of Library and Information Sciences* (3rd Ed.). CRC Press, Boca Raton, FL, USA.