ICSTI Public Conference 2007 – Session 2

Tools and Metrics

Keynote speech

Alternative metrics from a publisher's and a librarian's perspective

Donald King
School of Information and Library Sciences, University of North Carolina at Chapel Hill

This presentation will explore what librarians should know about journal publishing, what publishers should know about libraries, and what publishers and libraries should know about readers. Participation in many librarian and publisher meetings has shown that neither librarians nor publishers understood what the other was doing and their reasons for doing it. So, this presentation will try to go beyond the librarian's "Oh my god, they've raised their prices again" and attempt to explain why publishers have raised prices, why they went to electronic publishing and its consequences, and it will broach the issue of author-side payment and finally that of the readers.

When it comes to publishers, librarians should know about the consequences of major publishing policy decisions (print vs. electronic vs. both print and electronic; subscription vs. author-side payment). They should also know about typical publishing cost, journal pricing and journal financing on which some publishers base their decisions. Publishing cost involves fixed costs or first copy cost, that do not vary with circulation; variable costs (printing, subscription, maintenance, mailing, access, etc.), and they exist for both print and electronic; and indirect costs or overheads (administration, financing, marketing, R&D, etc.). Average cost is the total of the costs divided by output i.e. circulation. The typical publishing costs of a 150 article journal, for print only, electronic only and both – and electronic is not cheaper, as was first thought – are presented as well as the typical average cost per subscription at 500, 5,000 and 50,000 subscribers.[1] The average fixed publishing cost per article turns out to be close to 3,000 USD, slightly higher for electronic than for print. Some of the literature indicates that author-side payment should do away with fixed and variable costs and that with 50,000 subscribers, around 2.5 million dollars would be saved. As far as the pricing of journals is concerned, there are reasons for the steadily increasing prices and the price variations between journals and between publishers. According to a price tracking study of a set of journals conducted for the National Science Foundation and covering the period 1960 to the late seventies, prices have increased more than inflation every year. As a reminder, in 1960, the average price of a journal was less than 10 USD and the price of personal and library subscriptions were almost similar. Resource cost (labor, printing, etc.) has increased more than inflation and the same is probably still true today. Journal size has increased and nearly tripled in a few years (1977: 820

---

[1]Detailed figures are available in D.W. King, The cost of journal publishing: a literature review and commentary, *Learned Publishing* **20**(2), 2007.

pages/title; 2001: 2,216 pages/titles). More special features have been included (graphics, images, chemical compounds, equations and value added features such as journal linkage and retrospective access to journals, etc.).

Revenues have been decreasing due to changes in author payment practices and this caused a spiraling effect. Something that people tend to forget is that some years ago, over 50% of science journals used to require some form of author payment (about 760 USD in today's dollar). Since commercial publishers were not allowed such charges, authors started drifting towards non-charging publishers, and as a consequence publishers such as ACS started to lose revenues and had to make up for it. This lead to a spiraling effect: when prices are increased, subscribers and therefore revenues are lost, so prices are increased and so on. The reasons for the price variations among journals and publishers are difficult to sort out. Prices must reflect the average cost per subscription. Commercial publisher prices are clearly higher than those of non-profit publishers. Some journals or publishers serve small audiences and a commercial publisher median circulation is much lower than that of a society (up to 5,000 subscribers per title). There is a strong argument that the marketing power of commercial publishers might explain their high prices, and some authors and a recent EU study make a strong point for that. Alternative strategies exist such as discriminatory prices where personal subscriptions cost less that library subscriptions, but personal subscriptions are based almost exclusively on variable costs. Also large publishers will accept revenue-losing journals and make up for it with high-revenue journals. This is also true of a journal with some highly read articles and others less so. One of the most important things that happened over the last ten years is licensing and this will be covered later. Commercial publishers have to make a profit and often non-profit publishers make a surplus i.e. a profit and oftentimes societies will subsidize the journals they publish. We have to recognize the price and demand relationships and the fact that library subscriptions are much less sensitive to price changes than personal subscriptions. The flow of funds among the different participants in the system with author-side payment also has to be considered and, preferably, not in the simplistic way they are considered today. Financing journal publishing is not cheap. The data presented here are based on the work developed by Jack Meadows, Gillian Page and Bob Campbell.[2] They show that it takes 100,000 USD to start a journal and four to ten years to break even and it took 40 years for *Nature* to break even. Some costs such as editing and reviewing are incurred before the revenue comes in, and these costs have to be financed and author-side payment may help. R&D costs have skyrocketed but they mean better services. Non-profit publishers often finance those costs by borrowing or getting subsidies from their organizations.

What should publishers know about libraries? Libraries have major decisions to make such as purchasing journals or borrowing copies of articles from another source, ordering print or electronic or both, purchasing licenses instead of single subscriptions. Based on a recent study of the speaker, library costs go beyond acquisition costs with processing, user-related and use-related costs such as training to access electronic resources. Libraries have fixed costs that don't vary (acquisition/licensing and processing) and variable costs such as photocopy and reshelving for print, and equipment and printout for electronic. When it comes to purchasing versus borrowing, borrowing is based for both print and electronic on the price of the journal and the breakeven number of uses, proportional to the price. The author-side payment also has an effect especially for libraries serving large academic institutions with contributing authors. When comparing print and electronic journals, a certain number of factors come into play such

---

[2]G. Page, R. Campbell and J. Meadows, *Journal Publishing*, Cambridge University Press, 1997.

as resources used, life-cycle cost per title, cost per reading, and finally contingent valuation which is not discussed here in detail. Cost includes all the resources used such as acquisition, staff, facilities, equipment and systems, photocopies, binding and so on. Cost per title can be misleading when looking at electronic collections compared to print collections because there can be overlaps with licensing. An interesting fact is that some of the most used titles in a licensed electronic collection might be titles previously cancelled because of their cost. The cost per reading over the life cycle of a title is also an important factor to be taken into account. Global return-on-investment studies for library journal collections can be made by means of contingent valuation, an economic method used to assess the benefits of non-priced goods and services (e.g. libraries or specific library services) by examining the implication of not having the product or service. Results show a positive ratio of return-on-investment of about 3 to 1 (return in net benefit divided by total organization investment).

When it comes to what publishers and librarians should know about readers, readers make decisions as to where to get information, how to find articles and where to obtain those articles once they have identified them. Past experience with journals or information sources, availability and accessibility and most importantly the cost in their time and money, are factors in their decisions. Donald King and Carol Tenopir did a study in which they asked what were the sources last used to find a substantive piece of information. The results showed that main source was journal articles, followed by informal discussion with colleagues, and then conferences and workshops. So far, sources as eprint servers and author websites are not very popular, but they certainly will be in the future. There are many ways to identify articles (browsing, online databases, citations, colleagues, etc.) but browsing and searching online databases come first. Personal subscriptions (mostly print) have been going down while library collections have been going up as sources of articles while preprint are still little used. Scientists, in universities and elsewhere, are reading more and relying increasingly on libraries. The contribution of electronic collections is important since electronic library collections have broadened journal availability and save reader time. Reading of older materials has increased mainly through the availability of systems such as JSTOR. Libraries also contribute to the usefulness and value of articles. For research purposes, library collections are more reliable, help achieve the goals of parent organizations and readers tend to be more productive than non readers.

Acknowledgements are given to José-Marie Griffiths, Carol Tenopir, Sarah Aerni, Carol Montgomery and Roger Schonfeld for their work in the field and the large amount of survey data provided.

Discussion

The main issues that were raised focused on the shifting roles of publishers and librarians, and on the future of libraries.

Evaluation based on publications: Experiences in Physics

Denis Jérome
CNRS, Académie des Sciences

Evaluation of publications is a major topic today. Lately, it has become quite important in France and a group is currently working on this issue at the Académie des Sciences. Like most countries needing to fund basic research, France is caught in a loop: Public funding is needed for research. Publications are

the outcome of research and publications are needed to disseminate knowledge and advance research. Publications are also needed to evaluate researchers, research laboratories and institutions, and the country in which the research is performed. Such evaluation is then used for public funding. So there is some kind of a loop system at work. The situation between evaluation and publications is rather intricate not to say tricky and one might ask the question of whether or not publications can be used to evaluate research results. The following presentation will be based essentially on the field of Physics. The situation of evaluation is quite paradoxical. While the scientific contribution of Europe is important in many domains of hard sciences and especially Physics, it is a minor contributor to scientific publishing. The number of publications such as Letters published in Europe is much smaller than in the United States and the largest share of publications submitted to American publishers such as the American Physical Society (APS) or the Optical Society of America (OSA) comes from European contributors. For example, publications in *Europhysics Letters* have remained steady over the years while they have steadily increased in *Physical Review Letters*, and fifty-percent of APS publications are of European origins. Production started changing around 1996 and European Union (EU) production has been overtaking US production, the United Kingdom, Germany and France being leading within the EU. There is a strong need for plurality and too much concentration, as is the case for the APS, could be dangerous. Some variety in editorial policies is needed, local and historical contexts have to be taken into account and publishers should be in close contact with the various communities. Plurality is also needed to avoid editorial bias and leave more freedom of choice to researchers. Direct contact between authors and publishers should be established to avoid a time-consuming and bureaucratic evaluation process. Computer-assisted evaluation might help, but computers should not control.

Basic research is to a large extent supported by the state which decides how monies are allotted for research positions, equipment and operating costs. So, an evaluation of individuals, research laboratories, universities and research organizations is needed. Also, researchers expect the recognition that comes from a good ranking of their publications in order to get good positions and promotion. So far many evaluation committees still rely on peer review when it comes to analyzing scientific output but, nowadays, they also rely on bibliometric tools such as the famous or infamous Impact Factor. The Impact Factor is at the core of the evaluation issue. Today, the Impact Factor is that of journals and currently it is calculated on a two-year basis. It is calculated every year for Thomson's ISI *Journal Citation Reports (JCR)*. One might question the arbitrariness of the two-year period and the fact that only 6,000 Science and Technology journal titles are taken into account, even though Physics is rather well represented. The Impact Factor might be a good indicator for publishers but it can be misused to evaluate individuals and the distribution of citations is quite skewed by the notoriety of a journal. For example, articles in *Science* and *Nature* accounted for a high number of citations per articles in 2003–2004 while the number of citations was much lower for *Physical Review Letters* and *Europhysics Letters*. The editorial policies of prestigious journals also come into play. For example, scientific domains such Physics are much less represented in *Nature* and *Science* than fields such as Biomedicine.

All in all, the Impact Factor should be banned since it can be misused when only the notoriety and popularity of a journal is taken into account instead of the notoriety of the articles themselves. Other indicators are needed to deal with the importance, quality and impact of the scientific research, besides the notoriety of a journal. Physicists publish in a small number of journals nearly all listed in the *Web of Science (WoS)*. A test was conducted in four major labs at the University Paris XI which, by the way, ranked first for France and twenty-seventh in the world in the latest Shanghai ranking of world

universities. Physics is rather well represented because for the past thirty years, it has published mostly in English which might not be the case for other disciplines such as the Humanities. Today, there are web-based databases with inverted bibliographic data that can be used for citation analysis. Some of them such as Google Scholar and NASA ADS are free. NASA ADS cover mainly Astrophysics but Physics is well represented. Other databases such as the Thomson ISI databases (*WoS*, *JCR*) are especially useful for journals, while Google Scholar is better for books and proceedings. Using the "h-index" (Hirsch)[3] and the "g-index" (Egghe)[4] might prove a better way to rank articles and, according to an experiment conducted with the US Academy of Sciences, the g-index seems to be more reliable than the h-index and the Physics section of the Académie des Sciences compared rather well with the American Academy of Science. The advantages of the h-index are that it can smooth out big hits and review articles and shows good correlation with total production. It can be extended to journals and scientific fields and it takes time dependence into account. With the g-index, an even small number of papers, in case they are excellent, are taken into account and it can reveal precursor papers ("sleeping beauties"). It seems in any case a useful complement to the h-index.

Some technical problems must be mentioned. First, there is the problem of common name homonymy. For example, the name Goldman will pull out over 2,000 citations and it takes a physicist to know which Goldman wrote in the field of Physics. Hyphenated names can also be a problem. Then, there is the problem of self-citation, but only when you are close to the h-index. In the case of team work you have to normalize the h-index. Negative citations to incorrect works, cronyism and networks of friends can also be a problem but the most important is the quality of the citation, that is to say who cites an article. All this means that this system of evaluation has to be handled by scientists.

In conclusion, as far as evaluation in the field of Physics is concerned, databases are rather exhaustive and the same can be said of Chemistry. Bibliometrics can be used besides peer review but it should be used with caution and must be checked by scientists. The results must be validated by the persons being evaluated and bibliometrics should not be used to evaluate young scientists as they may perform very poorly with "h" and "g" but might have done excellent PhD work. There is also the need for international comparison. The Impact Factor should be banned when assessing individual performance. Several indicators related to citations must be used such as the h-, g- and g/h-indices, among others. Finally, to go back to the beginning of the presentation, there is a strong link between evaluation and the poor situation of publications in Europe, especially in Physics. Everything is due to evaluation and Europe has to find a way through this situation. Open Access should be very useful since it will make all publications equally visible on the Web.

Bibliometric approaches to evaluation, with an overview of French scientific research

Mary Van Allen
Thomson Scientific

Science has been growing quickly and it is quite beyond our capability to keep up with it. There has been an estimation of 20,000 papers published a day and there might even be controversies whether

---

[3] http://en.wikipedia.org/wiki/Hirsch_number.
[4] http://en.wikipedia.org/wiki/G-index.

that number is underestimated or not. The sheer volume of scholarly information makes it difficult for scientists, information professionals and scientific and technical policy makers to efficiently keep abreast of what is happening. Furthermore, the nature of today's scientific research is much more global and multidisciplinary than in the past, and that is both an exciting and daunting aspect of science today. New and emerging fields that are the result of some of the multidisciplinary efforts also make it difficult to determine relevancy. And finally, there is the issue of selectivity. How can researchers or librarians determine what to read or subscribe to, what area to focus on and which projects and researchers should receive more support than others.

Today, we are awash in metrics such as Cybermetrics, Scientometrics, Informetrics, and Webometrics. The main issue is knowing for what purposes those metrics are used and how to provide definitions around this concept of metrics. A good definition of bibliometrics is that of Dr. Henry Small, Chief Scientist at Thomson Scientific which says that: "bibliometrics is the statistical analysis of bibliometric data, mainly in scientific and technical literature. It measures the amount of scientific activity in a subject category, journal, country, topic or other area of interest". Bibliometrics most importantly gives a landscape of scientific communication and productivity, but one must not forget its statistical aspect. However, this statistical aspect is important since, as Lord Kelvin said: "If you can measure that of which you speak and can express it by a number, you know something of your subject; but if you cannot measure it, your knowledge is meager and unsatisfactory". Bibliometric statistics help provide base definitions and parameters for analysis, and provide quantitative data that administrative and funding agencies can understand to make rational and scientific decisions. Bibliometric methods provide various approaches to evaluation based on the theory of citation that says that citations are a form of peer recognition. These approaches can be basic and relative measures, look at units of analysis, i.e. the elements within the bibliographic records, and aim at providing measures of interaction. The basic measures are based on papers, citations and average citations per paper (ex.: Impact Factor). Relative measures relate to baselines for subject categories. Particular measures are expected citation rates for articles, percentiles, which are basically the frequency of distribution of citations across a subject category, and measures of interaction (showing the multidisciplinary nature of the citing paper and collaboration opportunities found through the citing paper). It is important to have those basic and relative measures in place and well defined, so that these metrics can be repeatable and reproducible not just by scientists but by the decision makers who are going to fund those scientists.

Those who benefit from bibliometrics are policy makers, research directors, university and government administrators, information specialists and the researchers themselves. However, librarians and information specialists, who have an important role as information conduits, have not sufficiently expressed themselves about bibliometrics. Many nations and agencies are using Thomson Scientific citation information for some type of assessment within their given groups. This proves that there is some value in citation analysis even though this is still subject to controversy. There is not one metric that can fit everyone's needs since the needs of a given population and the metrics available to them have to be taken into account. The *Web of Science*[5] is only one of the products offered by Thomson Scientific and everybody knows the *Journal Citation Reports*[6] that is heavily used by publishers. There is also the *Essential Science Indicators*[7] which highlight the top 10% (the cream of the crop) of researchers in institutions

---

[5]http://scientific.thomsonreuters.com/products/wos.

[6]http://scientific.thomsonreuters.com/products/jcr

[7]http://scientific.thomsonreuters.com/products/esi.

and which can be used to create benchmarks of where an institution stands compared to the top echelon of its respective fields.

The last 26 years of scientific publication output for the country of France will be used to assess French research. The average growth over the world for these 26 years was 2.9% each year while in France it was slightly above at 3.32%. France continues to have a strong participation in world research with a slight dip in the five more recent years or so, and it would be interesting to correlate that fact with French national or regional policies. For the purpose of this presentation of French research, three fields were selected: Geoscience, Neuroscience and Behavior, and Plant and Animal Sciences. There has been a decline over time in Plant and Animal Sciences papers. This does not mean that French research is dying out in that field but rather that researchers might be broadening their scope and writing in journals in neighboring fields. On the other hand, Geoscience has shone an impressive growth. In the last 10 years, previously under-represented areas such as Engineering and Economics have started to grow. Both initial and end numbers must be shown to get an accurate picture. Numbers have to be considered in context and according to the practices in the field. The French impact by selected field, i.e. the average citation and cite rates for the selected fields, shows that Neuroscience and Behavior is rising rather strongly while Geoscience and Plant and Animal Sciences are somewhat flattening out. As far as collaboration is concerned, a special run of the indicators program showed that, for example, the number of papers coauthored by French and British authors in Clinical Medicine has been rising and that the citation rates were higher for papers coauthored by French and Spanish authors.

For a detailed evaluation of French research based on individual papers, the following metrics were used: (1) Cites, or the total citation count for a selected paper; (2) Cites2, or a second generation cite count based on total citations received by the *citing* articles; (3) Expected Citation Rate, or the average citation rate for all papers of a given type, in a given journal for a given year. This might look similar to the Impact Factor, but the difference is that it is used to valuate the citation counts in a journal for a specific publication year; (4) Ratio, or the ratio of expected cites to actual cites that were accrued for a specific paper; (5) Field, or the subject area for the journal in which the paper appeared, which is useful to pinpoint an article in a specific field such as Physics published in multidisciplinary journals such as *Nature*, *Science* or *PNAS*; and (6) Percentile, or the % position of a given paper based on citations in the same field, i.e. a frequency distribution of where a given paper places with the other citations in that same field. As an example, in 2002, all French authored articles from the *Journal of Polymer Science* were used to illustrate these metric methods and they showed that a specific paper performed 7 times better than its peers in that journal for that publication year and that it placed in the top 0.9% of all the papers published in the field of Polymer Science for the year 2002.

The following metrics were applied to a group or set of papers in the field of Geoscience in France for publication years 2002–2006: (1) Total number of papers and total number of cites combined for the specific set; (2) Mean times cited i.e. total cites divided by total papers (average impact); (3) Mean times cites i.e. midpoint for citations; (4) H-Index or the number of papers (N) in a given dataset having N or more citations; (5) C-index or the sum of all actual citations divided by the sum of all expected citations; (6) Average Percentile or the average of the field percentile measures which are based on the field and year of publication; (7) Disciplinarity or the level of multidisciplinarity in a set of papers, ranging from 0 to 1 where the lower the number, the greater the multidisciplinarity (the Herfindal index). One point to keep in mind is that 'like must be compared with like'. For example, the same periods must be used when comparing cites accrued over a ten-year period between fields, individuals or institutions. This might be evident but is not always the case. Also, one must remember that different article types will be cited differently (reviews vs. letters or editorial material). The examples given show how different fields

have different dynamics leading to different performance cycles and averages. The issue of self-cites and negative cites can be handled through the peer review process. Self-cites often reflect the cumulative nature of an individual's research and are acceptable if not excessive. As far as the skewed distribution in bibliometrics, the 20/80 rule seems to hold where 20% of papers get 80% of cites. About 50% of papers are never cited.

To conclude, the elements driving the future are the growth in metrics (e.g. based on downloads), the diversity in measurement tools and definitions (which does not preclude the need to have some standardization), the changing demands based on the funding community (returns on investments) and the regional and local interests (rise of institutional repositories and regional databases). Today, nations are taking a hard look at how to provide substantive information about what is happening in their scientific communities. New metrics are emerging in a changing environment and the need for independent information is essential. There is not one metric that can fit everybody, but there is a need for exchanges to try and build some consensus.

Discussion

The issues that were raised mainly concerned the coverage of disciplines and the percentage of French publications in the *Web of Science*.

COUNTER and Usage Factor: Providing new insights into the value of journals

Peter Shepherd
Project COUNTER

One of the issues that has come up in terms of scientific journals is that of value and where it lies in the case of online journals. When talking about citation-based metrics, it is assumed that the value lies in the article. Today, there are many features in online journals that go beyond that of mere collections of articles. Online journals are not used by researchers only. They are used by "practitioners" that do not publish and do not cite. Increasingly, many academic institutions are finding that the rapidest rise in usage of online journals is among undergraduate students. Of course undergraduate student usage is not the same as that of researchers, but librarians feel that it is part of the value they offer their institutions. Traditional data-based methods to measure value do not take usage into account but usage statistics have their role. COUNTER usage statistics[8] could be used as a basis to calculate what is called the Usage Factor.

For years now, publishers and vendors have been providing their customers with usage statistics. The problem is that these statistics are not comparable because they do not use the same definitions, as is even the case with simple notions like *fulltext*. However, it had been clear for some time that libraries and library consortia need usage statistics to help assess the value of the different online products and services to which they subscribe in order to support collection development and justify their materials expenditures.

Things were not easy in the beginning. Publishers were not convinced of the desirability of easily compared statistics and were somewhat reluctant to see their statistics compared to that of others. For publishers to subscribe to the project and provide the data needed, COUNTER had to be practical,

---

[8]http://www.projectcounter.org.

reliable and in context. COUNTER started by producing a series of Codes of Practice, first for journals and databases then for books and reference works, and publishers had to comply with these Codes of Practice in order to be COUNTER compliant. These Codes of Practice provided definitions of the terms used, specifications for the Usage Reports, guidelines for data processing, compliance, maintenance and development of the Code of Practice, and COUNTER governance. COUNTER also instituted an independent auditing process to check the publishers' compliance with the guidelines. The reason for what COUNTER has achieved so far is that it was not purely an initiative from publishers or librarians or subscription agents. All three stakeholders were involved and despite tough discussions at the beginning, it was possible to move forward in a consensual sort of way.

So what is the situation today? The second release of the Journals and Databases Code of Practice came out in January 2006 and today, it is a widely adopted standard by publishers and librarians alike.[9] Over 60 publishers and subscription agencies, and not the least (Elsevier, Springer, Ebsco, . . .) are now COUNTER compliant and usage data cover over 10,000 fulltext journals. Librarians are using COUNTER usage figures for collection development and publishers are using them to demonstrate the value of their products and services. Publishers are required to submit 5 reports, the main being Report 1 that shows fulltext article requests by month and journals, with separate totals for html and PDF. The first release of the Books and Reference Works Code of Practice came out in March 2006. Six reports are required. Only 6 vendors are compliant to date but people are starting to use the data. However, some refinements are needed since relevant usage metrics are clearer for journals than for books. Many books are still available only as PDF files while for other books, such as encyclopedias, it is possible to download individual entries. So, to reflect the greater diversity in book publishing, Report 1 concerns the number of successful requests by month and title, and Report 2 is for the number of successful section requests by month and title. COUNTER also implemented an audit process, where audits can be conducted online and the first audits are now taking place. The audits are conducted by ABCE, a UK organization with a good track record. Once the audit process is fully implemented, it will enhance the credibility of the statistics.

When it comes to applications of COUNTER, two types of metrics (local and global) can be derived from the journal reports. Local metrics, which are developed by individual librarians for collection development, and global metrics, which can be used for funding decisions. Today, the COUNTER data is being used in the UK by the JISC,[10] a government funded organization that helps higher education institutions by negotiating access to information and provides network services and support. JISC also supports projects to develop innovative solutions. One of their projects is to provide a national overview of online journal usage using COUNTER Journal Report 1. COUNTER data for 2003–2004 was analyzed in terms of usage and price band (low, medium, high). Analysis in terms of subject category could not go down to a granular level because publishers among themselves, and publishers and librarians could not agree as to what journals fell into what fields. The most interesting statistics were those for fulltext article downloads which showed that for all the institutions involved, the price per use came down. The plan now is to use the methodology that was developed for an ongoing analysis of the usage and cost per downloads for these academic institutions. Similar studies are being conducted elsewhere (such as a Couperin study in France).

As far as global metrics are concerned, the Impact Factor is well-established and endorsed by funding agencies and researchers even if it is considered by some as being misused, overused and misrepresented.

---

[9]Release 3 came out in August 2008.

[10]http://www.jisc.ac.uk.

There is a growing body of usage data and now might be the time to see if a usage-based alternative perspective to the Impact Factor could be valuable. *Science Citation Index (SCI)*[11] is by nature selective. Librarians subscribe to many journals not covered in *SCI*. Also citation data on the whole measure the value to the research community that publishes and cites, but do not measure the value to practitioners and students.

To try to find solutions, the United Kingdom Serial Group (UKSG) has funded a project to assess the feasibility of developing and implementing Usage Factors through a survey to determine the level of support from author, librarian and publisher communities. The survey was completed in 2007 and the report submitted to UKSG.[12] The survey was conducted in two stages. First, a number of in-depth interviews were conducted with some 20 opinion-makers in the scholarly publisher, librarian and author/editor communities, then a wider-based survey covered 155 librarians and 1,400 academic authors. All surveyed authors and librarians thought that the usage factor would be helpful in assessing the value status of journals but the feedback from publishers was more mixed. Right now, publishers are quite comfortable with the way their journals are ranked on the base of Impact Factors. The question of the specified period that should be covered is still open.

Lack of trust between publisher-only and librarian-only organizations still hampers the compilation of Usage Factors and there are some concerns that this could be the base for another profit-making venture or industry. However, there is stronger than expected support, especially from publishers. Nevertheless, there are some caveats. It might be easier to inflate online usage data than to inflate citation data. It is easier to cover citations than to cover usage data because online publishing is becoming more distributed. The usage is no longer captive on a few big publishers since there are institutional repositories and other sources of articles.

To conclude, COUNTER has to face a number of count issues. A number of things will affect usage, such as interfaces. COUNTER has implemented data filters to solve the problem of duplicate usage such as different formats for the same article, and is coming up with a report that covers the number of unique article requests in a session. Also, there is the issue of archives since librarians are now purchasing back years separately from publishers. A new report has been created specifically for archives that cover specified time periods. To be credible in the future, COUNTER has to have usage reports for repositories. NISO is developing SUSHI,[13] a new standardized automated usage harvesting protocol. Also, at the request of consortia, the consortia reports are being improved. And finally, despite some reluctance, COUNTER is going to adopt XML for its next releases. And COUNTER is now looking into covering other content categories. Lastly, as an important reminder, COUNTER is only concerned with usage statistics and is not planning to move into other areas. Also COUNTER is supported by its members so new members are more than welcome and membership application forms are available on the COUNTER website.

Discussion

The issues that were raised concerned mainly individual journal subscriptions that are still mostly in print format and therefore not counted, and online journals with small circulation where the cost is likely to be higher.

---

[11] http://scientific.thomsonreuters.com/products/sci.

[12] For more information about the study, see: http://www.uksg.org/usagefactors.

[13] http://www.niso.org/workrooms/sushi.

Scholarly impact: From ranking to assessment

Johan Bollen
Los Alamos National Laboratory

Scholarly evaluation IS very important for research organizations since it is used to make appointment and funding decisions, monitor trends to identify emerging foci and prioritize activities. Decisions are mostly subjective and qualitative methods are difficult to scale, which explains the strong demand for quantitative indicators such as performance, quality and status metrics. One thing that is very often overlooked when talking about scholarly status or impact is that there are many possible facets to scholarly impact such as popularity, prestige, notoriety, production rate or novelty. It is then a question of combining these facets into indicators of similar aspects of scholarly impact. The focus should not be on one particular metric or type of metrics but rather on how these metrics correlate and correspond to the scholarly aspects we are trying to measure. We have to look at the scholarly life-cycle to see when we can acquire useful evaluation data. From an idea, the researcher goes on to reading existing literature, does his research, writes and submits the paper, is peer-reviewed, published and cited. There are three moments in the process when evaluation data can be acquired. First, during the reading stage when researchers go to the library to read up on the topic (usage data), then when the paper goes through the reviewing process (review data) and finally when the paper is published (citation data). In the context of the MESUR project,[14] the focus has been on usage and citation data. An important point is the delay between the time a paper is published and the time it is cited. In fact, when we look at citations we are looking at the past and this is a serious problem for young researchers. Their best papers could have been published a year ago and may have not yet accrued citations even though they have already been read many times.

Today, bibliometrics provide a one dimensional model where ranking is based on a single criterion of value, i.e. citation rates (ISI's impact factor) derived from a single data source (citations). But scholarly impact is a complicated notion, a multifaceted concept that cannot be captured by monolithic metrics. We need to shift to "assessment 3.0", assessment fit for the digital era, and situate resources in a value landscape of different value indicators. There is a need for a multidimensional rather than linear model of scholarly status assessment that is relational rather than frequentist and where notions of trust, centrality, reputation and so on dominate. The question is what dimensions to choose. The term "dimensions" instead of "metrics" is used on purpose because each of these dimensions could be linear combination or other combinations of many indicators deemed to be relevant.

This presentation will be in two parts. First, the MESUR project which explores how to combine usage, citation and bibliographic data in a very large semantic network to survey the widest possible range of indicators of scholarly impact, their relations, their correlations and to extract the dimensions from that correlated structure. And second the study of the peer review process.

The MESUR project, funded by the Andrew W. Mellon Foundation, is designed to investigate a large variety of different metrics to determine which facets or aspects of scholarly impact they best correspond with. We are doing this because the present landscape of scholarly assessment is dominated by citation data which suffers from a number of limitations in terms of their applicability and what facet of scholarly impact they indicate. Citation data only concerns a subset of the scholarly community's authors and the gathered journal citation data concerns a relatively small subset of the total artifacts produced. As an example, there are no citations for Linus Torvalds who created the Linux operating system, yet millions

---

[14] http://www.mesur.org.

of downloads for Linux. Citation data is furthermore delayed in time (+ one year) due to significant publication delays. In terms of metrics based on citation data the impact factor is most often relied on; it is a simple average citation rate that may or may not be the best indicator of scholarly impact from citation data.

On the other hand, usage data has many advantages. It can be applied to a much larger subset of the scholarly community as it covers non-publishing practitioners, more artifacts are accessible and usage is recorded nearly immediately upon publication at the point of service. Usage data can be used to calculate numerous highly informative usage-based metrics. Given the value of usage data various initiatives have been started to standardize its recording, aggregations and application (COUNTER, IRS (*Interoperable Repository Statistics*), SUSHI, CiteBase). However, several important issues related to usage data are left unexplored at this point. First, usage data tends to reflect the usage of the sampled communities. Usage-based scholarly assessment is thus subject to sample bias. Second, a multitude of metrics can be calculated from usage data. Which best indicate which facet of scholarly impact? These problems can be addressed by creating a representative reference data set of usage data that is sufficiently large-scale to be representative of the scholarly community so that metrics can compete and be evaluated on the same standardized substrate. When different metrics are calculated on the basis of this reference data set, the difference in rankings is not due to sample biases in the underlying usage data, but the metrics themselves.

One of the objectives of the two-year MESUR project is to acquire usage and bibliographic data for the largest possible scholarly community, create the largest possible artifact collection, create reference data sets of citation, usage and bibliographic data on the basis of which we can calculate a large variety of different metrics, and then correlate these metrics and determine whether they express clusters corresponding to similar or dissimilar aspects of scholarly impact. There are four phases to the project.

Phase I concerns the modeling of the scholarly communication process by means of an RDF/OWL ontology.[15] The MESUR ontology is based on three basic notions: Documents, Agents and Contexts. The latter are sub-classed into Events and States where the State is relatively permanent and the Event of a particular duration. The MESUR ontology was publicly released and can be accessed at: http://www.mesur.org/schemas/2007-01/mesur/.

Phase II, the "begging phase", concerned the collection and aggregation of usage data from some of the world's most significant publishers, aggregators and institutional consortia to achieve a representative sample of the scholarly community. The reference data set we collected involves about 70 million users, 50 million documents including metadata and references, 700 million usage events and about 1 billion citations. Relationships now amount to 10–15 billion triples. This sample originated from a variety of providers, i.e. an international large-scale sample of publishers, aggregators and institutions and thus required extensive normalization and deduplication. Publisher data often consisted of COUNTER statistics but also included large-scale, article-level usage data; aggregator data was mostly article level data; institutional data was easy to obtain and often came in the form of SFX logs, but also included some custom log formats with demographic and location data for individual users. Link resolvers are a real blessing since they have a standardized way of recording, representing, harvesting and aggregating usage data. Compared with the alternatives, they are not only relatively free of commercial interests but the institutions that own the data are often more willing to share them since issues of ownership and user privacy are lessened by how link resolvers record usage data.

---

[15]http://www.w3.org/TR/rdf-concepts, http://www.w3.org/TR/owl-features.

Phase III concerns data representation and storage. Once the data has been acquired, normalized and deduplicated, it needs to be stored. Relational databases are not suited to data of such order. The MESUR ontology defines the representation of our data in terms of RDF triples. An RDF triple is a beautiful notion; it is in fact the most fundamental expression of a fact, a combination of three items (subject, predicate, object). Once our citation, usage and bibliographic data have been converted into RDF triples, they can be stored as a network of triples, i.e. a semantic network. The later allows us to perform complex inferences on the basis of our data, which is quite cumbersome to do with relational databases. In recent years, there has been a surge of commercially viable triple store databases. MESUR actually works with Franz's AllegroGraph.[16] The access interface is similar to any other sequel databases but with the addition of Reasoners.[*]

Phase IV concerns the characterization of our reference data set and survey of a wide variety of different usage- and citation-based metrics. It is the phase where we perform a statistical analysis of the created networks, derived from both citation and usage data. The availability of these networks enables a new paradigm of scholarly assessment. Metrics based on simple counts and averages are no longer sufficient. We want to move to scholarly assessment that acknowledges the intricate networks of resource relationships that can be deduced from how users interact with online services. This is not new since Amazon.com, Flickr, netflix.com and the like already connect users and resources on the basis of usage patterns and use these relations to build advanced services. Our metrics can do the same from the networks we extracted from our usage data; they can examine existing connections between people and resources from many different angles and rank resources on the basis of their network positions rather than use simple "hit" counts or averages. The last part of phase IV concerns the metrics survey. Briefly, we have found in several case studies that different metrics can indicate different aspects of scholarly impact. MESUR has calculated many different metrics of scholarly impact. What do they mean? Our investigation is based on the following observation: when different metrics produce similar rankings, they express similar aspects of scholarly impact. When they diverge, we know they express dissimilar aspects of scholarly impact. Clusters of metrics with similar outcomes thus indicate an important facet of impact that is measured by several different metrics. A taxonomical analysis of metric similarities and a principal component analysis of their correlations will reveal which underlying aspects of impact exist, and which metrics best measure those aspects. When we cross-validate these results to the ISI impact factor and the COUNTER statistics we can interpret them in terms of what we already know.

In conclusion, the MESUR ontology has been defined and is now available online. MESUR has achieved agreements with 4 major university consortia, 3 major aggregators and about 5 major publishers, and enjoys the greatly appreciated support of the COUNTER project. Usage data is being loaded and normalized and is being contributed among others by the University of Texas (about 3 million for one year) and BioMed Central (24 million for two years) and one of the anonymous providers has been contributing over 50 million usage events par week. The objective is to achieve a database of about 10–15 billion triples. It is obvious that there are issues and challenges. Publishers are not keen to part with data since they are concerned over ownership and business implications, aggregators are willing to share but with publishers' agreement, while institutions are most willing to share data and are interested in the results of studies to demonstrate the value of their communities. Noise and spurious events are being weeded out. Initial results of the comparison of about 50 different metrics of scholarly impact are expected in the next few months.[**]

---

[16]http://agraph.franz.com/allegrograph.

[*]As demonstration, a SPARQL query is shown that calculates the ISI impact factor.

[**]Due to lack of time, the presentation did not go into the peer review process.

Discussion

The discussion mainly concerned clarifications of some points of the presentation, the issue of time stamps and the eventual use of demographic data for sociological studies.