

How to assess the impact of an electronic document? And what does impact mean anyway? Reliable usage statistics in heterogeneous repository communities

This document is a preprint of the formal publication with the same title, written by the same authors in:

OCLC Systems & Services 26 (2), p. 133-145

www.emeraldinsight.com/10.1108/10650751011048506

DOI: 10.1108/10650751011048506

Publisher: Emerald Group Publishing Limited

Authors:

Ulrich Herb - Saarland University and State Library, Saarbrücken, Germany (corresponding author)

Eva Kranz - Saarland University and State Library, Saarbrücken, Germany

Tobias Leidinger - Saarland University and State Library, Saarbrücken, Germany

Björn Mittelsdorf - Saarland University and State Library, Saarbrücken, Germany

Purpose

Usually impact of research and researchers is tried to be quantified by using citation data: Either by journal-centred citation data as in the case of the journal impact factor JIF or by author-centred citation data as in the case of the Hirsch- or h-index. The paper discusses a range of impact measures, especially usage based metrics. Furthermore the authors report the results of two surveys. The surveys focused on innovative features for open access repositories – with an emphasis on functionalities based on usage information.

Design/methodology/approach

The first part of the article analyzes both citation-based and usage-based metrics. The second part is based on the findings of the surveys: One in form of a brainstorming session with information professionals and scientists at the OAI6 conference in Geneva, the second in form of expert interviews mainly with scientists.

Findings

The results of the surveys indicate an interest in the social aspects of science like visualizations of social graphs both for persons and their publications. Furthermore usage data is considered an appropriate measure to describe quality and coverage of scientific documents, admittedly the consistence of usage information among repository has to be kept in mind. The scientist that took part in the survey also asked for community services, assuming these might help to identify relevant scientific information more easily. Some of the other topics of interest were personalization or easy submission procedures.

Originality/value

This paper delineates current discussions about citation-based and usage-based metrics. Based on the results of the surveys it depicts which functionalities could enhance repositories, what features are required by scientists and information professionals and whether usage-based services are considered valuable. These results also outline some elements of future repository research.

Acknowledgments

The authors would like to thank Philipp Mayr, Sven Litzcke, Cornelia Gerhardt, the experts who prefer to remain anonymous, and all participants of Breakout Group 6 at the OAI6 conference.

Introduction

As Harnad (2008) explains, the meaning of an impact measure can only be determined by correlating said measure with either another measure (construct validity) or an external criterion (external validity).

But which data should be employed to check impact measures like the Journal Impact Factor or the Hirsch-Index?

The range and divergence of potential validating data sets, respectively their object selection, object granularity, and complexity of calculation instructions, reveal that the scientific value of a document has multiple dimensions (Moed 2005b). The actual choice depends on the perspective from which the impact (usefulness) question is asked.

Galyani Moghaddam & Moballeghi (2008) give an extensive overview over possible methods. But a matter seldom addressed is the concrete motivation for impact measurement, a question that can help defining what impact should mean in a specific context.

Statistical predictions and especially quality assessments can become self-fulfilling prophecies, especially if the numbers are already in use. If we use the height of academics as quality criterion in calling new staff members, academic teams naturally will become taller. A later study of height and institutional quality will find a high correlation of quality and height not because of the inevitable working of things but because this relation was man-made and the variables were confounded to begin with. Nicholas addresses this issue commenting on the Journal Impact Factor in an interview conducted by Shepherd (2007).

Scientometric perspective: Eugene Garfield devised the Journal Impact Factor (JIF) as a filter criterion to determine whether a journal should be included in the Science Citation Index (SCI) sample (Garfield 2006). At that time each journal in the sample meant a serious amount of work. The restriction on a finite number of journals was not only a matter of quality but also of practicability. The assumption is that a higher rate of citations indicates a higher importance/quality/impact of an article but more importantly the journal. JIF is –in the context of journal assessment- presumably superior to simple publication counts as quantity does not depend on quality, but it can be argued that JIF rather describes a journal's prestige, concentrates in established topics, and depends on a certain amount of honesty while it can be easily misunderstood by the naive or corrupted by the dishonest (Brumback 2009).

Evaluation perspective: "The use of journal impacts in evaluating individuals has its inherent dangers. In an ideal world, evaluators would read each article and make personal judgements." (Garfield 2006)

There are two main problems in evaluation:

- Scientific quality and researcher value are assumed to be multidimensional, including characteristics outside the publication activity (Moed 2005b).
- Subjective or personal statements, i.e. evaluation by peers, are not reproducible, and have in modern days the air of bias and distortion about them.

Jensen et al. (2009) investigate career predictors for French scientists in the national research organisation CNRS. Promotions are decided by a peer committee. The correlation of promotion on the one hand and publication and citation measures on the other is highest for the Hirsch-Index. Nevertheless the amount of promotions correctly predicted by h is only

48%. This gap might result from human failure h cannot predict or measurement bias the experts do not succumb to, but presumably a mixture of both.

Therefore the questions are:

- Which variables should be collected in addition to citation metrics?
- How should the variables be weighted?
- How to maximize fairness and openness, or can objective measures and human-made recommendations be synchronised?

The need for multidimensional evaluation is shown by Shepherd (2007) who reports that over 40% of their web survey sample perceive the JIF as a valid measure but over 50% regard the JIF as over-used.

Journal perspective: The motivations of journal editors can be assumed to be purely economic, as only economically sound journals can compete with economically managed journals in a spiral of competition. Mabel et al. (2007) investigated the attitudes of medical editors towards the JIF and their handling of independent variables which are likely to increase their journal's JIF rating. Editors relied chiefly on quality increase of their staff to boost author recruiting, article selection, and author support.

JIF was accepted as status quo but editors expressed their concern that JIF is not useful to impress their practising readership. Thus, they could not solely rely on optimizing their JIF scores. They hope that complementary metrics that represent clinical impact or public advancements will be implemented. Empirical analysis of the interrelation of journal characteristics and journal performance (McWilliams et al. 2005) seem to contradict some of the medical editors' statements. It is rather likely that different circumstances in management science and medicine account for these discrepancies. Further assessment of the properties of different disciplines will improve the transfer of insights.

Library perspective: Electronic documents fundamentally change the mechanisms in libraries. Whereas in former times the library was the keeper of objects and controlled and monitored – sometimes even created- the processes of search, localisation, and usage, they have become an intermediate agent nowadays. These changes might be as trivial as people being able to receive a text without physically visiting the library leading to bulletin boards not being read anymore. Librarians have to adapt by offering telematic versions of their services (Putz 2002). On the other hand the easily adaptable electronic reception desk offers opportunities of personalisation, customisation and portalisation. (Ghaphery & Ream 2000) and (Ketchell 2000) warn that personalisation appeals rather to the professional or heavy user whereas customised views, centred on a specific topic or even specific academic classes, aid the average student user, who shuns high investments (e.g. login procedure, training period) and has to change focus of interest quickly. Additionally metrics can aid subject librarians in the compilation of resources.

Another issue is billing. As libraries no longer have control over objects, they have to rely on external server statistics provided by the publishers and hosts to make licensing and acquisition decisions.

The COUNTER standard (Project COUNTER 2008) is widely used to generate usage reports. Its granularity is rather unfit to help with acquisition decisions. Usage is reported per journal, not per volume, making it impossible to identify irrelevant vintages that would better be bought by article on demand and not included in a flat rate licensing bundle.

COUNTER tends to distort the actual usage, for example Davis & Price (2006) report how interface characteristics of the publisher portal can unjustly increase usage frequency.

Educational perspective: Educational science research in general focuses on the classroom aspects of digitisation.

Collaborative work environments, online instructions, and testing environments are designed and evaluated to enhance the lecturers' efficiency for example with homework management and on the other hand boost student to student and student to tutor communication (Appelt 2001). Electronic resources are produced by students or prepared by the lecturer to be stored and possibly versioned in the system. Course reserve collections are often created independently from library activities as many coursework software systems are designed as closed applications, which cannot be easily connected with other services. The aim of education is on the one hand to teach the curricula but on the other hand emphasis is placed on teaching information and communication technology (ICT) competence (Sikkel et al. 2002). As education relies heavily on text books, contemporary citation measures are not applicable.

The usability perspective is a specialised point of view that can complement the paradigms described above. It is most obvious with education as most education research explicitly includes investigations of ease of use and practicability. On the other hand institutions and organisations in a competitive environment (libraries, universities, and publishers) can improve their strategic position by increasing user efficiency. These can be purely technical aspects (e.g. a user achieving his goal with fewer steps requires less server computation time) but in general it has to be discussed whether the service does fulfil the request at all and whether it meets the user's needs.

Much discussed aspects of usability in information dissemination are recommender services, though their main application is in the commercial area (Montaner et al. 2003). The vast amount of works already available and the increasing growth rate can be assumed to overload the faculties of non-elite information hunters and gatherers (i.e. most students, practitioners, interested private persons, and persons concerned). Even a professional academic researcher can overlook an article and be informed in peer review about his non-optimal library search. But recommenders can not only help to clarify a topic. Content providers are very interested in recommenders that show the visitor alternative objects of interest hoping that he spends more time with the provider's products. This can be a benefit by itself as straying users increase the amount of views and visits, which is reflected in COUNTER statistics as well as in revenues for paid advertisements.

Other aspects of usability include among others visualisation of data, personalisation including user notes and comments being saved for later visits, see Expert Interviews and Brainstorming Session later in this article for further examples.

But valid usage statistics are valuable to all of the perspectives: To scientometry it is an additional database that enables research in construct validity and sociological aspects of citation habits though it has to be emphasised that there is no mono-variant relation between usage and citation (Moed 2005a). Possibly citations and usage are independent dimensions of multi-dimensional impact.

Access of an electronic resource can be measured in real-time and to a certain extent in-house. This should appeal to evaluation committees as well as developers (and usability testers) of educational methods and academic services. Methodologically speaking access and to a lesser degree usage are observable whereas questionnaires and even references/citations are susceptible to bias and distortions based on human language, beliefs and self-consciousness (Nicholas et al. 2005).

Libraries and publishers have always counted their users' activity. It is a simple result of billing. And of course these numbers were used to advertise the journal following the logic of the tyranny of the majority: The journal read by many should also be read by you.

There are problems that have to be addressed:

The observable event in a repository of digitised objects reached via HTTP is the client computer request for said object to the web server.

Neither a human intention nor successful delivery is strictly necessary. There are visits that result from search engines updating their search indices, errantry, and prefetching.

But attributing requests to different individuals is further hampered by technologies like thin clients and proxy servers but also by public search terminals. Thin clients allow their users to interact with software located and executed on a central infrastructure. In case of web browsers this implies that all browser instances serving one thin client cluster are routed via one IP address. Intransparent proxies -nowadays mainly an important aspect of network security- pose the same problem. The obvious solution is to determine one unique user not only via the request IP address but also by utilising session identifiers transmitted as cookies or dynamically created URL arguments. However there is no reliable way to tell to visitors apart who use the same physical machine and account. This is common in educational facilities with search terminals located for example in libraries. It would be necessary to clean the browser's cache each time before a successor begins his work or to identify the account as belonging to multiple persons for example by an appendix to the user-agent header field. Furthermore, aggregated statistics (e.g. author statistic) suffer from multiple instances of one document, but also from print-outs, private sharing of articles, making it very hard for the statistics provider to produce an ecologically valid parameter (Nicholas 1999), see Stassopoulou & Dikaiakos (2007) for a dynamic approach to robot detection.

The heterogeneity of perspectives strongly indicates that a single measure, even a single method, is hardly a reliable decision base. Furthermore this diversity implies that even if one perspective were to reject usage analysis based on scientifically valid reasons this cannot automatically extend to other motivations.

Open Access Statistics (OA-S) and similar projects

Interoperable Repository Statistics (IRS) is a British project tailored to the British repository context. Utilising PERL scripts and the software tool AWStats EPrints and DSpace repository access logs can be analysed. The strength lies in the well prepared presentation possibilities offering various kinds of graphs and granularities (IRS 2007).

MESUR is a research project which has established a large set of usage data by logging activities on publisher and link resolver servers. They aim towards creation and validation of usage-based metrics as well as validity testing of other impact measures (Bollen et al. 2008).

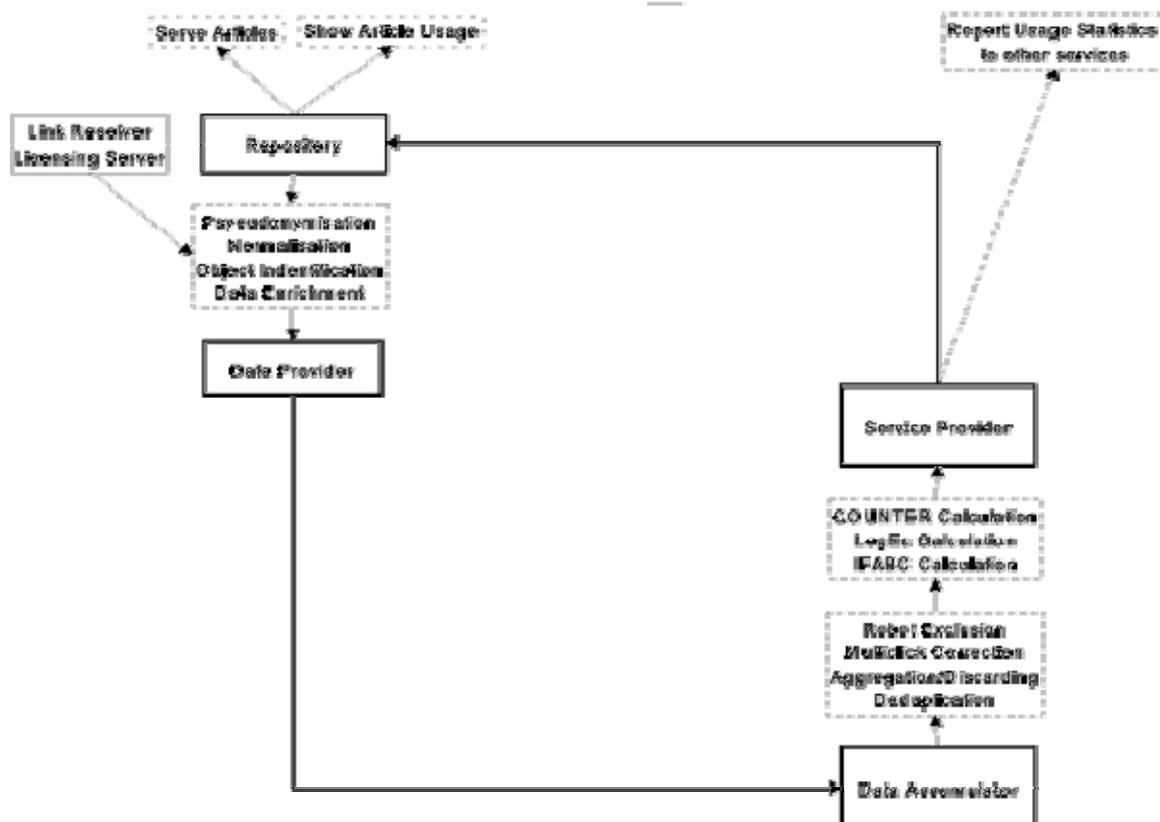
The PEER project investigates the relation between open access archiving, research, and journal viability (Shepherd 2008). To this end PEER measures usage of a closed set of journals made open access by publishers for the project's duration.

PIRUS is developing a standard for article level usage reports that adheres to COUNTER Codes of Practice. A prototype and abstract description were created to enable document hosts to report the raw access data to an aggregating server or process it themselves (Shepherd et al. 2009).

Open Access Statistics (OA-S, <http://www.dini.de/projekte/oa-statistik/english/>) is a project funded by the German Research Foundation (Deutsche Forschungsgemeinschaft DFG) and conducted by the Project Partners State- and University Library Göttingen (Georg-August Universität Goettingen), the Computer- and Mediaservice at Humboldt-University Berlin, Saarland University and State Library, and the University Library Stuttgart. OA-S aims to (1) establish protocols and algorithms to standardise calculations of usage frequency on web-technology based on open access repositories (2) create an infrastructure to collect the raw access data, process it accordingly and (3) supply the participating repositories with the usage metrics. In contrast to IRS, statistical parameters are not calculated locally - so in addition to article level measurements, parameters beyond document granularity can be implemented, like author-centred statistics or usage aggregation over different versions or multiple instances of the same publication like preprint vs. postprint or self-deposit vs. repository copy. This flexibility in scope should ease the combination and comparison of usage statistics and bibliometric indices. Methods and experiences are similar to those of the PIRUS project, but OA-S concentrates on the centralised aggregation strategy and faces an even more diverse repository software ecosystem.

In addition to the bibliometric perspective, the project will specify functionalities to enhance the usability of repositories, e.g. quality valuations, document recommendations, based among others on usage data.

In order to focus on the services most yearned for a questionnaire survey will be conducted to determine actual user priorities. It should be noted that neither the interviews nor the brainstorming were limited to a special perspective or methodology. All ideas were accepted equally.



Expert Interviews

The expert interviews were conducted according to the guidelines given by Bogner et al. (2005). Most experts were identified through their publications. To add publisher and user perspective persons who are involved in journal production and are situated at Saarland University were contacted, too. 5 out of 10 candidates agreed to participate in a loosely structured interview. Interview length ranged from 12 to 57 minutes. No person-centred presentation of results is given to ensure privacy. Most interviews were conducted via phone. All were recorded with explicit consent from the participants and afterwards transcribed to text.

The following list consists of the experts' ideas and inspirations from the interviews:

- 1.Recommender (high-usage-quota-based)
- 2.Freshness-Recommender (recent-publication-based)
- 3.Minority-Recommender (low-usage-quota-based)
- 4.Profile-Recommender (based on profile similarities)
- 5.Subject-Recommender (thematic-proximity-based)
- 6.Usage-Similarity-Recommender (clickstream-similarity-based)
- 7.Citation-Recommender (citation-intersection-based)
- 8.Favourites-Recommender (Users'-favourites-lists-based)
- 9.Recommendation of central authors
- 10.School-of-thought Recommender (Scientific social network graph)
- 11.Author-centred usage statistics
- 12.Repository-centred usage statistics
- 13.Subject-centred usage statistics
- 14.User-centred usage statistics
- 15.Reordering links (usage-quota-based)
- 16.Collapsing links in large result sets (usage-quota-based)
- 17.Re-Rendering Result List layout
- 18.Dead link identification
- 19.Users' Quality Statements, i.e. comments (free text)
- 20.Users' Quality Statements (rating)
- 21.Quality Statements (usage-based)
- 22.Ensuring document accessibility (bridging the gaps between different storages)
- 23.Automated Retro-Digitalisation requesting
- 24.Automated translation requesting
- 25.Feed Notifications
- 26.Notifying friends manually (e.g. via e-mail)
- 27.Search Phrase Recommender
- 28.Search Result commenting

Brainstorming Session

A brainstorming was conducted as part of Breakout Group 6 at the OAI6 conference in Geneva (Mittelsdorf & Herb 2009). In contrast to the expert interviews, many proposals were concerned with interface design and data visualisation and presentation. This possibly results from the fact that many participants are situated in libraries. The ideas are grouped and precluded by buzz word labels (Arabic numbers) to heighten readability.

- 1.Authority/Standardisation
 - a.Central Unique Author Identification
 - i.Author
 - ii.Identification/Profile

- iii. Picture
 - iv. Projects
 - v. Competence
 - b. Network of Authors
 - i. Social
 - ii. Professional
 - iii. Expertise
 - iv. Field of interest
- 2. Visualisations/Indexing Dimensions
 - a. Paper's Context
 - b. Visual Social Graph
 - c. Show development of ideas (network graph displaying publication times for a document set)
 - d. Visualisation of publication's geo-location
 - e. Position publication in the „landscape of science“
 - f. Project's social map
 - g. Visualise Data and Connections
 - h. Semantic classification
 - i. Numerical Semantics (Speech Independence)
- 3. Barrier Reduction
 - a. Connect to the world (link between science and application)
 - b. Publication-News-Binding
 - c. Solicit further research
 - i. Need stack
 - ii. Wish list
 - iii. Request notification
 - d. Practicable access to repositories not only via modern PC capabilities and resolution (e.g. mobile phones, hand helds, OLPC, etc.)
- 4. Reception Tracking
 - a. Consistent access statistics
 - b. Re-use tracking
 - c. Enhanced (complex) metrics (for better evaluation)
- 5. Assistance (author and user):
 - a. Automatic update/linking pre-print + new version
 - b. Thumbnail/snapshot creation (first page display)
 - c. Integrate everything (integrate information and processes and results) seamlessly working
 - d. Modification of the document catalogues' structures
- 6. Assistance (author):
 - a. Real-Time assistance in form fill-in
 - b. Automatic Metadata Creation/Lookup
 - c. Reduce Redundant Work (intelligent submission)
 - d. Dynamic publication (version management of production and reception; collaborative production)
 - e. Easy submission process
 - f. Dynamic publication list (exportable)
 - g. Bonus Point System
 - h. Easy feedback from authors to repository
 - i. Repository as workspace
 - j. Repository as research/production environment
 - k. Educational assistance/encouragement for new authors (How-tos)

- l. Automatic/easy classification/positioning of new publication
- m. Automatic citation generation
- 7. Assistance (user):
 - a. Track/pursue other searchers' way through the repository
 - b. User recommendations as part of repository
 - c. Graph/Image extraction from papers
 - d. Dataset extraction
 - e. Assign personalised searchable attributes
 - i. Personal comments
 - ii. Pictures as bookmarks
 - iii. Memory aids
 - iv. Relevance statement
 - f. Transparent result display relevance criteria

Comparison:

Simple numbers in this section indicate items from the expert interview set, while number-letter-combinations belong to the brainstorming set.

Both samples expressed a strong awareness of and interest into the social aspects and laws shaping modern science, calling for social graphs of publications and authors (10. and 2.a.-f.).

Statistics were perceived as an information source to judge quality and coverage (11.-14.) whereas the brainstorming group emphasised inter-repository consistency of statistics as a precondition (4.a.) and possible benefits to evaluation (4.c.).

Many ideas revolved around community sharing, assuming a positive shift in the amount of work required to identify an interesting paper (7.a.; 7.b. and 27.). These trends are probably inspired by the widely perceived impact of user-generated content and community content management. The same is probably true of 7.b., 7.e.i-iv, 28., 19., 20., and 8., but in addition this strong demand for personalisation mechanisms implies that users have the impression of many redundant steps in repository handling (e.g. search for a specific piece of text in a previously read article).

Overall the experts accepted the repository interface as it is in contrast to the brainstorming group. Most technical and bureaucratic proposals came from the latter. Possibly because a majority is employed in the library/knowledge management sector. The experts interviewed on the other hand emphasised that not only the amount of services is important but also the service's success rate. All of them would tolerate recommender systems with an accuracy of 90% or more but would rather not be bothered by the noise produced by an inaccurate service.

There seems to be a demand for complex measures and the unfiltered presentation of complex interrelations instead of simplifications. The persons interviewed no longer believed in the promise of simple numbers describing the world without loss of information.

Future Research

To investigate the desirability order of the collected ideas quantitative methods will be used. The questionnaire will have three logical parts:

- demographic questions for identifying user subcultures and later data re-use
- general "attitude towards different kinds of service" questions. These are filter questions that cause blocks of questions for specific services to be asked.
- specific questions. A participant will have to answer a number of thematic blocks based on his general attitudes.

The questionnaire will be a set of HTML forms. Adaptive testing is easily implemented using dynamically generated HTML pages. Adaptive testing reduces the number of items presented, which helps to prevent participants from giving random answers to questions they are not interested in.

In electronic testing there is also no need to manually transcribe answers from hard-copy forms into the computer, thus eliminating the risk of transcribing errors.

Execution via HTML forms is today the cheapest and most efficient way to conduct a survey targeting a large and international sample. There will be at least a German and an English version.

Conclusion:

The ideas presented in this paper provide especially those persons concerned with usability improvement and the creation of new services with valuable hints to the library or interface perspective. The informative value will greatly increase as the results of the questionnaire survey can be quantitatively interpreted.

The benefit to the other perspectives should not be underrated. Aside from designing specialised tools for evaluators, the data needed to implement added-value services and the data generated by visitors utilising these services can be integrated with established data sources' increasing validity and the amount of variance explained.

Usage data can be used to analyse the validity of bibliometric constructs. New modes of synchronous and asynchronous communication can help libraries and universities –even publishers- to tailor their stock to their clients' demands and even to rectify content or reference structures for example. A stronger awareness of the social aspects of the publishing process can renew peer communication and make peer review more transparent if not completely open. Educational as well as non-academic personnel is not only a beneficiary but as shown in the brainstorming can be a source of major transformations assuming that it is supported by students, academics, and bureaucrats.

Additionally the use of open protocols and standards for object description and data transfer are strictly necessary: Different solutions can aid the innovation process but this should not be an excuse for implementing the same algorithm on a different set of objects without retaining interoperability with other providers. The OAI standards as well as standards as IFABC, ORE, and OpenURL context objects need to be employed and further refined.

References

Appelt, W. (2001), "What groupware functionality do users really use? Analysis of the usage of the BSCW system", in *Parallel and Distributed Processing, 2001, Mantova, 2001*, IEEE, pp. 337-341

Bogner, A. Littig, B. and Menz, W. (2005), *Das Experteninterview: Theorie, Methode, Anwendung*, VS Verlag für Sozialwissenschaften, Wiesbaden.

Bollen, J., Van de Sompel, H. and Rodriguez, M.A. (2008), "Towards usage-based impact metrics: first results from the MESUR project", in *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries, Pittsburgh, 2008*, ACM, New York, pp. 231-240.

Brumback, R.A. (2009), "Impact Factor Wars: Episode V – The Empire Strikes Back", *Journal of Child Neurology*, Vol. 24 No. 3, pp. 260-262.

Davis, P.M. and Price, J.S. (2006), "eJournal interface can influence usage statistics: Implications for libraries, publishers, and Project COUNTER", *Journal of the American Society for Information Science and Technology*, Vol. 57 No. 9, pp. 1243-1248.

Galyani Moghaddam, G. and Moballeggi, M. (2008), "How Do We Measure Use of Scientific Journals? A Note on Research Methodologies.", *Scientometrics*, Vol. 76 No. 1, pp. 125-133.

Garfield, E. (2006), "The History and Meaning of the Journal Impact Factor", *Journal of the American Medical Association*, Vol. 295 No. 1, pp. 90-93.

Ghaphery, J. and Ream, D. (2000), "VCU's My Library: Librarians Love It. ...Users? Well, Maybe", *Information Technology and Libraries*, Vol. 19 No. 4, pp. 186-190.

Harnad, S. (2008), "Validating Research Performance Metrics Against Peer Rankings", *Inter-Research Ethics in Science and Environmental Politics*, Vol. 8, pp. 103-107.

IRS (2007), "IRS: Interoperable Repository Statistics", available at: <http://irs.eprints.org/> (accessed 17 July 2009).

Jensen, P., Rouquier, J.-B. and Croissant, Y. (2009), "Testing bibliometric indicators by their prediction of scientists promotions", *Scientometrics*, Vol. 78 No. 3, pp. 467-479.

Ketchell, D.S. (2000), "Too Many Channels: Making Sense out of Portals and Personalization", *Information Technology and Libraries*, Vol. 19 No. 4, pp. 175-179.

Mabel, C., Villanueva, E.V. and Van Der Weyden, M.B. (2007), "Life and times of the impact factor: retrospective analysis of trends for seven medical journals (1994-2005) and their Editors' views", *Journal of the Royal Society of Medicine*, Vol. 100 No. 3, pp. 142-150.

McWilliams, A., Siegel, D. and Van Fleet, D.D. (2005), "Scholarly Journals as Producers of Knowledge: Theory and Empirical Evidence Based on Data Envelopment Analysis", *Organizational Research Methods*, Vol. 8 No. 2, pp. 185-201.

Mittelsdorf, B. and Herb, U. (2009), "Breakout group 6. Access Data Mining: A new foundation for Added-value services in full text repositories", available at: <http://indico.cern.ch/contributionDisplay.py?contribId=72&confId=48321> (accessed 17 August 2009).

Moed, H.F. (2005a), "Statistical Relationships Between Downloads and Citations at the Level of Individual Documents Within a Single Journal", *Journal of the American Society for Information Science and Technology*, Vol. 56 No. 10, pp. 1088-1097.

Moed, H.F. (2005b), *Citation Analysis In Research Evaluation*, Springer Netherlands, Dordrecht.

MONTANER, M., LÓPEZ, B. and DE LA ROSA, J.L. (2003), “A Taxonomy of Recommender Agents on the Internet”, *Artificial Intelligence Review*, Vol. 19 No. 4, pp. 285-330.

Nicholas, D., Huntington, P., Lievesley, N. and Withey, R. (1999), “Cracking The Code: Web Log Analysis”, *Online & CD-ROM Review*, Vol. 23 No. 5, pp. 263-269.

Nicholas, D., Huntington, P. and Watkinson, A. (2005), “Scholarly journal usage: The results of deep log analysis”, *Journal of Documentation*, Vol. 61 No. 2, pp. 248-280.

Project COUNTER (2008), “COUNTER Codes of Practice”, available at: http://www.projectcounter.org/code_practice.html (accessed 17 July 2009).

Putz, M. (2002), *Wandel der Informationsvermittlung in wissenschaftlichen Bibliotheken*, University of Applied Sciences for Library and Information Management, Eisenstadt, 17 July 2009.

Sikkel, K., Gommer, L. and Van der Veen, J. (2002), “Using Shared Workspaces in Higher Education“, *Innovations in Education and Teaching International*, Vol. 39 No. 1, pp. 26-45.

Shepherd, P. (2007), “Final Report on the Investigation into the Feasibility of Developing and Implementing Journal Usage Factors”, available at: <http://www.uksg.org/sites/uksg.org/files/FinalReportUsageFactorProject.pdf> (accessed 17 July 2009).

Shepherd, P. and Wallace, J.M. (2009a), “PEER: a European project to monitor the effects of widespread open access archiving of journal articles”, *Serials*, Vol. 22 No. 1, pp. 19-23.

Shepherd, P. and Needham, P.A.S. (2009b), “PIRUS Final Report”, available at: http://www.jisc.ac.uk/media/documents/programmes/pals3/pirus_finalreport.pdf (accessed 17 August 2009).

Stassopoulou, A. and Dikaiakos, M.D. (2007), “A Probabilistic Reasoning Approach for Discovering Web Crawler Sessions”, *Lecture Notes in Computer Science*, Vol. 4505 No. 1, pp. 265-272.

Biographical Notes:

Ulrich Herb studied Sociology at Saarland University, Germany. He is member of the electronic publishing group of Saarland University and State Library.

Affiliation: Saarland University and State Library, Saarbrücken, Germany

Eva Kranz is studying Bioinformatics at Saarland University, Germany, where she also has been working as a student assistant for Open-Access-Statistics, since 2008. Ms. Kranz is actively involved in the open source project Collabtive where she is responsible for development, documentation and community management.

Affiliation: Saarland University and State Library, Saarbrücken, Germany

Tobias Leidinger is studying Computer Science at Saarland University, Germany. He is working for several electronic publishing projects at Saarland University and State Library (e.g. OPUS 4 and Open Access Statistics), since 2006.

Affiliation: Saarland University and State Library, Saarbrücken, Germany

Björn Mittelsdorf has been a member of Open-Access-Statistics since 2008. Previously he spent two years at the Institute for Psychology Information, Trier, Germany, where he was involved in digital preservation of primary research data.

Affiliation: Saarland University and State Library, Saarbrücken, Germany