Qualitative and Quantitative Methods in Libraries (QQML) 1: 321 –330, 2014

A method for journal collection management and the limitations of reality

Miriam Lorenz[1]

[1]Cologne University of Applied Science, Faculty of Information Science and Communication Studies

Abstract: Academic libraries are constantly facing new challenges, which require as much time and money as other processes to be more effective. A big part of the media budget is spent on licensing (electronic) journals. Even though libraries invest a lot of time in selection processes of journals, it is still not possible to combine the data (like statistics or JIF) of the different actors of scholarly communication in a multidimensional model and to produce a balanced base for decision making. This work shows the requirements of such a model and explains, which deficits exist at present and why the Data Envelopment Analysis (DEA) is a suitable method for the selection process of (e)journals.

Keywords: Journal Collection Management, Journal Management, eJournals, Usage Statistics, Journal Impact Factor (JIF), Data Envelopment Analysis,

1. Introduction

Communication is the fuel of science. In most cases, the method of choice is publication or reading of an article in a commercially produced scholarly journal. So one of the responsibilities of academic libraries is to provide access to scholarly journals. The processes of journals are traditionally organized in a way that publishers offer journals, which are licensed by libraries and are provided to scientists, who in turn publish in that journal.

Since the first electronic journals were published more than ten years ago, the conditions of this circle are changing constantly. At the moment around 2/3 of all journals (depending on the field) are additionally or exclusively available electronically (Johnson, Luther, 2007) and their bibliographical data can be found in search engines, databases or publisher portals for free. But on the other hand there are full texts that can only be read against payment (License, Pay-per-View). The broad access to bibliographical data whets the appetite: scientists expect a direct, quick and free access to all full texts they've found.

Received: 28.5.2013 / Accepted: 29.3.2014 ISSN 2241-1925

Although through the distribution of the open access movement around 20% of all journals are available for free (Björk et.al, 2010), the rising journal prices and the stagnant budgets of libraries cause an on-going conflict.
Academic and special libraries often cannot finance the broad access required by users (CIBER Research Limited, 2011) and therefore need to make large-scale licensing decisions and justify them to users and funders.
The goal of libraries in journal collection management is to achieve optimal coverage of electronic journals with the given money.
This study tries to give an overview of the measures and methods with their strengths and weaknesses used for journal management in Academic libraries. The first step is an analysis of functions, actors and general conditions of journals. Furthermore the different methods are analyzed specifically looking at formal and content issues.
As a conclusion "Data Envelopment Analysis (DEA)" will be presented. It was developed for measuring production efficiency. The intention is to show that DEA is able to reduce the effects of deficits of the methods and measures used so far and that DEA is not only useful but necessary for a collection management that really meets user needs in libraries.

2. Journals: functions, actors and processes

„The scientific and technical community should be proud of the journal and its significant contribution to the advancement of science and technology." (Walker, Hurt 1990). This emotional sounding last sentence of a chapter written by Walker and Hurt underlines the exceptional position that journals have in scientific society.
Journals fulfill four „traditional" functions: (Rowland, 2002): registration, certification, dissemination and archiving. With the publishing of a manuscript in a journal, the findings are linked to an author and build his reputation. At the same time the quality of the findings is controlled through the peer review process and / or the specialists in their role as an editor. This selection process is an assessment of the current relevance of the topic, besides the guaranty of scientific quality. The formal structure of journal titles allows the detection of an article so it can be identified by a citation. Furthermore these contents should be able to be stored, found and read permanently.
All of these functions are identical to the functions of scholarly communication and thereby show the importance of journals for the communication of scientists.
Journals are more than just a collection of scientific findings. Journals are central means of communication in science and are determined by a variety of actors. The journal collection management of libraries holds a key position for distribution of journals. Figure one shows the main actors, the needed processes and the relation of the operations of every actor.

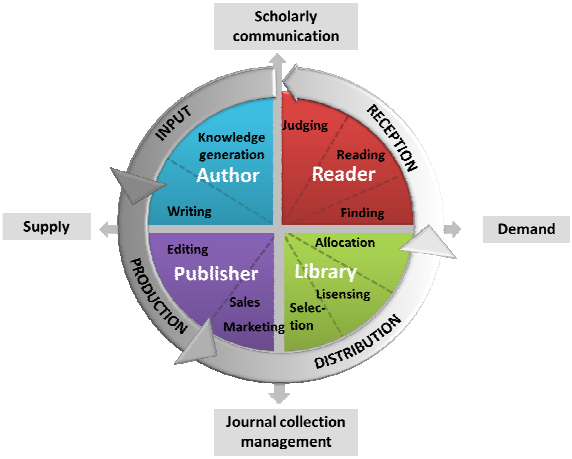

Fig. 1 Value-added chain of journals

Due to opportunities of electronic publishing and the open access movement the scientific community experiences fundamental structural changes, even though the tasks and processes needed stay the same; only the responsibilities for and the organization of single actors are changing. These roles follow the processes of commercial journals. Other actors like subscription agents are excluded, because their role or operations differ very little from those of publishers.
As shown in figure one, the value-added chain of journals starts with knowledge creation of the scientist, who writes a manuscript and delivers it to the publisher as an author. The publisher distributes it to other scientists so it can be reviewed or edited by them. After that the publisher is responsible for the professional "look" of that manuscript. That means for example formatting a special layout, bundling different articles to one issue, in some cases even printing or presenting the bibliographic data in different search engines and databases. The selection and licensing of journals with access to full texts (or to print versions), the provision of bibliographic data of the journal collection in OPACs or special reference tools like the „Elektronische Zeitschriftenbibliothek (EZB)" or LinkResolver systems. Therefore it is possible for the scientist to find, read and judge high-quality material and to cite it in his own publications. These processes in context build also a kind of cycle, where the linking element is the scientist as the reader and the author at the same time.

So there are complex structures and multidimensional requirements which have to be handled by the journal collection management of librarians. They need to buy journals with given money in a monopolistic, concentrated market (Schimank, Volkmann, 2012) and must try to meet the needs of readers, who are going to be authors.
By doing so, every actor produces data and indicators deliberately or as a side effect, which should be used in journal collection management of libraries

because of the dependencies described. But in fact they are taken into consideration with different intensity and priority.

3. Measured values of journals

For every actor in the journal process (author, reader, publisher) the library provides access to different data:

Author

The choice of authors where to publish and what to cite has a special position.
Already in 1934 the distribution of articles of a single topic in different journals (Bradford's Law of Scattering) was developed to identify core journals of a field. The law bases on the probability that scientists prefer specific journals for the collaborative problem solving. It is the mathematical proof of the known Pareto principle of libraries (80% of the needs, can be met by 20% of the collection).
In bibliometric analysis, the citations used in articles by the authors are evaluated. The author's decision to cite a publication is therefore interpreted as an aware and particular positive judgment of the cited work. The number of citations is used to generate different indicators, whereby the number of citations of a publication, journal or author is related to a given period of time or population.
The Journal Impact Factor (JIF), which was developed by Eugene Garfield and Irving H. Sher in 1961, gains the broadest circulation. It is defined as the number of citations of the articles of a journal in a certain period of time, divided through the number of possible „citationable" articles of this journal in the last two (or five) years. The JIF was introduced, „to help select additional source journals [for the Science Citation Index]." (Garfield, 2006) Even though Garfield himself points out repeatedly the limited significance of the JIF and the calculation method or the use of the JIF is criticized in numerous studies (vgl. z.B. Vanclay, 2012), the JIF became the standard indicator for some fields (for example medicine, biology, ...). The JIF is used as a sign of quality for journals, the valuation of applications (Garfield, 1998) or even the selection of journals in libraries.
Because of the constant criticism and motivated by the numerous options of potential applications, a lot of different indicators were developed as a supplement or alternative to the JIF. While the Journal Impact Factor uses only the data of the (today commercialized) database „Web of Science", the so called Altmetrics should evaluate the citation based on the social web (Priem et. al., 2010). All in all, the number of developments prove the importance of the citation of a publication in the scientific community.
Furthermore there exist many indicators like the Hirsch factor (Hirsch, 2005) or Lotka's Law (cf. Lotka, 1926), which try to create statements about the scientific status of the author himself. But because these statements just affect journals indirectly, they are not able to be used in the journal collection management.

Reader

While the evaluation of citation data have belonged to the daily routine in academic life for decades, user generated data have been taken into account since the electronic form of journals allows an automatic counting.

The measurement of usage of print journals was and is connected by extreme effort: in libraries, they counted orders and circulations, introduced temporarily day loans, observed the user or pleased him to tear off a piece of paper at every usage (Schümmer, 1999) or measured the level of displacement of a journal at the end of every day (Umstätter, Rehm, 1984). The required effort caused that data was collected very seldom and with small samples.

This didn't change after introducing and distribution of electronic journals. The statistics of the different publishers were too different and not comparable. The success story of usage statistics started with the common project of libraries and publishers „COUNTER" in 2004. The project has set and is still developing standards in structuring tables, defining numbers, terms or also file format and delivering form.

Even though COUNTER defines different numbers for journals, most of the libraries concentrate on the number of full text downloads per month per journal. An analysis of downloads based on articles – like it is done in statistics in the context of open access – does not (yet) happen for commercial journals. Altogether the ratio of criticism and expansion at the same time seems to be comparable to the dynamics in behalf of the JIF. So there are deficits in technical use, form and content for the standard itself and also for the creation and use of the statistics (cf. Lorenz, 2010). But at the same time the usage statistics have been used for the „Usage Factor Project" since 2007. The goal is to establish an impact factor based on download statistics in addition to the JIF or the Altmetrics (Shepherd, 2012).

Despite the "real usage" of journals, the needs and wishes of users are more or less systematically recorded via talks, interviews or surveys (cf. Kent et. al., 1979). As well as the described quantitative data, this method includes strengths and weaknesses. Surveys offer the opportunity to collect knowledge of field experts and to get new information and perspectives. On the other hand there is also the opportunity that the respondents just look at their personal situation, financial support or specific topic or that their answers are just motivated by the lack of time or a political strategy. Furthermore the answers could be biased through interpersonal effect or experiences with the libraries in other areas (Blake, Schleper, 2004).

Publisher

Apart from the input or reception based data of authors and readers, the publishers provide a huge number of production data. Even though the interest focused on price, information like the number of issues, the year of the first publishing or the total number of all journals of a publisher, could be useful to evaluate a journal. Surprisingly the price is the hardest data to get. Even when a pricelist is published, the specific price is hard to calculate because of contract

types like the „Big Deal“, special conditions for archives and discounts in the context of consortional agreements (Day, Davis, 2009).

Interim conclusion
There is a lot of different data about journals that is tested, discussed and developed in a vital research landscape. On the other hand every number is just a sign for one aspect of the process or for one motivation of one actor. Also all data, even if it is measured or raised, have deficits and just give evidence to small sections of journals. What does that mean for the use of data in the journal collection management? How much do the current procedures respect the limitations?

4. Journal collection management in libraries

According to a survey of 155 libraries within the Journal Usage Factor Project 2007/2008 (Shepard, 2011) are the reasons for or against the licensing of journals (in descending order; most important first):

- requests of the library user,
- usage statistics,
- price,
- cost per download,
- reputation and
- Journal Impact Factor.

Studies in the USA (Baker et. al., 2008; Cox, 2011), New Zealand (McDowell, 2004) and Austria (Pavlovic, 2012) have shown that usage statistics are the main (and only) factor of decision making in libraries.
Simultaneously studies demonstrate that the decisions differ widely if only one data is used. (Nisonger, 2000; Shepherd 2012).
Chung as well as Day and Davis did complex studies in 2009, where they tried to put different data in one model together for decision making. Chung converted the single numbers into 100er-units. The results showed that the equal weight of all factors lead to unbalanced outcomes. In the study of Day and Davis, the single data of every journal had been so different that in the end the decision was based on just one factor (usage).
At the moment, the heterogeneous and partly conflicting indicators about journals neither can be analyzed simultaneously for the journal collection management nor can a single factor inducing reliable results or the deficits of the indicators displayed.

5. Reasons for Data Envelopment Analysis as the ideal method for journal collection management.

A method, which is used successfully in the journal collection management, needs to meet the following requirements: According to the focus of a library, there has to be the opportunity to weigh the indicators differently. Moreover it should be possible to consider several inputs or outputs with different units at

the same time. The results should be ordinally scaled at minimum (i. e. sortable), without a software installation and feasible as easy as possible.
The Data Envelopment Analysis (DEA) allows the combination of different data in one ranking. It determines the best practicable efficiency conditions and helps to identify the „best practice“ for every unit.
In 1957 DEA was applied for the first time by Farrell (Farrell, 1957) to optimize the efficiency of production processes. The following period of worldwide economic growth and due to the feeling of unlimited resources the analysis of efficiency lost its importance. So the model was revived in the middle of the 70es by Charnes and Cooper and refined by Rhodes in 1978 (Charnes et. al., 1978). This time it weren't production processes, but hospitals and health care centers that should be tested according to their efficiency. This total different application with a service instead of a technical focus and the huge number of units and factors made some revisions necessary. This basic form of the model (CCR (Charnes, Cooper, Rhodes)) is still preferred today, besides numerous other variants.
The analyzed unit (Decision-Making-Units, DMU) can be defined as variable. Besides operational units also single processes or projects of a special kind can be sorted as DMU and can be taken as a base for modeling (Hoffmann, 2006).
The implementation of a DEA is not just pure calculation: it is a process, which includes the analysis of the processes or DMUs for the identification of the ideal input or output factors, the interpretation of the meaning of efficiency, the discussion of the results and the development of strategy for the realization of Best Practice.
Because of these attributes, there are more than 2000 studies published on DEA and it was used in many different areas like during the financial crises (Streit, 2009) or in Pangasius farms in Vietnam (Le Thai Hanh, 2009), but also to compare research performance. The DEA had also been successfully established for the comparison of libraries (Mann, 1997; Noh, 2011) and universities (for example in Germany in public funded projects like HELENA and EUMIDA).
The DEA meets every criteria and the next step of the project will be an exemplary application in journal collection management.

6. Reasons against Data Envelopment Analysis as the ideal method for journal collection management

The applicability of the Data Envelopment Analysis for journal collection management seems to be likely, by just looking at the theoretical elements. But in practice the experiment of implementing the DEA failed before it really started. Analyzing the processes of journal management all over the world showed that in the actual business models and negotiation practices of 'big deal', packages and large consortia, licensing single title lists is neither usual nor possible in most cases. And if possible, it's more expensive.
Publisher and librarians doesn't seem to be interested in offering the most relevant journals but all journals. The importance of selection decreased a lot in times of electronic journals, high-speed internet connection, discovery systems, flat rates, patron driven acquisition and new business models like Open Access.

In this project, there were no reasons found for investing a lot of effort and time in collecting indicators, learning how to use DEA and knowing what could be the ideal journal titles. Because reality shows that there is no place for the an individual title list, whether this list is made by DEA or any other method.

6. Conclusion

Journals take a special position in the scientific community as the most important communication media, which includes a lot of different actors. In journal collection management, academic libraries play the role of the distributer, who – despite of the monopoly position of the publishers and the limits of their own budgets – needs to find the perfect collection for the user. In some cases the libraries collect data from every actor with different significance and units at their disposal for the decision. The strengths and weaknesses of the data and the influence of every actor cause that a decision, based on one factor, becomes unilateral. But in most cases libraries are not licensing a collection, which is most specific but is as broad as possible.

The Data Envelopment Analysis would meet all of the requirements for a combination of the data. So if there will be a need for single title licensing in the future ever again, we are prepared.

References

Björk B-C, Welling P, Laakso M, Majlender P, Hedlund T, et al. (2010): *Open Access to the Scientific Journal Literature: Situation 2009.* PLoS ONE 5(6): e11273. doi:10.1371/journal.pone.0011273

Blake, J. C. and Schleper, S. P. (2004). *From data to decisions: Using surveys and statistics to make collection management decisions.* Library Collections, Acquisitions, and Technical Services (28), Nr. 4, (S. 460-464).

Charnes A.; Cooper, W. W. and Rhodes, E. (1978). *Measuring the Efficiency of Decision Making Units*. European Journal of Operational Research (2), Nr. 6, (S. 429–444).

Chung, H.-K. (2009). *An analysis model of creating a core journal collection for academic libraries.* Library Collections, Acquisitions, and Technical Services (33), Nr. 1, (S. 17–24).

CIBER Research Limited (2011a*). Access to scholarly content. gaps and barriers*. URL: http://rinarchive.jisc-collections.ac.uk/system/files/attachments/gaps_final_report_low_res.pdf [22.03.2013].

Day, A. and Davis, H. (2009). *Collection Intelligence: Using Data Driven Decision Making in Collection Management.* Proceedings of the Charleston Library Conference, (S. 64-80).

Farrell, M. J. (1957). *The measurement of productive efficiency*. Journal of the Royal Statistical Society (120), (S. 253–281).

Hirsch, J. E. (2005). *An index to quantify an individual's scientific research output.* Proceedings of the National Academy of Sciences (102), Nr. 46, (S. 16569–16572).

Hoffmann, C. (2006). *Die Data Envelopment Analyse (DEA) und ihre Anwendungsmöglichkeiten zur vergleichenden Effizienzanalyse im Forstwesen,* Wien: Universität für Bodenkultur Wien.

Garfield, E. (1998). *The Impact Factor and Using It Correctley.* Der Unfallchirurg (48), Nr. 2, (S. 413).

Garfield, E. (2006). *The history and meaning of the journal impact factor*. Journal of the American Medical Association, 295(1), 90–93

Johnson, R. K. and Luther, J. (2007). *The e-only tipping point for journals. What's ahead in the print-to-electronic transition zone.* URL: http://www.arl.org/bm~doc/electronic_transition.pdf [22.03.2013].

Kent, A. and et al. (1979). Use of Library Materials: The University of Pittsburgh Study. Dekker, New York.

Le Thai Hanh, B. (2009). *Impact of financial variables on the production effiency of Pangasius farms in An Giang province, Vietnam*, Vietnam: Nha Trang University.

Lorenz, M. (2010). *Elektronische Nutzungsstatistiken. - oder: Zu Fuß ist es kürzer als über den Berg.* In Mittermaier, B. (Hrsg.).), *WissKom2010 "eLibrary - den Wandel gestalten",* (S. 137-149).

Lotka, A. (1926). *The frequency distribution of scientific productivity.* Journal of Washington Academy of Science (19), (S. 317–323).

Mann, G. M. (1997). *Efficency Evaluations of North American University Libraries by Data Envelopment Analysis,* Montreal: McGill University.

Nisonger, T. E. (2000). *Use of the Journal Citation Reports for serials management in research libraries. An investigation of the effect of self-citation on journal rankings in library and information science and genetics.* College & Research Libraries (61), Nr. 3, (S. 263–275.).

Noh, Y. (2011). *Evaluation of the resource utilization efficiency of university libraries using DEA techniques and a proposal of alternative evaluation variables.* Library Hi Tech (29), Nr. 4, (S. 697-724).

Pavlovic, K. (2012). *Verwendung von Nutzungsdaten für Lizenzentscheidungen bei Netzpublikationen.* URL: http://edoc.hu-berlin.de/series/berliner-handreichungen/2012-334/PDF/334.pdf. [22.03.2013]

Priem, J.; Taraborelli, D.; Groth, P. and Neylon, C. (2010). *Altmetrics: A manifesto (v.1.0).* URL: http://altmetrics.org/manifesto. [22.03.2013]

Rowland, F. (2002). *The peer-review process.* Learned Publishing (15), Nr. 4, (S. 247–258).

Saha, S.; Saint, S. and Christakis, D. (2003). *Impact factor. a valid measure of journal quality?* Journal of the Medical Library Association : JMLA (91), Nr. 1, (S. 42–46).

Schimank, U. and Volkmann, U. (2012). *Die Ware Wissenschaft: Die fremdreferentiell finalisierte wirtschaftliche Rationalität von Wissenschaftsverlagen.* In Engels, A. und Knoll, L. (Hrsg.), *Wirtschaftliche Rationalität – Soziologische Perspektiven,* (S. 165–183), Wiesbaden: Springer.

Schümmer, V. (1999). *Nutzungsanalyse von mathematisch-naturwissenschaftlichen Print-Zeitschriften an der Universitäts- und Landesbibliothek Düsseldorf. Mit einem Ausblicka uf die Nutzung elektronischer Zeitschriften.* Bibliotheksdienst (33), Nr. 9, (S. 1475–1494

Seidenfaden, L. (2007). *Ein Peer-to-Peer-basierter Ansatz zur digitalen Distribution wissenschaftlicher Informationen,* 1. Auflage, Göttingen: Cuvillier.

Shepherd, P. (2011): *The Journal Usage Factor project. results, recommendations and next steps.* URL: http://www.projectcounter.org/documents/Journal_Usage_Factor_extended_report_July.pdf [22.03.2013]

Shepherd, P. (2012). *A new metric for measuring journal impact – Usage Factor.* STM Innovations Seminar 2012 – Innovation Impacts in STM: New Metrics, Big Data, Cool Apps. *URL: http://www.stm-assoc.org/2012_12_07_Innovations_Shepherd_Usage_Factor.pdf.* [22.03.2013]

Streit, B. (2009). *Lehren aus der Finanzkrise. Individualvergütung anhand der Data Envelopment Analysis.* URL: http://www.postbank.de/csfiles/711_60_streit_5.pdf. [22.03.2013]

Umstätter, W. and Rehm, M. (1984). *Bibliothek und Evaluation.* Nachrichten für Dokumentation (35), Nr. 6, (S. 237–249).

Vanclay, J. K. (2012). *Impact factor: outdated artefact or stepping-stone to journal certification?* Scientometrics (92), Nr. 2, (S. 211–238).

Walker, R. D. and Hurt, C. D. (1990a). *Scientific and technical literature. An introduction to forms of communication*, Chicago: American Library Association.