



Journal Publishing in Developing, Transition
and Emerging Countries

Proceedings of the 5th Belgrade International Open Access Conference 2012

Belgrade, Serbia, May 18-19, 2012

Edited by Pero Šipka



Journal Publishing in
Developing, Transition, and Emerging
Countries

**Proceedings of the
5th Belgrade
International Open Access Conference 2012**

Belgrade, Serbia
May 18-19, 2012

Pero Šipka (ed.)

Publisher
Centre for Evaluation in Education and Science
Kneza Miloša 17
11.000 Belgrade
Serbia
tel. +381 11 32 38 506
www.ceon.rs



Copyright information

© 2013 Centre for Evaluation in Education and Science

ISBN 978-86-89475-01-2

ISSN (print): 2334-7880

ISSN (online): 2334-7880

Cover by PeteF

Ordering information

Print-on-demand by Amazon.

Web access and citation information

This entire proceedings can also be viewed on the web at

<http://boac.ceon.rs/public/full/5th-bioac.pdf>. Each paper has a unique identifier (DOI) which can be added to citations to facilitate access.

The DOI # should not replace the full citation. Each e-version contains instruction on how to cite a paper.

Programme Committee Members

Agathe Gebert

GESIS – Social Science Open Access Repository (SSOAR), Köln
(*Germany*)

Bhaskar Mukherjee

Banaras Hindu University, Department of Library and Information
Science
(*India*)

Biljana Kosanović

National Library of Serbia, Belgrade
(*Serbia*)

Eleonora Dagiene

Vilnius Gediminas Technical University
(*Lithuania*)

Giannis Tsakonas

University of Patras, Library & Information Center, Patras
(*Greece*)

Goran Nenadić

University of Manchester, School of Computer Science
(*United Kingdom*)

Irina Kuchma

eIFL: Electronic Information for Libraries
(*The Netherlands*)

Jovan Zubović

Institute of Economic Sciences, Belgrade
(*Serbia*)

Marcin Kozak

University of Information Technology and Management, Faculty of
Economics, Department of Quantitative Methods in Economics, Rzeszow
(*Poland*)

Miloš Radovanović

University of Novi Sad, Faculty of Sciences
(*Serbia*)

Nikos Houssos

National Documentation Centre / NHRF, Athens
(*Greece*)

Pippa Smart

PSP Consulting, Oxford
(*United Kingdom*)

Primož Južnič

University of Ljubljana, Faculty of Arts, Department of Library and
Information Science and Book Studies, Ljubljana
(*Slovenia*)

Siniša Subotić

NGO “Persona”, Banja Luka
(*Bosnia & Herzegovina*)

Stela Filipi

University of Belgrade, University Library
(*Serbia*)

Vasiliy Vlassov

Moscow Medical Academy, Moscow
(*Russia*)

Pero Šipka

Centre for Evaluation in Education and Science, Belgrade
(*Serbia*), The President

Contents

Preface	ix
Full Papers	
Enhanced Journals: A Case Study with General Remarks <i>Andras Holl</i>	1
Progressive Opportunities for Research Journal Publishing <i>Eleonora Dagiene</i>	11
Journal Classifications Based on Citation Data: The Comparison and Suitability of Three Distance Measures <i>Dejan Pajić</i>	25
“Face Validity” Differences between Locally Published and Highest Level International Journals Based on the In-Text Citations Approach <i>Siniša Subotić</i>	37
Open Access Journals in India: A Study Report with Special Reference to Computer Science Journals <i>Bhaskar Mukherjee</i>	49
Maps of Science Based on Keywords of Articles’ Antecedences, Presences, and Consequences: Application of the CEON/CEES Model of Multi-Perspective Description of Articles <i>Tanja Jevremov, Dejan Pajić, Miša Sotirović & Pero Šipka</i>	61
Supporting Open Access Nationwide <i>Jadranka Stojanovski</i>	75
Does the Open Access Business Model Have a Significant Impact on the Citation of Publications? Case Study in the Field of Civil Engineering <i>Teja Koler Povh, Primož Južnič & Goran Turk</i>	89
Computer Science and Information Systems: Publishing an International Open Access Journal in a Developing Country <i>Mirjana Ivanović, Miloš Radovanović, Vladimir Kurbalija & Jovana Vidaković</i>	101

Applicability of the Nano-publication Concept for Fostering Open Access in Developing and Transition Countries <i>Adam Sofronijević & Aleksandra Pavlović</i>	113
Output in WoS vs. Representation in JCR of SEE Nations: Does Mother Thomson Cherish All Her Children Equally <i>Biljana Kosanović & Pero Šipka</i>	125
Short Papers	
Towards Automated Enrichment of Scientific Publications with Complex Meta-data <i>Aleksandar Kovačević, Zora Konjović & Goran Nenadić</i>	139
Origin of Plagiarized Manuscripts Submitted in the Croatian Medical Journal <i>Ksenija Baždarić, Lidija Bilić Zulle & Mladen Petrovečki</i>	145
Challenges for an Emerging New Scientific Journal in Romania <i>Marius Drugaș, Gabriel Roșeanu & Elena Bonchiș</i>	149
Publishing Scientific Journals in the Conditions of a Transition Society <i>Živan Živković & Ivan Mihajlović</i>	153
Democratisation of Knowledge through Open Access Journals and Repositories in Transition Countries. Case study: Singidunum Journal of Applied Sciences and Singipedia <i>Gordana Dobrijević & Svetlana Stanišić</i>	157
Bibliometric Quality of Serbian Journals 2002-2011: More Than Just a Dress for Success <i>Pero Šipka</i>	161

Preface

The 5BIOAC Programme Committee decided to put academic journals into the focus of the 2012 conference. We primarily had in mind journals from transition and developing countries, in line with the earlier defined scope and range of the conference, its organizer, and traditional participants. The idea was that Belgrade is the right place to mix and match different and distant players having their opinion on this important subject.

The idea was also to mix different contents – from full paper presentations, through mad minutes and workshops, to open forum discussions. And indeed, we succeeded in assembling a decent number of scholars from the developed and the developing world, world experts and local enthusiasts, big and small publishers, IT professionals and journal editors, to join a versatile program where everybody was welcome to voice their opinion. This could be sensed during the conference, and we hope it will be repeated in response to the articles published in this volume.

We received altogether 39 submissions, significantly more than we could possibly accommodate during the two days of sessions. The Program Committee, along with a number of external referees, did an outstanding job of reviewing the papers, giving detailed comments and suggestions to the authors. There were 14 full oral presentations and 7 posters reported also within a mad minutes session. Only 11 full papers and 6 poster papers that met high methodological and publishing standards are included in the proceedings. We are grateful to all the authors for their effort.

The Programme Committee is most sincerely grateful to the invited speakers, Stevan Harnad, James Testa and Ana Marušić, for their outstanding contributions which gave a special dimension to the conference. Their inspiring presentations in both .PDF and video formats are available at the 5BIOAC conference website. Mr. Testa, along with David Tempest, provided an additional contribution by exposing – on a special open forum – the general position of LPJs in the policies of their respective publishing houses, Thomson Reuters and Elsevier. The program was enriched with two valuable workshops, one on ScholarOne Manuscripts, held by Ian Potter from Thomson Reuters, and the other on the new CrossRef services, held by Ed Pentz from CrossRef accompanied by his local partners.

Our special thanks are due to Thomson Reuters, Elsevier, and CrossRef, for their generous donation to the conference. Session chairs Miloš Radovanović, Stela Filipi, Agathe Gebert, Irina Kuchma, and Primož Južnič did a great job in keeping the conference on track. Finally, our warm thanks to Saška, Tanja, Cica, Naki, Boža, Mateja, Zoka, Jole, Miša, Nebojša, Nikola Pavlović and, last but foremost, Nikola Stanić, who all accomplished their organizational tasks with full commitment and infectious enthusiasm.

Pero Šipka
CEON/CEES President
5BIOAC Programme Committee Chair

Enhanced Journals: A Case Study with General Remarks

András Holl

holl@konkoly.hu

Library of the Hungarian Academy of Sciences, Konkoly Observatory, Hungarian Academy of Sciences, Arany J. Str. 1, Budapest (Hungary)

Abstract

The Information Bulletin on Variable Stars – a small astronomy journal published in Hungary – was an early Open Access journal. Now it offers several enhanced features to its reader community. Relying on the rather unique publishing environment existing in the field of astronomy, and on software developed locally, this journal is markedly different from other enhanced journals in certain aspects.

We explore the key features of enhanced and common electronic journals: reference linking, database connections, data linking, multi-media content, feedback from the reader community, quality control. We argue that while exploring new avenues of scientific publishing, one should conservatively preserve some traditional values and features.

Some aspects of article disassembly – dealing with items smaller than the usual basic publication unit, the article – are explored as well. Figures, for example, are article components which might be re-used, i.e. used outside the original context.

Keywords: articles; astronomy; contents; data; Electronics journals; journal publishing; linking; Web; information

Introduction

The Information Bulletin on Variable Stars (IBVS) is a small electronic journal published by Konkoly Observatory, Hungarian Academy of Sciences. It covers a narrow research field in astronomy: the study of stellar variability. Originally it was started as a bulletin in 1961, it appeared on the web in 1994, and became a refereed journal around that time. The paper version was discontinued in 2011, and since then IBVS has been an electronic-only journal (<http://www.konkoly.hu/IBVS/IBVS.html>).

Enhanced journals & IBVS features

Features that enhanced publications offer usually include the following: i.) extensive linking of external resources; ii.) availability of research data; iii.) multimedia content; iv.) post-publication material (see the information on the [SURF Foundations EJME project](#)).

Reference linking

Reference linking is not considered as an enhanced feature. Most electronic journals offer linked references – the usual method is via CrossRef. IBVS does have reference links, but relies on a different support infrastructure: the bibliographic identifier standard in astronomy (Bibcode, see [Schmitz et al., 1995](#)) and the NASA/Smithsonian Astrophysics Data System (ADS, see [Kurtz et al., 2000](#)), a free bibliographic database. Bibcode identifiers are created from the source text of the article using technology initially developed at the Centre de Données astronomiques de Strasbourg (CDS), and tailored to the needs of IBVS. The rationale behind technologies used at IBVS is discussed in more detail by [Holl \(2012a\)](#). Reference linking is an essential feature. The IBVS method is probably not adaptable generally – it is beneficial for an express journal, but might not be suitable elsewhere.

Object linking

IBVS provides links to several astronomical object databases and electronically available catalogs, including CDS Simbad ([Genova, 2007](#)), the NASA Extragalactic Database and the General Catalogue of Variable Stars (Moscow). In this case the objects need to be tagged by the authors. Tagged objects are automatically reported to CDS and ADS (both of these astronomical databases link back to IBVS). All object links are generated from standardized object names. Object linking is a desirable feature for electronic journals – but it assumes the existence of object databases and lexical resources, preferably freely available.

Research data

Creating and maintaining repositories of data is a task much more difficult than in the case of publication. Data is best archived by specialists of the given field. IBVS decided to store and make available the data used

for the articles. Each data file has its own meta-data, and a unique identifier (IBVS does not assign DOIs yet, neither to the articles, nor to the data – we cannot afford even the minimal cost and manpower increase it involves). Small volumes of research data are usually present in the research articles – either in the text, or in tables or figures. These data are prepared for human consumption. IBVS tries to put a machine readable data file behind tables and figures, wherever it is applicable.

DOIs offered by DataCite could be used by other enhanced journals, if the data is made available in a repository. If a journal chooses to store research data related to the articles, some database functions need to be implemented, and questions about the long-term preservation of data need to be addressed.

Most of the time research data for IBVS is simple, and could be easily stored in plain text files. But more complex data structures present challenges. Journals might understandably defer from the task of archiving such data. Packaging would most probably be part of the solution, and external, dedicated data archives could be preferable for complex data.

Data linking

The Wide Field Plate Database (WFPDB, Sofia) contains meta-information on astronomical photographic plates stored at several plate archives around the world. For certain plates low resolution preview images are presented, and occasionally the full digitized plate is available too. IBVS offers links to WFPDB entries, in the cases where we could identify the serial numbers of plates used for a given article ([Tsvetkov et al., 1997](#)).

There are still not many data repositories offering free access. In regards to data the use of standard, unique data identifiers (DOI or other), and standard data formats should be emphasized.

Multimedia content

Multimedia is not widely used in astronomy publications. However, IBVS does have some animated GIF figures. Enhanced journals should make sure they use standard file formats to ensure platform-independence, easy access and the possibility of preservation. Wide availability of software

for the given format is also of great importance. In this particular case - animated GIFs can be viewed with any web browser.

Post-publication material

The only type of post-publication material IBVS offers are the errata. All the substantial errors noticed after publication result in an erratum, attached to the original article. It is possible for an enhanced journal to provide a forum for commentaries, but our view is that such comments have only a limited audience – those few researchers who visit the journal website regularly. We think that most readers access an article following a link from the references section of another article, or from a bibliographic database, and they often choose the PDF, thus commentaries mostly present in the HTML are not visible for them. The real forum for comments is the whole scientific literature: if an article needs serious commentaries or reflections, they should appear in other articles. On the other hand, while it is important to facilitate scientific discussions, their content is different from that of literature. Now it is also possible to preserve and make available the content of mailing lists. Our view is that what matters should appear in the mainstream literature, and some exchanges of opinions – while these play an important scholarly role – should be left to oblivion in the vast majority of cases. References to "private communication" or mailing lists should be the exception, not the rule.

Use of a third-party data visualization tool

IBVS articles often contain long lists of objects with coordinates and other characteristics which could only be visualized using a spatial information system. As such a system exists in astronomy – the CDS Aladin Sky Map ([Boch, 2011](#)) – IBVS relies on it. Charts and star lists are passed to Aladin to enable the reader to interact with the data. In our view, there is a need for such GRID-like tools in other areas as well. IBVS tables are transferred to Aladin in a standard data format used in astronomy: the VOTable ([Ochsenbein et al., 2004](#)).

Meta-data

Providing meta-data for the readers and for the bibliographic services is a must for every electronic (or indeed, printed) journal. IBVS presently

supplies meta-data for each issue in the tagged format of ADS. We are also ready to provide DublinCore meta-data through OAI-PMH.

Abstracts

Abstracts form a special part of meta-data. At IBVS, abstracts were introduced for the sake of ADS (and other bibliographic services). These abstracts are the usual, textual summaries of the paper. Some journals recently introduced visual abstracts – figures, which catch the eye of the reader and might tell more than a paragraph of text. Both kinds of abstracts are for humans. One might ask whether journals could provide abstracts for machines – abstracts which contain the summary, the most important conclusions of the article formatted in a formal language. IBVS regularly publishes short discovery notes on newly found variables. Such information should be digestible for machines too. In a semantic web scenario reports on local cricket or football matches should contain a computer readable abstract on the scores. Scientific journal articles should be linked to the data web as well, distilling the most important findings to machine readable format.

Quality control

Alternatives for the current peer review system are being considered. Reader voting – one of the proposed alternatives – does not provide means for improvement of the paper. Open review might have its advantages, but we think it does not relieve the editors of certain tasks: the competence and the independence of the referees must be established anyway. IBVS employs the traditional refereeing system (the single-blind version). However, IBVS benefits from being an OA journal, with a rapid publication cycle: errors caught by the readers are corrected in the Errata attached to the papers.

Enhanced journals present further problems from the aspect of peer review. How could referees examine all auxiliary materials? Providing the necessary tools should be the lesser problem – some tools at least should be provided for the readers anyway. Other tools – which help the systematic checkout of the material, and not present in the readers' toolkit, might be more difficult to provide. IBVS has some tools provided by CDS for object name checking and some other tools developed in-house for dis-

covery note validation and time series manipulation, but referees cannot use these. Peer review takes up precious time of the referee, and the whole process delays the publication of the article. Would the referee take the time to go through all the – possibly extensive – auxiliary material? Maybe this is the point where readers could be employed, after publication, as a second line of referees. At IBVS the referee usually sees the PDF of the article, while only editors scrutinize all auxiliary material.

Long term archiving

Enhanced features of the electronic journals depend on many external factors. The longevity of database links or the sustainability of external software are hard to guarantee for the long term. We think that preserving the enhanced features of the journals is a great challenge. Long term preservation is best served by following a proactive approach. Articles in enhanced journals contain informational elements (new data, new findings) and presentational elements (like multimedia), and the two are not easy to separate. Nevertheless, the former is more important, and easier to preserve than the latter. Traditional journals are easier to preserve than enhanced ones, and maybe it is enough to preserve the core components (and the data).

IBVS articles still have – and will have – a self-consistent core, the PDF article, which is preservable in the long run. PDF was chosen because it is analogous to the printed article. We plan to use PDF/A, and we are seeking for a suitable dark archive within the astronomical community – the obvious choice being the ADS. For the data, CDS could be a suitable place for archiving. Links in the HTML version, or the Aladin visualization of IBVS data, are not necessarily preservable.

Article disassembly, compound objects, aggregation

Scientific journal articles can be disassembled to smaller components – semantic building blocks – which might be reassembled to provide a different view of the article, or used independently (see [Holl, 2012b](#), in press). Certain components of IBVS articles could be accessed and used independently of the whole article, e.g. figures, tables and meta-data. All figures and some tables (data files are regarded as auxiliary tables) have their own unique identifiers through which they are individually accessi-

ble. Each figure and some tables have their own meta-data, including object names. We prefer to have the data available in tabular form for the relevant figures.

Figures are those article components which could be re-used, used out of the context of the original article. With proper attribution and reference to the origin, figures should be reproducible. While using CC licenses for scientific journal articles does not seem appropriate, it should be considered for figures.

Relations between article components and auxiliary material could be mapped using OAI-ORE. IBVS does not use this technology yet. Another possibility of creating compound objects from the article and its auxiliary material is to employ some kind of packaging technique. These are discussed in Enhanced Publications ([Vernooy-Gerritsen, 2009](#)).

Advanced search

There are two features that make IBVS search unusual. One of them is the ability to search and retrieve data files and figures. The other is that we have implemented a system which enables searching for object and author name aliases. Aliases could be fetched from external dictionaries.

Usefulness and cost of the enhanced features

Though we have not investigated the reception of the enhanced features thoroughly, it is possible to compare the impact of the journal with other journals in the same field. Such comparisons are favorable for IBVS, and access statistics showed an increase in downloads with the introduction of the HTML version. However, it is hard to disentangle the effect of the enhanced features. Enhanced journals are still in their infancy, and probably only time will tell which features are really important. In our case the enhanced features were developed by only a few people: the technical editor and some students, using only a fraction of their time. To some extent the development was an exercise which enabled the developers to learn about information technology, services and standards in the field of astronomy. Larger journals and future enhanced journal projects might follow a more rigorous development process.

Plans for improvements

There are several improvements we intend to make for IBVS. OAI-PMH meta-data distribution would help bibliographic databases, while RSS feed would be useful for readers. We plan to implement Open Researcher and Contributor ID-s as soon as possible. The advanced search tool would need a name preprocessing filter, and we plan to create a web submission tool for short notes.

Conclusions

Features offered by IBVS rely on the special environment of astronomy: field-specific standards & file formats are used and open databases provide linking opportunities.

Enhancements of scientific journals are presently offering “extras”, which are interesting, but mostly experimental features. A number of questions should be answered before these can become main-line accessories of the articles. The other difficulty enhanced journals face might be the lack of free resources: databases in the given field or data repositories. In this regard astronomy is in a very fortunate position. Experimenting with enhanced journals is nevertheless important.

References

- Boch, T., Oberto, A., Fernique, P., & Bonnarel, F. (2011). Aladin: An Open Source All-Sky Browser. In: *ASP Conference Proceedings*, 442-683. Retrieved from <http://www.aspbooks.org/publications/442/683.pdf>
- Genova, F. (2007). The Many Faces of SIMBAD. *ASP Conference Series*, 376(145). Retrieved from http://www.aspbooks.org/a/volumes/article_details/?paper_id=5271
- Holl, A. (2012). Information Bulletin on Variable Stars: Rich Content and Novel Services for an Enhanced Publication. *D-Lib Magazine*, 18(5-6). Retrieved from <http://dlib.org/dlib/may12/holl/05holl.print.html>
- Holl, A. (2012). Article disassembly: New ways to handle information in publications. *Serdica Journal of Computing* (accepted for publication). Retrieved from <http://sci-gems.math.bas.bg/jspui/bitstream/10525/1780/1/sjc-vol6-num1-2012-p59-p66.pdf>

- Kurtz, M.J., Eichhorn, G., Accomazzi, A., et al. (2000). The NASA Astrophysics Data System: Overview. *Astronomy and Astrophysics Supplement*, 143(41), [doi:10.1051/aas:2000170](https://doi.org/10.1051/aas:2000170).
- Ochsenbein, F., Williams, R., Davenhall, C., & et al., (2004). VOTable: Tabular Data for the Virtual Observatory. In: *ESO Astrophysics Symposia, Toward an International Virtual Observatory*, 118. [doi:10.1007/10857598_18](https://doi.org/10.1007/10857598_18).
- Schmitz, M., Helou, G., Dubois, P., et al. (1995). NED and SIMBAD Conventions for Bibliographic Reference Coding. In: D. Egret & M.A. Albrecht (Ed.), *Information & On-Line Data in Astronomy*. Kluwer. Retrieved from <http://cdsweb.u-strasbg.fr/simbad/refcode.pdf>
- SURF Foundation, Enhanced Journals...Made Easy. Retrieved from <http://www.surfoundation.nl/en/projecten/Pages/EJME.aspx>.
- Tsvetkov, M.K., Stavrev, K.Y., Tsvetkova, K.P., et al. (1997). The Wide-Field Plate Database: Present Status and Future Development. *Baltic Astronomy*, 6, 271. Retrieved from http://articles.adsabs.harvard.edu/cgi-bin/nph-iarticle_query?1997BaltA...6..271T&data_type=PDF_HIGH&whole_page=YES&type=PRINTER&filetype=.pdf
- Vernooy-Gerritsen, M., (ed.) (2009). *Enhanced Publications*. Amsterdam University Press. [doi:10.5117/9789089641885](https://doi.org/10.5117/9789089641885).

Progressive Opportunities for Research Journal Publishing

Eleonora Dagiėnė

eleonora.dagiene@vgtu.lt

Vilnius Gediminas Technical University, Sauletekio al. 11, LT-10223 Vilnius
(Lithuania)

Abstract

Research policy in new EU member states stands at the crossroads of the organization of science and research activities. A great number of research journals, predominantly in local languages, are relics from the past. Twenty years ago science was very much a local undertaking and publishing in international journals was rare and limited, as well as generally not supported by the government. This was the main reason why so many journals in local languages were established. It is understandable why editors would like to preserve them, but it is a question of national research policy to assess whether investing in local research journals is a worthwhile use of resources, or if it might be better to promote and support publishing in distinguished international journals. It is not only Lithuania but also the other post-Soviet countries which gained independence more than twenty years ago that find themselves at this crossroad. Rapidly evolving technologies present new challenges for academic publishers, being an integral part of modern scholarly communication. Unfortunately, the majority of academic publishers and journal editors remain unfamiliar with and lack interest in the rapidly advancing e-publishing technologies and innovations, and thus lag behind the global trends. Aiming to comply with international standards, they have to juggle a number of managerial challenges, including the use of limited funds for the acquisition of efficient e-publishing solutions. In 2008, Vilnius Gediminas Technical University (VGTU) Press initiated a project aiming to maximise journal publishing while keeping costs to a minimum, i.e. implementing the globally recognised CrossRef innovative and e-publishing systems (peer-review, production, online publishing) to further promote journal circulation. As a result, at the end of 2011, DOI, CrossCheck, Cited-by and CrossMark of CrossRef, as well as [Open Journal Systems](#) of Public Knowledge Project, were implemented.

Keywords: journal publishing; languages; online publishing; scholarly communication; Open Access; communications; e-publishing innovations

Introduction

Publishing of research journals is a key factor indicated in different models of scholarly communication. However, distinct forms have been used depending on the period. For instance, the Garvey-Griffith Model, which was introduced in the early 80s, perceived the research publication as a printed entity ([Garvey, 1979](#); [Garvey & Griffith, 1972](#)) with “real” connections. However, the 2020 model presented by Julie M. Hurd in the late 90s ([Hurd, 2000](#)) already employed digital technology terms such as: e-journal site, pre-print server, digital library, etc. Later on, Bo-Christer Björk ([2005](#)) states: “one interesting aspect of the scientific communication process is that it is a global and interconnected information system”. Thus, modern technologies have changed the means of scholarly communication fundamentally, and it has recently been suggested that the research publication be divided into nanopublications. “Data citation and the derivation of semantic constructs directly from datasets have now both found their place in scientific communication. The social challenge facing us is to maintain the value of traditional narrative publications and their relationship to the datasets they report upon while at the same time developing appropriate metrics for citation of data and data constructs.” ([Mons, et al. 2011](#)).

Scholarly communication is rapidly moving to the Internet, so innovative publishers are struggling to keep up and they are using all available modern technologies. It has been about two decades since highly ranked publishers have started providing articles with DOIs, using electronic peer-review systems and publishing journals online.

Currently, journal platforms have started implementing innovative scientometrics tools. In 2009, PLoS became the first publisher to add comprehensive usage and reach data to every published article, so that the entire academic community could form their own assessments of their value ([PlosOne, 2012](#)). The Article Level Metrics (ALM) program by PLoS provides performance data for every article, including the following data from the sources noted: *Article usage* – page views, PDF downloads (PLoS and PubMedCentral) and XML (PLoS); *Citations* – CrossRef, PubMed Central, Scopus, Thomson Reuters Web of Science; *PLoS reader evaluation* – Comments, Notes, Star rating; *Blog coverage* – Research

Blogging, Nature Blogs; *Social tools* – CiteULike, Connotea, Mendeley and Facebook. Last July, SAGE Open was launched. This portal also has an option for readers and the academic community at large to continue the peer review process after online publication ([SAGE Open, 2011](#)).

Publishers of research journals understand perfectly well that a journal is regarded to be good if it contains articles authored by well-known scholars. Subsequently, the opinion of authors needs to be heard when planning to engage in improvements regarding journal publishing. In the international Survey of Senior Researchers ([CIBER, 2005](#)), undertaken by the Publishers Association and the International Association of STM Publishers in 2005, responses to the question “*Choosing where to publish*” reveal that the prestige of the outlet is of central importance, as indicated directly by the journal’s reputation or, by proxy, by its impact factor. Practical issues – such as the nature of the readership for that journal and speed of publication, are also highly rated. While choosing a journal for their manuscript, respondents regard “*Online Manuscript Submission*” and “*Print and Electronic Versions*” as more important than various “*Permissions*”, such as: to post post-print version of their articles, to post pre-print versions or to retain a copy.

The journals using only traditional publishing methods and abstaining from modern publishing technologies do not have the possibility of getting into the flow of international research information and thus do not satisfy the high requirements set for research publications which are related to the quality of information transmission.

Research Journal Publishing in Lithuania

Research policy in new EU member states stands at the crossroads of the organization of science and research activities. A great number of research journals, predominantly in local language, are relics from the past. Twenty years ago science was very much a local issue and publishing in international journals was rare, limited and also generally not supported by the government. This was the main reason for establishing a large number of journals in local languages. It is understandable why editors would like to preserve them, but it is a question of national research policy to assess whether investing in local research journals is a worthwhile

use of resources, or if it might be better to promote and support publishing in distinguished international journals. Currently, issues regarding funding for publishing scientific journals and the way forward are being actively discussed and debated in Lithuania.

Some years ago, Lithuanian science policy-makers required national universities to strive for internationalism, thus centring the entire performance assessment of higher education institutions on international cooperation. Considering these requirements, national universities started strengthening their foreign relations departments. Yet, they continued to fail in utilising opportunities offered by e-publishing – the key factor of contemporary scholarly communication. Since e-publications (textbooks, monographs, etc.) became recognised as equal to printed scholarly outputs in Lithuania as late as in 2010, advancement has been seriously delayed. Nevertheless, academic publishing remains the key channel for scholarly communication, offering researchers opportunities to become known, appreciated and recognised.

Although Lithuania is a small European country with approx. 3 million residents, the number of research journals continues increasing just as in other world countries. In addition, previously printed journals have also become available online in a very simple way – websites with limited and necessary information only. In 2010, the research project “Online availability of Lithuanian research journals” revealed that approx. 80% of Lithuanian research journals were freely accessible online. However, due to primitive e-publishing solutions, the journal websites do not meet scholarly communication requirements for searchability, discoverability and author-researcher visibility. Towards the end of 2010, only 157 out of 219 Lithuanian research journals could be found on the Internet ([Dagiene, 2010](#)), while the remaining ones had a print version only. The current data shows that now, in 2012, 210 research journals have online versions and can be accessed on the Internet. However, the situation with research journal publishing is not undergoing change to improve publishing quality. Rather, some journals recently publicised on the Internet are completely new Open Access journals that are published online only. [Figure 1](#) shows that in 2012 the majority of journals that were searchable on the Internet were in the field of Humanities. A more in-depth analysis showed

that those were mostly simple websites rather than true e-publishing systems (portals).

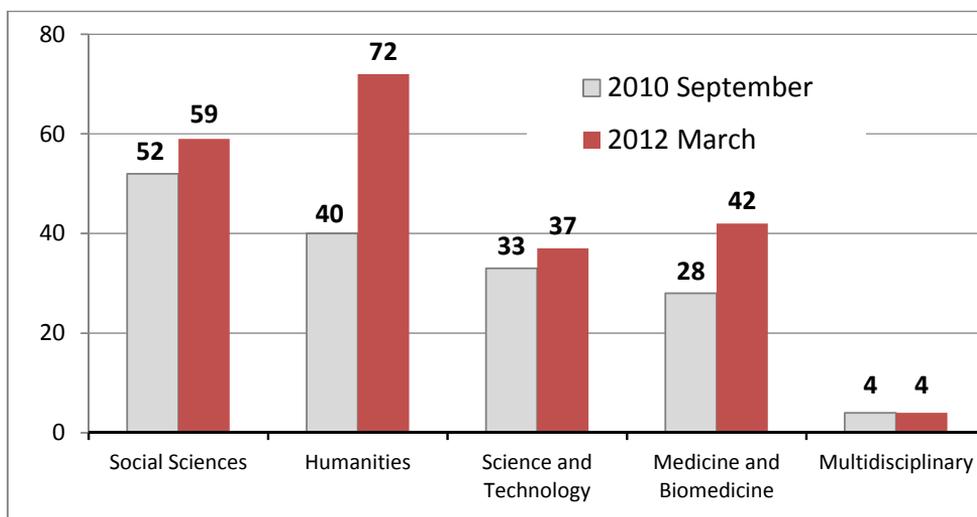


Figure 1. Distribution of Lithuanian journals, which are searchable on the Internet, according to the subject category of the journal

The Lithuanian research journal market is, in fact, oversaturated, which is evident by the large number of journals for such a small country. It could be presumed that this situation is a result of the minor requirements for researchers in Social Sciences and Humanities, who are able to earn scientific titles, be promoted, etc. without publishing internationally. Thus, the total number of researchers is increasing, which gives rise to an increasing number of domestic journals. The articles of local researchers are usually published by journals in the Lithuanian language and the manuscripts are reviewed by Lithuanian reviewers. Therefore, the majority of journals are not visible internationally and do not satisfy the current requirements of the local policy makers nor international publishing standards.

It should be noted that Lithuanian publishers of research journals continue operating on meagre budgets. Traditionally, printed research journals have experienced a lack of manuscripts and budget cuts on the basis of

non-conformity to high internationality standards, thus threatening the survival of such publishing houses. However, the need to improve journal publishing did not lead to an increase in the budgets; therefore a careful consideration of all possible outcomes in relation to the choice of innovative solutions was needed.

The majority of Lithuanian research journals are not published by experienced professional publishers as in the UK, but by academics who are researchers in different fields of sciences and who are responsible not only for content managing, but also for language polishing, typesetting, copy-editing, printing, dissemination, online publication, etc., which is naturally done by publishers in the UK. This is the reason why journals published in the UK are ranked so high around the world. Obviously, Lithuanian journals publicised by researchers are not properly presented on the Internet, as e-publishing requires professional knowledge. Currently, there are no commercial publishers of research journals in Lithuania and a rather large number of editorial boards of Lithuanian research journals approach well-known publishers aiming to expand the prestige of their journals. Some are already published by Springer, Taylor & Francis and Emerald. However, the majority of Lithuanian journal publishers are still in search of the most effective solution.

Pilot project on journal publishing improvement at Vilnius Gediminas Technical University (VGTU) Press 2008-2012

In the beginning of 2008, taking into account the changes taking place in international scholarly communication, it was decided to implement a pilot VGTU journal publishing improvement project which aimed to improve journal publishing while keeping the university costs to a minimum. A current analysis of journal publishing trends opened up opportunities for VGTU, such as implementing the globally recognised CrossRef and establishing innovative and e-publishing systems (peer-review, production, online publishing) to further promote journal circulation. To ensure an even distribution of funds and human resources, the development of journal publishing was divided into gradually implemented stages.

At all times, journal publishing at VGTU strives to correspond to the strict requirements for research journals, such as an international editorial board, peer-reviewing, and indexing in prestigious databases. The journal websites were launched in 1998. The VGTU journals have attained a high level of regard with 8 of the 19 titles being indexed in the Web of Science and 13 in Scopus. The VGTU journals still maintain a respectable level of impact according to the leading metrics (IF, SJR or SNIP).

Aiming to efficiently improve journal publishing with limited financial resources, the project engaged in the continuous monitoring and analysis of various papers on the topic of academic publishing experiences in leading developed countries. In 2008, the Research Information Network released the report “Activities, costs and funding flows in the scholarly communications system in the UK” ([Activities, 2008](#)), which provides an analysis of all the costs involved in the process of publishing, distributing and accessing journal articles, both globally and in the UK.

In The STM Report, “An overview of scientific and scholarly journal publishing” ([Ware & Mabe, 2009](#)), authors Mark Ware and Michael Mabe offered insights regarding the possible changes in the subsequent 3-5 years. Reading them now, in 2012, one can see that they were completely right. Yet another report – “E-journals: their use, value and impact – final report” ([E-journals, 2011](#)), by the Research Information Network, which is also especially relevant to research journals, helps to increase the understanding of the importance of e-journal publishing and its forms for scholars (both authors and readers) as direct consumers of research journals.

Digital Object Identifiers (DOIs)

The first stage (2008) – implementing Digital Object Identifiers (DOIs). For this purpose, Vilnius Gediminas Technical University became the first CrossRef member in Lithuania. As soon as DOI metadata registration in the CrossRef system was launched, VGTU journal websites had to be reorganised. DOI implementation and the reorganisation of the websites did not require grand investments, and it is very important for journals that are published online ([Meyer, C. A. 2008](#))

CrossCheck for plagiarism prevention

The second stage (2009) – CrossCheck is used by all editors of VGTU journals. Over the period of three years since the launch of this service, a number of possible plagiarism cases were identified, thus preventing more serious problems. The implementation of the service required a relatively small investment. Plagiarism is still a hot topic ([Science publishing, 2012](#)).

Partnering with a well-known academic publisher

The third stage (2010). As experience shows, well-known commercial systems demand for greater investments; thus, a publisher with a limited budget rarely has an opportunity to implement such solutions. To expand the prestige of journals and implement prominent online systems while keeping the same publishing costs, VGTU started looking for a publishing partner. Currently, 12 VGTU journals are published in partnership with Taylor & Francis, who use the ScholarOne Manuscript – a well-known online peer-review system. VGTU journals also use CATS, an electronic production system, and are available electronically on Taylor & Francis' publishing platform. Both CATS and the new online platform were made possible by substantial investments on the part of Taylor & Francis.

Implementing Open Journal Systems (OJS)

The fourth stage of the project (2011) – implementation of a publishing platform based on [Open Journal Systems \(OJS\)](#) for five Open Access journals, published by VGTU Press. Using OJS is described in an article ([Edgar & Willinsky, 2010](#)) which explores the possibilities and advantages of OJS. VGTU journals were provided with a unified style and separate OJS Managing Systems that support two languages, i.e. Lithuanian and English. Earlier journal volumes were loaded onto the system with DOIs. In addition, all journals were provided with a Cited-by widget which was written by a VGTU Press programmer, and one journal also has a Cross-Mark option ([Meyer, 2011](#)). A small budget was allocated for this stage. OJS can be easily adapted to meet innovative journal publishing needs without requiring large investments, and thus subsequently help presses such as VGTU attain the status of a contemporary publisher. All journals have unique URL addresses that have been registered with Google Ana-

lytics in the beginning of 2008. This way, it was possible to track the subsequent implementation of OJS and the registration of journals in the DOAJ database. Not only did the number of visitors increase, but also the list of countries expanded.

Stage four finalised the project that was initiated in 2008. As primary results suggest, the key objectives were reached: publishing of VGTU journals was improved and there has been an increase in prestige for authors from around the world and other Lithuanian publishers. Minimum investments were required to improve journal circulation and visibility in the global academic community. Nevertheless, as the process is still at a very early stage, final conclusions are still premature.

Implementing Open Journal Systems (OJS) in Lithuania

Up to 2010, only three journals in Lithuania used OJS as a publishing system and accepted manuscripts online. The EIFL grant ([Open Access, 2012](#)) served as a great stimulus for the project “Promoting Open Access through Implementation of [Open Journal Systems](#) in Lithuanian Research Journals”. The detailed report of the project is freely accessible on the Internet. During the project, more than 130 members of the academic community engaged in academic publishing attended seminars organised by the Association of Lithuanian Serials to learn about opportunities offered by OJS and CrossRef innovations. Following the seminars, 22 OJS installations were implemented in seven research institutions in Lithuania at the end of 2011.

In 2011, Kaunas University of Technology (KTU) implemented [KTU Open Journal Systems](#), which accommodates all 12 journals published by the University. Furthermore, only the most up-to-date volumes (since 2011) are provided in the OJS. The current situation – with volumes of the same journals spread across three different archives (older websites), is undesirable as it impedes indexing, promoting and searching, especially when a reader needs to find publications from different years or all articles from the same author in one journal, etc. The Director of the KTU Publishing House has relayed that the majority of activities were implemented by full-time employees of the KTU Publishing House and the University, using additional funding from a variety of sources.

As part of the Lituaniatika Project, the research journal *Baltistica* published by Vilnius University received a grant of the Research Council of Lithuania amounting to EUR 10,434. The journal was supplied with OJS Managing Systems that provides user environments in as many as six languages: English, Lithuanian, German, Latvian, French and Russian. However, articles and abstracts are only available in the original language of the article. It should be noted that the entire archive from the very first volume of 1965 has been uploaded onto the system. The Managing Editor, Vytautas Rinkevičius, was the project manager assisted by a full-time IT specialist working at VU Faculty of Philology and two students who were in charge of scanning the volumes and uploading them onto the system. The journal has no designated administrator; all reviewing and publishing processes are curated by the managing editor who initiated the implementation of OJS.

The perspective offered by Ben Hunter ([Hunter, 2010](#)) regarding the OJS Management system implementation is relevant when analysing the experience of OJS implementation and the improvement of e-publishing in Lithuania: the new system had to be customised, back issues had to be imported, and the editorial board had to be trained on the new system; all of which is reasonable and proven by our experience.

Conclusions

Contemporary e-publishing of research journals covers not only online publishing, but also a number of other aspects that are directly related to publications, authors, institutions, their assessment and representation. Consequently, the improvement of academic publishing is a large undertaking that demands a deep understanding of modern technologies, their opportunities, and importance. If such improvements are entrusted to people with insufficient skills, it would be best if they were to consult with specialists.

In aiming to be visible, easily found and acknowledged by the worldwide academic community, all research journals should implement well-tailored online platforms, but also register article DOIs, use CrossCheck, Cited by Linking and other relevant services and technologies designed for journal publishing.

To maximise results, the editorial boards of research journals should rely on assistance from professional and experienced publishers.

Acknowledgements

The author would like to thank Prof. Arunas Augustinaitis for generously sharing his valuable thoughts, one of the unnamed reviewers and Victoria Babbit for useful remarks and suggestions that helped to improve this publication. Any errors are the author's own responsibility.

Competing interests

Eleonora Dagienė is the Director of Vilnius Gediminas Technical University Press and the Chair of the Council of the Association of Lithuanian Serials.

The Association of Lithuanian Serials has received a grant from EIFL which helped introduce innovative ideas in Lithuania and abroad during 2011 and 2012.

References

- Activities, costs and funding flows in the scholarly communications system in the UK (2008). Retrieved from <http://www.rin.ac.uk/our-work/communicating-and-disseminating-research/activities-costs-and-funding-flows-scholarly-commu>, 2012 Mar 20.
- Björk, B. (2005). A lifecycle model of the scientific communication process. *Learned Publishing*, 18(3), 165-176. doi:10.1087/0953151054636129.
- Dagiene, E. (2010). Lithuanian Research Journals Moving to Open Access. In: *Knowledge Exchange Workshop at COASP, 21st September 2011*. Retrieved from <http://www.knowledge-exchange.info/Default.aspx?ID=479>, 2011 Nov 30.
- E-journals: Their use, value and impact - final report. A Research Information Network report (2011). Retrieved from <http://www.rin.ac.uk/our-work/communicating-and-disseminating-research/e-journals-their-use-value-and-impact>, 2012 Jan 25.
- Edgar, B.D., & Willinsky, J. (2010). A Survey of the Scholarly Journals Using Open Journal Systems. *Scholarly and Research Communication*, 1(2), 22. Retrieved from <http://dlib.org/dlib/may12/holl/05holl.print.html>

- Garvey, W.D. (1979). *Communication: The essence of science*. Elmsford, NY: Pergamon Press.
- Garvey, W.D., & Griffith, B.C. (1972). Communication and information processing within scientific disciplines: Empirical findings for psychology. *Information Storage and Retrieval*, 8, 123-126. Retrieved from <http://infocuib.laborales.unam.mx/~mt12s01j/archivos/data/2/42.pdf>
- Hunter, B. (2010). Moving Open Access to Open Source: Transitioning an Open-Access Journal into the Open Journal Systems Journal Management System. *Technical Services Quarterly*, 28(1), 31-40. doi: [10.1080/07317131.2010.500972](https://doi.org/10.1080/07317131.2010.500972).
- Hurd, J.M. (2000). The transformation of scientific communication: A model for 2020. *Journal of the American Society for Information Science*, 51(14), 1279-1283. Retrieved from <http://www.ou.edu/ap/lis5703/sessions/hurd.pdf>
- Meyer, C.A. (2011). Distinguishing published scholarly content with CrossMark. *Learned Publishing*, 24(2), 87-93. doi: [10.1087/20110202](https://doi.org/10.1087/20110202).
- Meyer, C.A. (2008). Reference Accuracy: Best Practices for Making the Links. *Journal of Electronic Publishing*, 11(2), doi: [10.3998/3336451.0011.206](https://doi.org/10.3998/3336451.0011.206).
- Mons, B., et al., (2011). The Value of Data. *Nature Genetics*, 43, 281-283. doi: [10.1038/ng0411-281](https://doi.org/10.1038/ng0411-281).
- New Journal Publishing Models: An International Survey of Senior Researchers, CIBER author survey. A CIBER report for the Publishers Association and the International Association of STM Publishers (2005). Retrieved from http://www.homepages.ucl.ac.uk/~uczciro/pa_stm_final_report.pdf, 2008 Jan 7.
- Open Access publishing in Lithuania and Serbia: Case studies (2012). Retrieved from <http://www.eifl.net/news/open-access-publishing-lithuania-and-serbia-c>, 2012 Mar 12.
- Open Journal Systems. Retrieved from <http://pkp.sfu.ca/?q=ojs>, 2011 May 4.
- PlosOne. Retrieved from <http://blogs.plos.org/plos/2012/02/new-data-sources-added-to-article-level-metrics>, 2012 Feb 20.
- SAGE Open. Retrieved from <http://www.sagepub.com/journals/Journal202037?siteId=sage-us&prodTypes=any&q=public+usage+metrics&fs=1>, 2011 Sep 25.
- Science publishing: How to stop plagiarism (2012, January 5). *Nature*, 481, 21-23 [Comment]. doi: [10.1038/481021a](https://doi.org/10.1038/481021a).
- Ware, M., & Mabe, M. (2009). *The STM report: An overview of scientific and scholarly journals publishing*. Oxford, UK: International Association of

Scientific, Technical and Medical Publishers. Retrieved from http://www.stm-assoc.org/2009_10_13_MWC_STM_Report.pdf, 2012 Apr 5.

Journal Classifications Based on Citation Data: The Comparison and Suitability of Three Distance Measures

Dejan Pajić

dpajic@ff.uns.ac.rs

University of Novi Sad, Faculty of Philosophy, Dr Zorana Đinđića 2, 21000 Novi Sad (Serbia)

Abstract

In the study presented in this paper, citation data from the Serbian Citation Index were used to calculate three types of proximity measures among Serbian scientific journals in the fields of social sciences and humanities. The measures were based on the frequency of inter citations among journals, journal co-citation counts, and bibliographic journal coupling. The clustering solutions derived from different distance matrices were compared and validated against the current subject classification of national journals. All three solutions were generally compatible with the list of major disciplines suggested by the Serbian Ministry of Education and Science, but the results indicated the need for a more precise subject classification. The most accurate journal classification has been achieved by using the bibliographic journal coupling method. Bibliographic journal coupling has primarily produced clusters of thematically similar journals, while inter citation and co-citation clusters also revealed some types of relationships not necessarily determined by subject similarity. The visualization of proximity data provided additional insight into the relations among disciplines and the status of several multidisciplinary journals. In general, the presented results indicate that the citation data are suitable not solely for journal classification tasks, but also as a mean of scientific domain analysis at the journal level.

Keywords: journal mapping; journal clusters; co-citation analysis; bibliographic coupling

Introduction

Monitoring and improving the quality of academic journals are among the most important issues of scientific policies, particularly those of small and developing countries. The commonly used indicators of journal quality

(i.e. *Impact Factor* and *Eigenfactor*) are usually based on the data provided by Thomson Scientific Journal Citation Reports, and are thus not suitable for the evaluation of journals from “peripheral” countries. This was the main reason why the project on the evaluation of Serbian scientific journals was launched in 2004. Since then, the Center for Evaluation in Education and Science, with the support of the Serbian Ministry of Education and Science, has annually published the *Journal Bibliometric Report (JBR)*. JBR provides quantitative indicators of impact and bibliometric quality for more than 370 Serbian academic journals. The indicators are based on the data extracted from *SCIndeks - Serbian Citation Index* ([Šipka, 2005](#)) and the *Web of Science*.

The interpretation of JBR indicators relies heavily on the appropriate definition of the reference groups of journals with similar expected (theoretical) impact. Such groups are usually formed by classifying journals into categories based on subject similarity. A generally accepted view is that the Impact Factor can be considered meaningful and valid only if journals are compared within a particular scientific field or research area ([Testa & McVeigh, 2004](#)). JBR supports two such journal disciplinary classifications: the Field of Science and Technology Classification from the Frascati Manual (FOS), and the classification provided by the authorized Field Committees of the Ministry of Education and Science of Serbia (MESS).

Generally, the problem of journal classification may be formulated as a question of whether to use *ex ante* or *ex post* classification ([Leydesdorff, 2006](#)). Journals, including those newly established, could be classified “by force” into one of the already existing groups or subject areas. Alternatively, relationships among journals could be analyzed by calculating different types of proximity measures in order to extract clusters of journals with a similar thematic scope or citation patterns. Two techniques are commonly used: co-citation analysis ([Small, 1973](#)) and bibliographic coupling ([Small & Koenig, 1977](#)). Recently, several novel methods have been introduced in order to increase the accuracy of classification and visualization of clustered journals ([Janssens, Zhang, Moor & Glänzel, 2009](#); [Boyack & Klavans, 2010](#)).

The analyses presented in this paper were performed on a sample of journals in the fields of social sciences and humanities (SS&H). The main in-

tention was to test the suitability of journal citation data for classification tasks in those disciplines, since it is well known that SS&H are more nationally oriented and rely to a greater extent on the non-periodical literature as a principal channel of scientific communication (Hicks, 2004). Furthermore, the two JBR classifications differ significantly, particularly in the fields of SS&H. Hence, the second goal of this study was to compare the existing journal subject categories with different types of journal clusters extracted from journal citation data.

Method

Data source

All data were extracted from the SCIndeks database. The first step was to create a list of journals in the fields of SS&H. Only active journals, having published at least one issue in 2010 or 2011, were taken into account. The motivation was to classify all active journals and, hence, there were no conditions set regarding the minimal number of journal citations or co-citations among journals. The final list resulted in 137 journal titles covering different subject areas, from psychology and sociology, to history, linguistics, and literature.

The second step was to create a set of articles. In addition to the articles published in the selected journals, the sample also included all articles that cited any of the articles published in the selected journals. This means that the sample contained articles published in journals not necessarily restricted to SS&H. The publication period was limited to 10 years (2002-2011). The final sample contained 22,863 articles.

Similarity measures

Three types of similarity measures among journals were calculated. The first measure was based on *journal intercitation relations* (IC). The similarity of two journals was defined as the number of articles from the first journal that cited or had been cited by any article from the second journal. The second measure was based on *journal co-citation frequency* (CC). The linkage strength of two journals was defined as the number of articles citing both of them. Finally, the third measure was based on *bibliographic journal coupling* (BC). The strength of journal coupling is usually defined

as the number of documents cited by the two journals ([Small & Koenig, 1977](#)). However, when applied to the selected sample, this type of distance measure resulted in a sparse similarity matrix which could not be used for further analyses. As an alternative, the strength of journal coupling was defined as the number of pairs of articles from two journals citing the same journal.

Journal clustering

The resulting matrices were normalized in order to obtain more accurate similarity measures. Intercitation and co-citation matrices were normalized using the *Jaccard Index*. Since the degree of similarity in the case of bibliographic coupling was actually the number of pairs of documents, the matrix was normalized by dividing the number in each cell by the maximum number of document pairs for two journals. For example, if the number of documents for journals A and B are 50 and 60 respectively, then the actual number of pairs of documents citing the same source was divided by the value of 3000.

Normalized similarity matrices were transformed into distance matrices by calculating the inverse of the cell values ($d_{ij} = 1 - a_{ij}$). Journal clusters were derived from each distance matrix by using the *Partition Around Medoids* (PAM) algorithm in the R statistical package ([Kaufman & Rousseeuw, 2008](#)). Additionally, *nonmetric multidimensional scaling* (MDS) was used to create journal maps and visualize the relations among clusters.

Results

Validation of journal classifications

Social sciences and humanities journals represented in the JBR are currently classified into eight major categories according to the MESS classification: 1. economics, 2. philosophy, sociology, and political sciences, 3. psychology and educational sciences, 4. law, 5. language and literature, 6. history, archeology, and ethnology, 7. other social sciences, and 8. other humanities. The main differences between the FOS and MESS classifica-

tions are that the first one treats the fields of psychology and educational sciences as separate clusters, and also philosophy is in the category of humanities. Both classifications were used as external criteria for the validation of different clusterings based on citation data. The *Adjusted Rand Index* (ARI) was used as a measure of similarity among five clusterings. Generally, ARI values can range from 0 to 1, where the value of 1 indicates that two classifications are exactly the same.

The ARI values are shown in [Table 1](#). They are relatively low, but it should be noted that this does not necessarily indicate a low similarity among classifications nor a poor clustering, since a large number of journals naturally belong to more than one category. All classifications based on citation data are more similar to the MESS than to the FOS classification. The clustering based on the BC method showed the strongest agreement with both JBR classifications.

Table 1. Values of the Adjusted Rand Index of similarity among five journal clusterings

	intercitations (IC)	co-citations (CC)	bibliographic coupling (BC)
MESS	0.43	0.34	0.53
FOS	0.36	0.30	0.37
IC		0.58	0.60
CC			0.49

MESS and FOS are current journal classifications used in the Journal Bibliometric Report

Average silhouette widths (ASW) were used as measures of cluster validity and, indirectly, as an estimate of the relative suitability of different distance measures. ASW values, as well as the silhouette values of each individual object, can range from -1 to 1. Larger values indicate stronger cluster structure and greater similarity among cluster members. The classification based on the IC matrix had the lowest ASW (0.10). However, even in this case, the clustering algorithm did manage to reproduce the overall structure of JBR classification, since the six major categories of journals were clearly present. The remaining two clusters were composed

of journals in the field of political sciences, and journals in the field of physical education. All eight clusters had at least two members with negative values of silhouette, which could indicate that those journals belong to another (or independent) cluster. The clusters of economics, history, and sociology had the largest number of such spurious memberships.

The clustering solution based on journal co-citation frequencies also had relatively low ASW (0.14). The extracted clusters were very similar to those generated from the IC matrix in terms of subject contents. The political sciences cluster was also present, but sport sciences journals were merged with the field of psychology and educational sciences. In contrast, this procedure isolated a cluster of journals in the field of arts. When compared to the IC solution, it was noticeable that a large number of journals changed their cluster membership simply by switching positions between related disciplines, e.g. from sociology to political sciences, from economy to law, or from law to political sciences. However, there were more than twenty journals which were obviously completely randomly assigned to clusters. A further inspection revealed that those were the journals that had very low citation rates and, consequently, weak or no co-citation linkages with other journals.

Finally, the clustering solution based on the bibliographic journal coupling method yielded the most reasonable classification of journals, with the ASW being 0.37. Six of the eight clusters were thematically the same as those extracted from the inter-citations and co-citations matrices. Compared with the classifications based on IC and CC data, the use of bibliographic coupling resulted with more “natural” clusters. Clusters were much more homogeneous and all of them had ASW values above 0.30, which was rather high given the nature of citation connections and journals as objects of classification ([Janssens et al., 2009](#)). The single exception was the cluster of journals in the fields of sociology, philosophy and political sciences. However, not even this cluster could have been labelled as spurious or inappropriate. It was obvious that those journals did belong to a single cluster, but because of their strong connections with journals from other disciplines (e.g. law and psychology), the average silhouette width of this cluster was only 0.18.

Optimal number of journal clusters

All three distance measures based on citation data have proved to be a solid basis for the separation of journals from the six major disciplines listed above. Low ASW values and the occurrence of several clusters of highly specialized journals could have been indicators that the division into eight clusters was not sufficient to provide an acceptable differentiation of disciplines. However, increasing the number of clusters in the PAM model did not improve the homogeneity of individual clusters, nor the overall quality of clustering. Moreover, in the case of bibliographic coupling, after setting the number of clusters to higher values, ASW constantly decreased. In the case of distance measures based on intercitation and co-citation data, increasing the number of clusters did result with higher values of ASW, but such improvement was obviously due to a high homogeneity of several clusters consisting of only three or four members. Additionally, at least two clusters based on intercitation data have apparently emerged as a result of the regional proximity of journal publishers.

ASW values for all three solutions became more or less constant for the number of clusters larger than 13, indirectly suggesting the optimal number of journal categories. Since it was rather difficult to assess the feasibility of clusters without broader insight into the relations among disciplines, the clustering results were visualized using nonmetric MDS. Only the maps based on bibliographic coupling and co-citation data are presented in this paper. Except for the lack of space, the reasons for such a decision were the high similarity between IC and CC maps, and the presumed higher objectivity of co-citation counts as a proximity measure. Initially, both maps were created for all 137 journals. However, the CC map suffered from several artifacts, so in that case the sample was limited to journals having at least ten co-citations with other journals. This requirement has reduced the number of displayed journals to 124. Figures [1](#) and [2](#) show the maps of journals based on the two clustering criteria. Clusters are marked with gray circles. The most representative members of each cluster (so called *medoids*) are displayed in larger font. Some journal titles are omitted in order to make the maps more readable.

The two MDS maps have very similar structure. The “core” of social sciences is positioned centrally. It contains journals in the fields of sociolo-

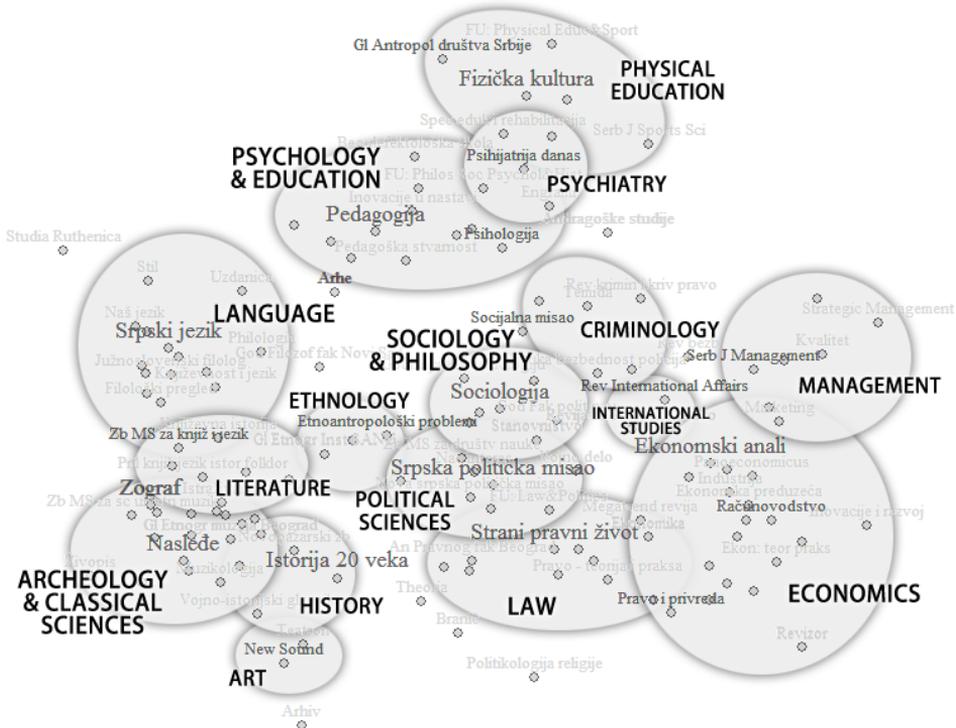


Figure 1. MDS map of 137 Serbian social sciences and humanities journals based on bibliographic journal coupling data (2002-2011)

gy, philosophy, law, history, economics, and political sciences. History and ethnology act as intermediaries or links between social sciences on one side, and humanities on the other. Both maps indicate the existence of a rather isolated cluster of disciplines obviously gravitating towards natural sciences (psychology, psychiatry, and physical education). This result is consistent with some previous findings regarding the position of psychology as a “hard” social science discipline (Ding, Chowdhury & Foo, 2000). The bibliographic coupling method has yielded clusters that were more homogeneous and compact. This was primarily due to a denser matrix when compared to the matrix of co-citations. However, it could be assumed that bibliographic coupling is probably a more appropriate and subtle technique for detecting subject similarities.

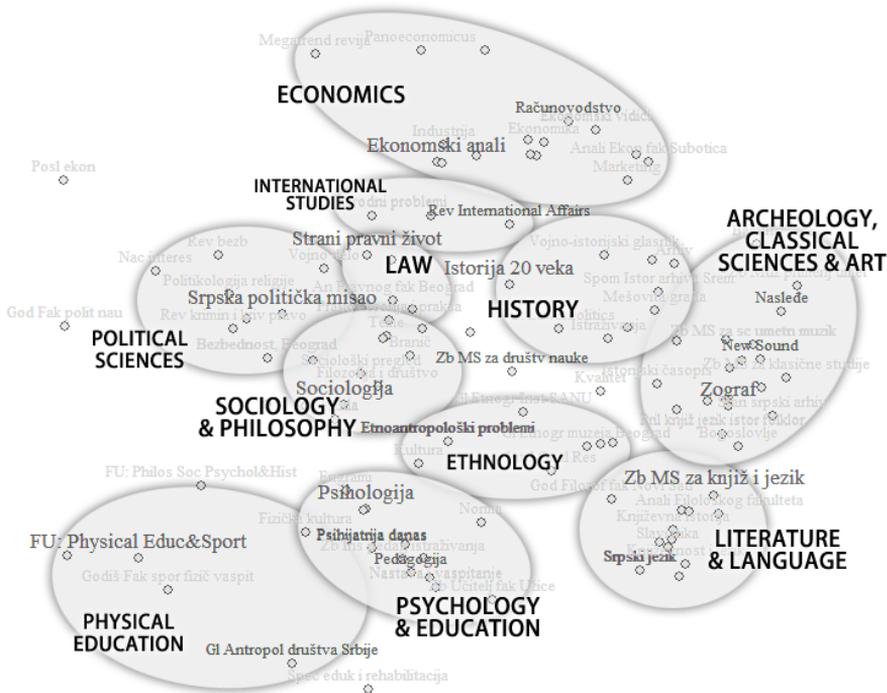


Figure 2. MDS map of 124 Serbian social sciences and humanities journals based on journal co-citation data (2002-2011)

Individual journal positions

The described methodology has revealed several “migrations” of journals among disciplines. For example, on the basis of co-citation patterns, the *Psihologija* (*Psychology*) journal is classified together with educational sciences journals, but on the basis of the sources used (BC) it is closer to psychiatry. Similar differences can be observed in the case of several history journals. Furthermore, it seems that some journals are far from their “parent” clusters. For example, the *Journal of the Anthropological Society of Serbia* (*Glasnik antropološkog društva Srbije*) is closer to sports sciences than to its current reference group of mainly ethnology journals. Finally, both maps revealed the central positions of several multidisciplinary journals.

Conclusions

The presented results have indicated that journal citation data are suitable not only for classification tasks in the fields of SS&H, but also as a mean of scientific domain analysis at the journal level. Bibliographic journal coupling has served as a good and robust alternative to co-citation and intercitation analysis, and has confirmed some previous findings regarding the accuracy of research front representation using the BC method ([Boyack & Klavans, 2010](#)).

The answer to the question about the optimal number of journal clusters cannot be simple and precise. The results have shown that the current MESS journal classification is valid, but should be more specific. However, increasing the number of groups could split up some heterogeneous but “natural” clusters, since there are many multidisciplinary journals. In that sense, it should be very useful to repeat this analysis on the sample of all national journals.

Finally, some previous research results have suggested that field normalization should solve the problem of short citation windows ([van Leeuwen, 2006](#)). Although this analysis was based on data covering a ten-year period, the resulting matrices were very sparse, and yielded several serious artefacts. Additional analyses for different and shorter periods of time would be very useful to shed more light on this problem.

References

- Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *JASIST*, 61, 2389-2404. [doi:10.1002/asi.21419](https://doi.org/10.1002/asi.21419).
- Ding, Y., Chowdhury, G., & Foo, S. (2000). Journal as markers of intellectual space: Journal co-citation analysis of Information Retrieval area, 1987-1997. *Scientometrics*, 47, 55-73. Retrieved from http://link.springer.com/chapter/10.1007%2F978-3-540-45175-4_45
- Hicks, D. (2004). The four literatures of social science. In H.F. Moed, W. Glänzel & U. Schmock (ed.), *Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems* (pp. 473-496). New York: Kluwer Academic Publishers. Retrieved from http://link.springer.com/chapter/10.1007%2F1-4020-2755-9_22

- Janssens, F., Zhang, L., Moor, B. D., & Glänzel, W. (2009). Hybrid clustering for validation and improvement of subject-classification schemes. *Inform. Processing & Management*, 45(6), 683-702. [doi:10.1016/j.ipm.2009.06.003](https://doi.org/10.1016/j.ipm.2009.06.003).
- Kaufman, L. & Rousseeuw, P. J. (2008) *Finding groups in data: An introduction to cluster analysis*, Hoboken: John Wiley & Sons.
- Leydesdorff, L. (2006). Can scientific journals be classified in terms of aggregated journal-journal citation relations using the Journal Citation Reports? *JASIST*, 57(5), 601-613. [doi:10.1002/asi.20322](https://doi.org/10.1002/asi.20322).
- Small, H. (1973). Cocitation in the scientific literature: A new measure of the relationship between two documents, *Journal of the American Society for Information Science*, 24(4), 265-269. [doi:10.1002/asi.4630240406](https://doi.org/10.1002/asi.4630240406).
- Small, H. G., & Koenig, M. E. D. (1977). Journal clustering using a bibliographic coupling method. *Information Processing & Management*, 13(5), 277-288. [doi:10.1016/0306-4573\(77\)90017-6](https://doi.org/10.1016/0306-4573(77)90017-6).
- Šipka, P. (2005). The Serbian Citation Index: Context and content. In; *Proc. of ISSI 2005 - 10th Int. Conf. of the SSI*, Stockholm, Sweden, July 24-28, 2005 (pp. 710-711). Stockholm: ISSI and KUP. Retrieved from http://ceon.rs/pdf/Sipka_SCIIndeks_proceedings.pdf
- Testa, J. & McVeigh, M.E. (2004) The impact of open access journals: A citation study from Thomson ISI. Retrieved from http://www.lib.uiowa.edu/scholarly/documents/ISI_impact-oa-journals.pdf, 2012 March 23.
- van Leeuwen, T. (2006). The application of bibliometric analyses in the evaluation of social science research. Who benefits from it, and why it is still feasible. *Scientometrics*, 66(1), 133-154. [doi:10.1007/s11192-006-0010-7](https://doi.org/10.1007/s11192-006-0010-7).

“Face Validity” Differences between Locally Published and Highest Level International Journals Based on the In-Text Citations Approach

Siniša Subotić

sinisasub@gmail.com

University of Novi Sad, Faculty of Philosophy, Dpt. of Psychology, Dr Zorana Đinđića
2, 21000 Novi Sad (Serbia)
NGO “Persona”, 78000 Banjaluka (Bosnia and Herzegovina)

Abstract

Locally published journals are often in a disadvantage when being compared to the highest level international journals using popular approaches such as the impact factor. This research explores an alternative approach, which focuses on the internal article’s “face validity” characteristics, based on the in-text citations measures. The sample comprised 157 empirical articles (psychology = 65, kinesiology = 32, psychiatry = 26, pedagogy = 23, and others = 11) from the journals published in Serbia in the period from 2004 to early 2010, which were compared to the matched 157 articles from the highest level international (English language) journals. ULS factor analysis conducted on the various in-text citations variables revealed two factors: *Integration* (containing variables such as: number of citations and references, number of citations that occur in multiple sections etc.), and *Superficiality* (containing: missing, imprecise and secondary citations). Large differences were obtained between the groups on both factors: Integration was significantly higher in the international group ($d = 1.49$), and Superficiality in the Serbian group ($d = 1.10$). Ward’s cluster analysis conducted on the two factors’ scores revealed that the articles can be classified in two clusters, with the first containing 117 international, and 9 Serbian articles (psychology = 4, psychiatry = 2, kinesiology = 2, and pedagogy = 1), and the second containing 148 Serbian, and 40 international articles (kinesiology = 16, psychiatry = 12, psychology = 5, pedagogy = 4, and others = 3), with the cluster distribution difference between the local Serbian and international articles being significant: $\chi^2(1, n = 314) = 154.61, p < .001, \phi = -.70$. Strong differences between the clusters on both factors were obtained, showing higher scores on Integration in cluster one ($d = 2.14$), and higher scores on Superficiality in cluster two ($d = 0.93$), thus the clusters were named as a *Higher tier* and a *Lower tier (of articles)*. It should also be noted that there is a certain trend of improvement in Serbian articles on the Integration di-

mension over the observed time span ($\rho = .29, p < .001$). In general, the results confirm that the in-text citations based approach, and subsequently derived latent dimensions can be used to differentiate locally published and the highest level international journal articles across various scientific disciplines, and that on the average, most of the Serbian articles fall behind most of the international articles, both on the grounds of internal citation integration, and citation superficiality and mistakes. These findings directly suggest the areas of improvement for journals' editorial boards, reviewers, and authors.

Keywords: local journals; journal quality; citation analysis; face validity; social sciences; Serbia; factor analysis; cluster analysis; impact factor

Introduction

Although there are some differences and exceptions between disciplines, scientific journal articles are the main and increasingly important source of scientific communication and diffusion of knowledge ([Larivière, Archambault, Gingras, & Vignola-Gagné, 2006](#)). Therefore, determining and continuously monitoring the quality of journals and articles is of the utmost importance. The most (in)famous and widely used tool for this purpose is definitely the impact factor (see [Althouse, West, Bergstrom, & Bergstrom, 2009](#); [Garfield, 2006](#)). However, the impact factor tends to vary across fields and over time ([Althouse et al., 2009](#)), making it non-optimal for interdisciplinary comparisons. Furthermore, although it can be considered as a valid measure for the quality assessment of scientific journals, it is not a valid measure for assessing the quality of individual papers and scientists ([Opthof, 1997](#)).

The usage of the impact factor is especially problematic when evaluating so called locally published journals (i.e. journals published outside of the most developed countries, and often not in English language), because they are generally underrepresented in the major citation indexing services (Archambault, Vignola-Gagné, Côté, [Larivière, & Gingras, 2006](#); [Šipka, 2003](#)), to the extent that the impact factor is arguably a completely invalid measure of these journals' quality ([Šipka, 2003](#)). Since locally published journals are a necessary complement to the highest level international journals, as they deal with the region, culture and language specific topics and serve as an educational and training tool for young researchers ([Šipka, 2003](#)), it is necessary to try and keep local journals at

least somewhat competitive with their international counterparts regarding quality. Therefore, different, less “unfair” ways of comparison besides the impact factor should be explored.

There are various possible ways to approach the evaluation of the local journals’ (formal) quality besides the impact factor (see [Šipka, 2003](#)). One of the less known concepts is in-text citations counting, which has recently been demonstrated as a successful way to improve the accuracy of assessing scientific contribution ([Hou, Li, & Niu, 2011](#)), and is a simple alternative to the complex context analysis approaches (e.g. [Huang et al., 2006](#)). Similarly to the impact factor, a possible weakness of this concept is large variability in the authors’ motivations to cite a source, as well as in the purposes that citations serve ([Bornmann & Daniel, 2008](#); see also [de Bellis, 2009](#)). However, none of this seems to call into question the role of a citation as a reliable measure of impact ([Bornmann & Daniel, 2008](#)).

[Šipka \(1996\)](#) proposed an in-text citations approach that takes into account not only the number of individual citations, but also the citations’ “re-coverage” (i.e. multiple occurrence) in different sections of the article, the citations’ age, etc. This approach, thus, focuses on the article’s internal citation characteristics, which arguably captures some of its immanent formal quality characteristics – [Šipka \(1996\)](#) refers to this as the “face validity”. [Šipka \(1996\)](#) successfully tested the “face validity” idea, showing that various in-text citations based variables group into two discriminant functions: *Article’s integrity* and *Thoroughness vs. superficiality*, and that based on them meaningful and reasonably precise discrimination between highest level international and locally published Serbian psychology articles can be achieved.

Considering the general impact factor based comparisons’ limitations ([Althouse et al., 2009](#); [Opthof, 1997](#)), and the specific limitations for local vs. international journals comparisons ([Šipka, 2003](#)), the problem of this research is to explore the alternative, in-text citations approach, and its usefulness in comparing the formal quality of local and the highest level international journals. Specifically, the goal is to revise and expand [Šipka’s \(1996\)](#) ideas, and test if such in-text citations based “face validity” measures can be used to compare the formal quality of local and interna-

tional journals from multiple scientific disciplines, since Šipka's (1996) findings were limited to the field of psychology only. If the results are satisfactory, these procedures could be considered as tools for monitoring and improving local journals' quality.

Method

Sample and procedure

The sample comprised 157 scientific articles from the journals published in Serbia (i.e. locally published journals) in the period from 2004 to early 2010, which were compared to the matched 157 articles from the highest level international (English language) journals¹. Following Šipka's (1996) original procedure, only the empirical, IMRAD format journal articles (i.e. articles containing Introduction, Method, Results and Discussion sections) were included in the research. This also included several Serbian ($n = 7$) and international ($n = 3$) articles that had slight deviations from IMRAD (but were logically equivalent to it), or that had Results and Discussion as a joined section. The latter was more characteristic for Serbian ($n = 50$) than the international articles ($n = 3$, with $n = 9$ "hybrid" structure articles²). Since these aspects were impossible to control, analogous to Šipka (1996), Results and Discussion were treated as one section for all the articles (referred to as: Results & discussion).

The selection of the scientific disciplines to include journals from was dictated by the Serbian sample. Because of various differences between scientific disciplines (Althouse et al., 2009; Archambault, et al., 2006; de Bellis, 2009; Fischer, Tobi, & Ronteltap, 2011; Larivière, et al., 2006; Martin, & Sugarman, 2009), the research was deliberately limited to social sciences journals (*humanities*, i.e. literature, philosophy etc. were considered, but were dropped due to the low number of empirical articles), with the addition of cross-discipline fields, such as psychiatry and kinesiology. All of the articles from the selected time span available on the Serbian Citation Index (Šipka, 2005) in full, and lexically searchable text (i.e. not "scans") were considered, minus the journals that had less than two publications in total, and the outliers ($n = 3$ articles had to be eliminated because they contained no in-text citations). This left 65 psychology, 32 kinesiology, 26 psychiatry, 23 pedagogy and 11 other³ arti-

cles; 26 were written in English, 16 in English and Serbian, and the rest were available in Serbian only.

The international group comprised the same exact number of articles for every field, while also being controlled for a year of publication. These articles were selected as a deliberate, heterogeneous sample⁴; the full list of included journals is available on request.

In-text citations were counted with respect to the articles’ sections.

Data analysis

All the analyses were conducted in R environment ([R Development Core Team, 2005](#)), using FACTOR ([Lorenzo-Seva, & Ferrando, 2006](#)), or manually. In-text citations counting was done manually, and reference parsings were done by the CEON Parser (see [Šipka, 2005](#)).

Results

Variable preselection

Initially, following and expanding upon [Šipka’s \(1996\)](#) research, a larger pool of variables was considered: number of citations (total number, and number and proportion in each of the articles’ major sections), number of references, coefficient of repetition (number of citations / number of references), number and proportion of sources cited various times (e.g. only once), number and proportion of missing citations (i.e. cited in text, but not on a reference list) and references (i.e. appearing on a reference list, but not cited in text), count of references occurring in two of the article’s sections (\cap), number of secondary citations (i.e. “as cited in”), number of imprecise citations, various measures of citations’ age, etc.

However, a lot of variables were omitted due to no significant contribution (e.g. citations’ age measures⁵), low discriminativity (e.g. coefficient of repetition), multicollinearity and / or curvilinearity issues (all the proportion based measures were severely problematic, except the proportion of missing references, which behaved better than its raw number counterpart). Hence, after a systematic preselection process and trial factor analyses, only the ten most optimal variables were kept in the final model (see the *Latent dimensions* subsection).

Latent dimensions

The ten selected variables were analyzed using an unweighted least squares (ULS) factor analysis. Optimal parallel analysis ([Timmerman & Lorenzo-Seva, 2011](#)) based on $n = 1000$ permutations suggested one (95th percentile criteria) or two (mean criteria) significant factors. However, the *hull* method ([Lorenzo-Seva, Timmerman, & Kiers, 2011](#)) favoured the two-factor solution, and since it was also easier to interpret it was selected as the final one⁶. The factors were rotated using *promin* rotation. The results of the factor analysis are presented in [Table 1](#).

The first factor comprises variables concerning citing (and referencing) more, and citing the same sources in multiple sections. While (due to a different variable preselection) not identical to [Šipka's \(1996\)](#) first discriminative function (*Article's integrity*)⁷, this factor is directly analogous to it, probably representing the citation characteristics of a well grounded and integrated scientific text. Hence, the factor is named *Integration*.

On the surface, the second factor only mildly resembles [Šipka's \(1996\)](#) second discriminative function (*Thoroughness vs. superficiality*). However, they seem to share a similar gist, which is arguably better represented in this solution, due to the inclusion of more error measuring variables. Basically, this factor reflects citational mistakes, with some tendencies towards lower thoroughness (i.e. lower literature coverage); thus, the factor is named *Superficiality*.

Group differences

Large differences ([Cohen, 1992](#)) were obtained between the two groups of journals on both factors' scores: Integration was significantly higher in the international journals group ($t(239.732) = -13.16, p < .001, d = 1.49$), and Superficiality in the local Serbian journals group ($t(194.046) = 9.65, p < .001, d = 1.10$).

To explore the “grouping tendencies” of the articles, Ward's hierarchical cluster analysis was conducted, using the two factors' scores. Multiple validation criteria (e.g. Silhouette and Connectivity; see [Handl, Knowles, & Kell, 2005](#)) suggested that the two-group solution is the most optimal. The first cluster contained 117 international, and only 9 Serbian articles

Table 1. Factor matrix

Variable	<i>Integration</i>	<i>Superficiality</i>	h^2	Variance
Number of citations	.99 (1.00)		1.00	1740.32
Number of references	.88 (.93)	(-.41)	.90	518.95
Introduction \cap Results & discussion	.75 (.74)		.56	30.33
Introduction \cap Method	.71 (.65)		.47	5.27
Number of single citations	.68 (.74)	(-.42)	.61	249.47
Method \cap Results & discussion	.61 (.57)		.34	4.00
Missing references (proportion)	-.33 (-.35)		.13	.02
Number of missing citations		.56 (.54)	.30	1.66
Imprecise citations ⁸		.56 (.54)	.29	.89
Number of secondary source citations		.55 (.54)	.29	6.60

Notes: Loadings in front of the brackets are factor pattern loadings, while the numbers inside the brackets are factor structure loadings (values $<.32$ were suppressed); Variables with \cap refer to a count of citations that occur in two of the article’s sections (i.e. citations’ “re-coverage”); h^2 = communality after extraction; $KMO = .72$; *Integration* eigenvalue = 4.20; *Superficiality* eigenvalue = 1.69; Variance before the rotation = 58.85%; Factors are negatively correlated in a lower medium intensity (Cohen, 1992): $r = -.25, p <.001$ ⁹

(psychology = 4, psychiatry = 2, kinesiology = 2, and pedagogy = 1), with two of them being in English, and the second cluster contained 148 Serbian, and 40 international articles (kinesiology = 16, psychiatry = 12, psychology = 5, pedagogy = 4, and others = 3). The cluster distribution difference between the local Serbian and international articles is significant and strong (Cohen, 1992): $\chi^2(1, n = 314) = 154.61, p <.001, \phi = .70$. Large differences (Cohen, 1992) between the clusters on both factors were obtained, with higher scores on Integration in cluster one ($t(169.433) = 16.51, p <.001, d = 2.14$), and higher scores on Superficiality in cluster two ($t(262.577) = -9.27, p <.001, d = 0.93$), thus the clusters were named as a *Higher tier* and a *Lower tier (of articles)*.

It should also be noted that there is a lower medium ([Cohen, 1992](#)) trend of improvement in Serbian articles on the Integration dimension over the observed time span ($\rho = .29, p < .001$).

Discussion

The main goal of this research was to explore the in-text citations based approach ([Hou et al., 2011](#); [Šipka, 1996](#)), as an alternative to the non-optimal impact factor based assessment of the locally published journals' quality ([Šipka, 2003](#)). In this regard, a “revived” and slightly modified [Šipka's \(1996\)](#) “face validity” concept turned out to be a useful tool for comparing the local (Serbian) and the highest level international scientific articles from several fields.

Two meaningful latent in-text citations dimensions (reasonably similar to the functions obtained by [Šipka, 1996](#)) concerning (1) internal citation integration and (2) citation superficiality and mistakes were obtained, and based on them it was shown that local Serbian articles fall strongly behind the international ones on both grounds, with only a handful of them reaching the higher tier of articles. While these findings should be replicated on other fields' (e.g. natural sciences & engineering) articles, and on local journals other than Serbian to allow for broader generalization, and while they can only be safely interpreted as differences in formal (citation) quality, they nevertheless suggest several areas of improvement (mostly) for local journals. First off, even though there is already a certain trend of improvement over time, that trend is not sufficiently large, thus local (Serbian) journals' editors and reviewers should explicitly insist on higher literature coverage and better integration of ideas and arguments between the sections of submitted manuscripts (i.e. aspects indirectly tentatively measured by the Integrity factor). Authors should also make a greater effort to follow those guidelines. Furthermore, more effort on everyone's part should be put into discouraging the use of secondary sources, and also the detection and correction of mistakes of any kind, including citation errors, even after publication – while this might seem trivial, the rate of error correction and retraction is actually a strong sign of journals' quality ([Fang & Casadevall, 2011](#)). Thus, journals should adopt a strict(er) peer and editorial review process, which would include citations and other formal mistakes checking procedures. It can be argued that higher formal

strictness during reviews should probably enhance the educational and training roles of local journals (Šipka, 2003), and complement other aspects of quality control, since “face validity” is necessary, but not enough to alone guarantee quality (Šipka, 1996).

Besides the already mentioned necessities for replications, since “face validity” is more of a component than a correlate of quality (Šipka, 1996), further research should aim to take some additional aspects of quality into consideration. Furthermore, some limitations of this research should be revised, e.g. more advanced mathematical procedures (that are not limited by curvilinearity, thus allowing the integration of more potentially useful variables) should be explored, and some additional indicators should also be considered. Finally, between / within disciplines specificities and “face validity” norms should be determined.

Acknowledgements

The author wishes to thank P. Šipka for his logistic help, and O. Dakić for help with English.

References

- Althouse, B. M., West, J. D., Bergstrom, C. T., & Bergstrom, T. (2009). Differences in impact factor across fields and over time. *Journal of the American Society for Information Science and Technology*, 60(1), 27-34. [doi:10.1002/asi.20936](https://doi.org/10.1002/asi.20936).
- Archambault, É., Vignola-Gagné, É., Côté, G., Larivière, V., & Gingras, Y. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics*, 68(3), 329-342. [doi:10.1007/s11192-006-0115-z](https://doi.org/10.1007/s11192-006-0115-z).
- Bornmann, L., & Daniel, H. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45-80. [doi:10.1108/00220410810844150](https://doi.org/10.1108/00220410810844150).
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159. [doi:10.1037/0033-2909.112.1.155](https://doi.org/10.1037/0033-2909.112.1.155).
- de Bellis, N. (2009). *Bibliometrics and citation analysis: From the Science citation index to cybermetrics*. USA: The Scarecrow Press. Retrieved from http://203.128.31.71/articles/0810867133_LIS.pdf

- Fang, F. C., & Casadevall, A. (2011). Retracted Science and the Retraction Index. *Infection and Immunity*, 79(10), 3855-3859. doi:[10.1128/IAI.05661-11](https://doi.org/10.1128/IAI.05661-11).
- Fischer, A. H., Tobi, H., & Ronteltap, A. (2011). When Natural met Social: A Review of Collaboration between the Natural and Social Sciences. *Interdisciplinary Science Reviews*, 36(4), 341-358. doi:[10.1179/030801811X13160755918688](https://doi.org/10.1179/030801811X13160755918688).
- Garfield, E. (2006). The History and Meaning of the Journal Impact Factor. *Journal of the American Medical Association*, 295(1), 90-93. doi:[10.1001/jama.295.1.90](https://doi.org/10.1001/jama.295.1.90).
- Handl, J., Knowles, J., & Kell, D. B. (2005). Computational cluster validation in post-genomic data analysis. *Bioinformatics*, 21(15), 3201-3212. doi:[10.1093/bioinformatics/bti517](https://doi.org/10.1093/bioinformatics/bti517).
- Hou, W., Li, M., & Niu, D. (2011). Counting citations in texts rather than reference lists to improve the accuracy of assessing scientific contribution. *BioEssays*, 33(10), 724-727. doi:[10.1002/bies.201100067](https://doi.org/10.1002/bies.201100067).
- Huang, S., Yu, Y., Xue, G. R., Zhang, B. Y., Chen, Z., & Ma, W. Y. (2006). TSSP: Multi-features based reinforcement algorithm to find related papers. *Web Intelligence and Agent Systems*, 4(3), 271-287. Retrieved from <http://iospress.metapress.com/content/84erkhh60q0vvnwfw/>
- Larivière, V., Archambault, É., Gingras, Y., & Vignola-Gagné, É. (2006). The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology*, 57(8), 997-1004. doi:[10.1002/asi.20349](https://doi.org/10.1002/asi.20349).
- Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods*, 38(1), 88-91. pmid:[16817517](https://pubmed.ncbi.nlm.nih.gov/16817517/). doi:[10.3758/BF03192753](https://doi.org/10.3758/BF03192753).
- Lorenzo-Seva, U., Timmerman, M. E., & Kiers, A. L. (2011). The Hull method for selecting the number of common factors. *Multivariate Behavioral Research*, 46, 340-364. doi:[10.1080/00273171.2011.564527](https://doi.org/10.1080/00273171.2011.564527).
- Martin, J., & Sugarman, J. (2009). Does Interpretation in Psychology Differ From Interpretation in Natural Science. *Journal for the Theory of Social Behaviour*, 39(1), 19-37. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1468-5914.2008.00394.x/abstract>.
- Opthof, T. (1997). Sense and nonsense about the impact factor. *Cardiovascular Research*, 33(1), 1-7. doi:[10.1016/S0008-6363\(96\)00215-5](https://doi.org/10.1016/S0008-6363(96)00215-5).

- R Development Core Team (2005). R: A language and environment for statistical computing (Version 2. 14. 1) [Computer Software]. Austria, Vienna: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>.
- Šipka, P. (1996). Procena "očigledne valjanosti" naučnih radova iz oblasti psihologije na osnovu navoda u tekstu [Estimation of "Face Validity" of psychological research papers based on in-text citations]. *Psihologija*, 29(4), 417-428. Retrieved from <http://scindeks-clanci.ceon.rs/data/pdf/0048-5705/1996/0048-57059604417S.pdf>
- Šipka, P. (2004). Vrednovanje časopisa na osnovu bibliometrijskih pokazatelja: Opis modela i implikacije za uređivački rad [The evaluation of journals based on bibliometric indicators: Description of a model and implications for editorial work]. *Srpski arhiv za celokupno lekarstvo*, 38-49 (special issue). Retrieved from http://ceon.rs/pdf/vrednovanje_casopisa_na_osnovu_bibliometrijskih_pokazatelja.pdf
- Šipka, P. (2005). The Serbian Citation Index: Context and content. In: *Proceedings of ISSI 2005 - 10th International Conference of the Society for Scientometrics and Informetrics*, Stockholm, SE July 24-28, 2005 (pp. 710-711). Retrieved from http://ceon.rs/pdf/Sipka_SCIndeks_proceedings.pdf.
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, 16(2), 209-220. doi:10.1037/a0023353.

Notes

- [1] Šipka (1996) tried to differentiate between the “academic” and “non-academic” local journals, but his results showed some problems with such a distinction. Therefore, it was not considered here.
- [2] I.e. containing several Studies, but some with Results and Discussion separated, and some with them joined.
- [3] Containing mostly articles on geography and economy.
- [4] A random selection was not optimal, because some sort of rudimentary content control was necessary. For example, while pedagogy exists in Serbia, it is not an internationally established discipline, so the international subsample had to be selected out of the educational psychology articles dealing with specific topics, such as educational

measurement and tests, etc. Only titles, abstracts and key words were used in the content matching process. Furthermore, the Serbian group usually had one dominant field-specific journal, dealing with a wide range of topics, while international journals are generally more specialized – to match this, international articles were roughly proportionally (quasi-randomly) selected from multiple (two to three) highly ranked journals from each field. Following the described methodology, roughly 50% more international articles were preselected for each field (controlling for a year of publication); from this preselected pool, the final articles were selected randomly.

- [5] Even though irrelevant here, the citations' age should probably be standardly considered in future comparisons.
- [6] Factor invariance was impossible to check due to sample size, but it is plausible to assume that, while relevant on the Serbian and joined sample, the second factor would be suppressed on the international sample alone.
- [7] Note that this function also had a proportion of missing references negatively loading on it.
- [8] This variable does not include spelling errors, but rather cases where a cited source has multiple authors, but only one name is cited in-text, or when there are multiple works by the same author(s), from the same year, but they are not differentiated in text (i.e. "a, b, c..." were not used), etc. Note that this variable was truncated to 5 points (0, 1, 2, 3, 4+ imprecise citations) for two reasons: 1) 96.5% of the articles had a score of 3 or less, with a handful of articles with the extreme values; 2) in certain cases, the exact number of imprecise citations could not be precisely determined (this was always a case with articles already having a definite score of at least 3). Thus, coding the mentioned "problematic" articles simply as "4" was a convenient way to deal with those issues.
- [9] Note that this correlation is not a result of a proportion of missing references loading on the first factor (when that variable is omitted, the correlation drops off only for .005).

Open Access Journals in India: A Study Report with Special Reference to Computer Science Journals

Bhaskar Mukherjee

mukherjee.bhaskar@gmail.com

Banaras Hindu University, Department of Library & Information Science, Varanasi -
221005 (India)

Abstract

The paper provides a descriptive analysis of open access publishing in India, with a special focus on publishing in the field of computer science. There are 373 open access journals in India, with a noticeable number of titles appearing in the period 2009-2010. The most prominent subject areas are health and computer sciences. Open access with the author-pays model, has become successful in India. Although the number of articles published this way may be quite high, their citation index is modest.

Keywords: articles; citations; computer science; India; Open Access; providers; publishing; Web; impact

Introduction

India is one of the most academically prominent countries in Asia, with an education system dating back to the ancient universities of Nalanda and Takshashila in the 5th century BC. Academic and research infrastructure in India has increased in the past decade, reaching 400 universities, and 1000 colleges, with numerous research and development (R&D) laboratories in the fields of industry, agriculture, medicine, environmental sciences, space research, military engineering, energy and ocean studies. These institutions are the major driving force of the scholarly communication. In addition, India also has a very well developed publishing industry, with around 15,000 registered publishers and an annual production of 70,000 titles published in the 24 official languages of India, 18 of which are used by millions of people ([Khatri, 2009](#)). With the advent of Web technology, major steps were made in digitizing earlier scholarly resources, many of which are available in the open access (OA) domain.

Although India is on the 12th place in the world according to the number of journals, when it comes to OA, it holds a very high 5th, well ahead of countries with larger journal production such as the Netherlands, China, Germany, and Australia ([Sahu & Parmer, 2006](#)). This can be explained by the fact that a considerable number of journals in India are published by learned societies and associations, with editors having strong academic interest in publishing without the involvement of a commercial publisher. Furthermore, the Indian government shows interest and encourages the OA production. The most prominent examples include the Indian Academy of Sciences, with 11 scientific journals all of which are freely accessible on the Web, and the Indian National Science Academy (INSA) with 4 journals, proceedings, and monographs in open access. The INSA OA resources were launched in December 2003, as a part of the project “Building Digital Resources: Creating Facilities at INSA for hosting S&T Journals Online”. Last but not the least are the e-print archives of the Indian Institute of Science comprising research papers, both preprints and postprints, technical reports, unpublished findings, and journal articles.

In the R&D sector, there is the National Chemical Laboratory, a research lab in India that provides free access to their data, including data from the National Collection of Industrial Micro-organisms and the National Centre for Biodiversity Informatics. Equally relevant is the “National Health Information Collaboration” project by the Indian Council for Medical Research and the World Health Organization, which provides a portal for Indian health data and information and free open software that can be used for various other purposes. Another example is The Council of Scientific and Industrial Research, responsible for scientific industrial research in India, whose R&D unit aims to provide open access to Indian patents and medicinal plants information.

In the commercial sector, *Bioline International* – which is a not-for-profit collaborative application of the University of Toronto Libraries (Canada), the Reference Center on Environmental Information (Brazil), and Bioline (UK), provides access to 14 Indian journals on their primary site, as well as access to archives of these journals at the *Bioline EPrints Archive*.

Similarly, there is *Medknow Publications*, a commercial publisher providing free publishing and access services for over 30 biomedical journals.

In observing these developments one has to take into account that, while the majority of the e-publishers in the developed countries started between 1995-1999, in India e-publishing started only five to eight years ago. Hence, the concept of e-publishing is still fairly new for publishers in India compared to countries such as the USA or the UK, where it is already an established industry. Although the take up of open access publishing in India, as outlined above, is very impressive and seems to be very promising, in order to assess these services and their sustainability, a more detailed quantitative and qualitative analysis and evaluation is required.

Objectives

The overall aim of this study is to explore the status of Indian OA publishing in general and the status of OA publishing in the field of computer science and technology in particular. The specific goals of this study are:

- (1) To gather information about OA journals in terms of their age, publisher, and subjects;
- (2) To find out, based on information on the official journals' websites, what types of content management strategies were applied; and
- (3) To learn, through Google Scholar (<http://scholar.google.com/>) and the Indian Citation Index (<http://www.indiancitationindex.com/>), about the citation patterns of OA journals.

Method

The data collection started with the identification of open access journals. The identification of the OA journals was originally planned to be based on two sources: the Directory of Open Access Journals (DOAJ) and IndianJournals.com. DOAJ, one of the leading online directories, enlists scholarly, open access journals throughout the world available under various subjects. In February 2012, DOAJ listed 373 OA journals from India. IndianJournals.com is another web portal of interdisciplinary Indian journals including OA journals, maintained by Divan Enterprise. Upon the in-

initial assessment of IndianJournals.com it was established that no reliable information could be extracted from this service because of the fact that it did not provide unique OA journal titles consistently. Therefore, it was decided to base this study on the DOAJ data only. Upon acquiring information about journal titles, details about the age, sponsoring agency, host organization, language and subjects were collected from the journals' official websites. In order to study the content management process of the journals, a toolbox of criteria was developed, based on earlier published studies (Rousseau, 2002; Mukherjee, 2007). The citation pattern of OA articles has been measured by counting and analyzing the citations of Google Scholar and the Indian Citation Index.

Results & Discussion

Figure 1 shows the annual increase of OA journals in the period 2003-2011.

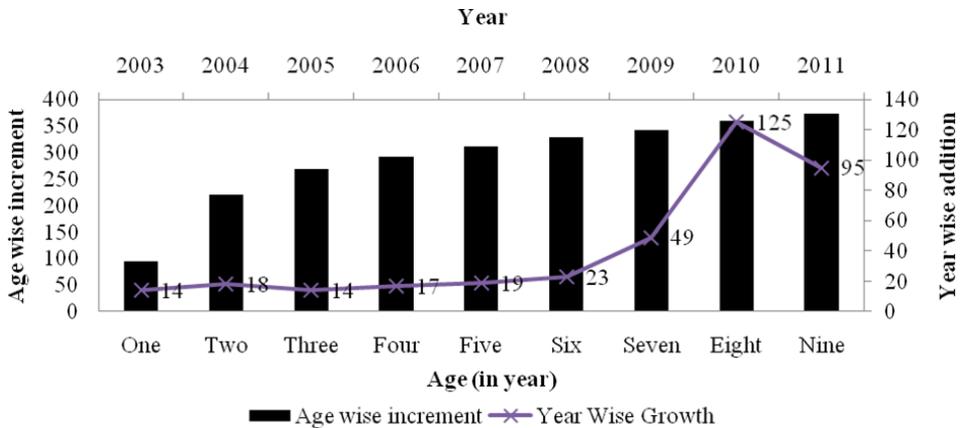


Figure 1. Growth of OA Journals in India

According to DOAJ, there were no OA journals in India in 2002, and following this period an increase to 378 by December 2011 can be observed. Between 2002 and 2008 there was an average annual increase of 18 journals per year. However, from 2009 to 2011 there was an annual rise of 49, 125 and 95 journals per year respectively, with an average of 90 new journals per year. This rapid change may be attributed to the intervention of various national donor agencies, such as the Department of Science and Technology and the Department of Biotechnology. In addition, the

governing bodies of the major research councils, such as the CSIR, UGC reached the conclusion that the results of all publicly funded research should be made available through self-archiving and encouraged open access journal publishing initiatives.

More detailed investigation of the journal growth rate showed that medical science and computer science are the two major fields in which the annual increase of journal titles was more prominent than in other subject fields.

[Table 1](#) shows the number of OA journals under ten broad subject areas. It should be noted that DOAJ categorises the 378 journals of India into 103 subject areas which are subsumed to broader categories primarily for the convenience of presentation. From the data shown in [Table 1](#), it is evident that nearly 50% of OA publishing in India is concentrated in the field of health and allied sciences, followed by 12% in the field of computer sciences and 10 % in Biological sciences. It should be noted that most of the journals of the two fields have started their publication from 2009 onwards. It is clear that the distribution of OA publishing across subject areas is very uneven and OA is yet to gain popularity among scholars in subject fields such as agriculture, arts etc.

Table 1. Leading Subject Domain of OA Publications in India

Leading Broad Subjects	No.	Percentage
Health Science & Allied	184	48.68
Computer Science	44	11.64
Biology & Allied	39	10.32
Sciences & Allied	27	7.14
Social Science & Allied	25	6.61
Engineering & Allied	21	5.56
Chemistry & Allied	17	4.50
Arts & Allied	9	2.38
Agriculture & Allied	8	2.12
Environmental Sciences	4	1.06

The next step was to identify the publishers playing the leading role in OA dissemination in the scholarly domain in India. [Table 2](#), shows the top ten publishers.

Table 2. Leading Publishers of OA Journals in India

Leading Publishers	No.	Percentage
Medknow Publications	81	21.15
Academy & Industry Research Collaboration Center (AIRCC)	23	6.01
NISCAIR	15	3.92
Bioinfo Publications	13	3.39
Indian Academy of Sciences	8	2.09
Kamla-Raj Enterprises, Delhi	7	1.83
Integrated Publishing Association	5	1.31
Engineering Journals Publications	4	1.04
Advanced Research Journals	3	0.78
Indian Academy of Sciences, Springer	3	0.78
Other (204 publishers)	221	57.70

The highest number of OA journals is being published by Medknow Publication (21.15%), followed by the Academy & Industry Research Collaboration Center [AIRCC] (6.01%), and NISCAIR (3.92%). However, it is important to note that there are 204 publishers who are publishing only a single journal in OA form. [Frantsvåg \(2010\)](#) analyzes the distribution of journals by publisher in the DOAJ. He finds that the distribution is strongly skewed, with nearly 90% of publishers represented by only one journal title and larger publishers representing only a quarter of the titles in the directory. Arguing from the point of view of economic theory, [Frantsvåg \(2010\)](#) concludes that the high proportion of single-journal OA publishers is a sign of an inefficient OA marketplace that is unable to take advantage of economies of scale in production and distribution.

From [Table 1](#) it is clear that health science and computer science are two major disciplines in Indian OA publishing. Since India has shown the po-

tential for high-caliber computer science research ([Kumar & Garg, 2005](#)), a more detailed analysis of the journals in the field of computer science only was conducted. [Table 3](#) shows data mostly concerning various web-site and technical quality variables for 44 computer science journals.

Table 3. Technical Quality Variables of Websites of Computer Science Journals

Evaluation Criteria	Frequency	Percentage
Whether the journal has an independent website	Yes	100
Nature of journal domain	.org	25
	.com	10
	.info/info.in/in	07
	.ac.in	01
	.net	01
Whether the Journal is peer-reviewed	Yes	100
Whether the journal has an International Editorial Board	Yes	100
Indian role in publication	: as publisher	42
Chief-Editor of the journal is from India		02
Availability of Journal:	: Print and online both	32
	: only online	12
Form of contents includes	: research article/survey article	44
		100.00
	: review article	16
	: case study	03
	: technical note	02
	: conference	01
	: letters	02
Submission process of article	: through email	35
	: online	05
	: both email & online	04
Timeliness of Publication	: on time	Yes
Author Coverage	: mostly Indian authors	26
	: mostly international authors	0
		0.00
	: mixed authors, national & international	18
		40.91
Type of OA model	: Complete OA	09
	: Author pay OA	33
	: Partial OA	02
		4.55

The question – do the resources have some reputable organization or expert behind them? – always remains an important issue in determining the quality of a journal. This question may be even more relevant in this case as articles are available free of cost. What could be defined as a positive side of OA journals in computer science is that most of them are being published by some Indian organization. Most of these non-profit organizations extend their academic activities by publishing OA journals. This may be viewed as a healthy sign for their future. The inclusion of an international editorial board and the peer-review policies of these journals, as mentioned in the journals' documentation, are some additional positive sides of these OA journals. Stevan Harnad once mentioned that “Once peer review is in place on the Net, once the quality hierarchy has been established, serious scholars will no longer have reason to hesitate to confer their best work to the electronic-only medium” ([Harnad, 1996](#)). While most of these journals are widely indexed by various international bibliographic services such as Google Scholar, EBSCO Publishing, ERIC, Index Copernicus, GeonD, Genamics, Open J-Gate, JournalSeek, ProQuest, Scirus, Ulrich Web, and University bibliographic databases such as those of Cornell University Library, California State University Library etc., the author coverage of these journals is basically either solely Indian (59%) or mixed (not necessarily in terms of joint authorship): Indian and international authors both (41%). However, one of the most important results is that most (75%) of these journals followed the author-pays model of OA in which publishers expect authors to pay handling charges of accepted papers before publishing. John Willinsky (as cited in [Zimmermann, 2009](#)) makes the point that there are many different models of open access publishing, for example sponsored OA, OA supported by author fees, and embargoed OA. Author-pays is now one of the models of OA publishing mostly present in developing countries. For instance, one of the publishers of a computer science journal, AIRCC, asks authors to pay \$120 (for the maximum of 20 pages) for publishing articles in their journal, while another publisher – Bioinfo, asks \$275 for publishing articles in their journals. These article processing charges replace subscription charges, and allow publishers to make the full-text of every published article freely

available to all interested readers. The majority of journals are asking between \$100-125 as handling charges for publishing articles. Some incomes needed to maintain the service also come from the sale of printed volumes, as most (72%) of these journals are also available in print form.

[Table 4](#) shows the citation pattern of OA journals in the field of computer science from Google Scholar. All investigated journals were indexed by Google Scholar and only one was indexed in the Indian Citation Index. Therefore, it was not possible to take into account the Indian Citation Index. Of the total 44 journals in computer sciences (indicated earlier in [Table 1](#)), the publication of 1 journal started in 2004, 14 in 2009, 21 in 2010, and 8 journals in 2011, the age of most (85%) of these journals thus being either two or three years.

Table 4. Citation Pattern of Computer Science Journals (Measured by Google Scholar)

Journals having Articles	No. of Journals	Total articles	Mean % of Cited Article	Total citations	Mean Citation/article
up to 25	8	170	78	60	0.34
> 25 and <=50	10	374	89	236	0.64
>50 and <=75	10	621	84	490	0.76
>75 and <=100	3	272	93	187	0.70
>100 and <=200	6	907	96	556	0.67
> 200 and more	7	2920	78	1289	0.45

As indicated in [Table 4](#), OA journals that published 50 to 100 articles in the last three years have a better citation rate than journals that published more than 100 articles. So, the quantity of articles cannot be the only criterion for getting more citations. Overall, all these journals received citations with an average rate of 0.54 citations per article, which is rather modest. The finding reveals that although open access articles have large accessibility, the accessibility does not necessarily attract more citations. Having said that, most of these journals are fairly new (2-3 years old) and are yet to build their scholarly reputation that may lead to more citations. This finding contradicts earlier research by [Lawrence \(2001\)](#), according to which computer science OA journals received more citations than traditional journals.

Anomalies

Although the publishing pattern of open access journals in India appears quite promising, the study has revealed some anomalies:

- (1) The number of articles per issue is rather large for some journals. A monthly journal – *International Journal of Computer Science and Information Technologies*, volume 2, issue 5, contains 132 articles. It seems to be unrealistic to maintain a quality reviewing system for a monthly issue with 132 articles.
- (2) There are at least two journals of AIRCC having the same name with different journals published by Medwell in Pakistan. Journal titles obviously need an additional identifier.
- (3) In Google Scholar (GS) most of the journal names are very roughly indexed: some articles are indexed with full journal name, while most of them are indexed with a different/abbreviated/ incomplete journal name. One should put complete caution while searching citations through journal name in the advanced search option of GS. Additionally, not all articles are indexed in GS in spite of the fact that the journal in which they are published is indexed in GS.
- (4) While searching the citation data of these journals through phrase search in GS, the searched results for most of these journals were quite surprising. In a number of instances, the GS results were for different journals having small variations in name, published by other publishers. One must take full care while searching in GS to avoid false citation statistics.
- (5) A number of articles that are published in these journals are also accessible through web archives like Citeseer, arXiv etc. which leads to duplicate citations of the same article.

Conclusion

The present study indicates that the OA publishing model in India has a rather large scope which may be interpreted as a sign of its vitality and robustness. In addition, the keen interest in OA demonstrated by Indian scholars is another important factor that may play a key role in the future

of open access scholarship. Author-pays is emerging as one of the successful models and the number of research articles being submitted for publication is quite high for all journals based on this kind of service. At the same time, it is essential that the journals' publishers should shift their attention from the quantity to the quality of articles being published. A coordinated effort and academic endorsement on the national level is essential in promoting and coordinating open access publishing systems and improving awareness. Most of the referenced journals in the field of computer science show a fairly low citation index, but this may be attributed to the fact that they are fairly new and are yet to gain wider academic reputation.

Acknowledgements

I am thankful to Prof. Aida Slavic, Editor-in-chief of the UDCC for her technical assistance and Department of Science & Technology, India and BHU for travel support.

References

- Frantsvåg, J.E. (2010). The size distribution of open access publishers: A problem for open access. *First Monday*, 15(12), Retrieved from <http://firstmonday.org/ojs/index.php/fm/article/view/3208/2726>, 2012 Feb 28.
- Harnad, S. (1996). Implementing peer review on the Net: Scientific quality control in scholarly electronic journals. U R. Peek & G. Newby (Ed.), *Scholarly publication: The electronic frontier* (pp. 103-108). Cambridge MA: MIT Press. Retrieved from http://cogprints.org/1692/1/harnad96.peer_review.html
- Khatri, N.K. (2009). E-Journal Publishing in India: A Case Study of Indian Statistical Institute Publication: Sankhya - The Indian Journal of Statistics. In: *Proceedings of the International Conference on Academic Libraries (ICAL-2009)* (pp. 386-88). Delhi: University of Delhi. Retrieved from http://crl.du.ac.in/ical09/papers/index_files/ical-65_186_390_1_RV.pdf
- Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature*, 411(6837), 521-521. [doi:10.1038/35079151](https://doi.org/10.1038/35079151).
- Mukherjee, B. (2007, Spring). Evaluating E-Contents beyond Impact Factor: A Pilot study selected Open Access Journals In Library and Information Science. *Journal of Electronic Publishing*, Retrieved from <http://hdl.handle.net/2027/spo.3336451.0010.208>, 2012 Feb 28.

- Kumar, S., & Garg, K.C. (2005). Scientometrics of computer science research in India and China. *Scientometrics*, 64(2), 121-132. [doi:10.1007/s11192-005-0244-9](https://doi.org/10.1007/s11192-005-0244-9).
- Rousseau, R. (2002). Journal Evaluation: Technical and practical issues. *Library Trends*, 50(3), 418-439. Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/8418/librarytrendsv50i3i_opt.pdf?sequence=1
- Sahu, D.K., & Parmar, R. (2006). The position around the world: open access in India. In N. Jacobs (Ed.), *Open Access: Key Strategic, Technical and Economic Aspects* (pp. 26-32). Oxford: Chandos Publishing. Retrieved from http://openmed.nic.in/1599/01/Open_Access_in_India.pdf
- Zimmermann, C. (2009). The Economics of Open Access Publishing. *Economic Analysis & Policy*, 39(1), 49-50. Retrieved from http://www.eap-journal.com/archive/v39_i1_4_zimmermann-intro.pdf

Maps of Science Based on Keywords of Articles' Antecedences, Presences, and Consequences: Application of the CEON/CEES Model of Multi-Perspective Description of Articles

Tanja Jevremov¹, Dejan Pajić², Miša Sotirović³ and Pero Šipka⁴

¹jevremov@neobee.net

University of Novi Sad, Faculty of Philosophy, Dpt. of Psychology, Dr Zorana Đinđića, 21000 Novi Sad (Serbia)

²dpajicns@gmail.com

University of Novi Sad, Faculty of Philosophy, Dpt. of Psychology, Dr Zorana Đinđića, 21000 Novi Sad (Serbia)

³sotirovic@ceon.rs

Centre for Evaluation in Education and Science, Kneza Miloša 17, 11000 Belgrade (Serbia)

⁴sipka@ceon.rs

Centre for Evaluation in Education and Science, Kneza Miloša 17, 11000 Belgrade (Serbia)

Abstract

A model of multi-perspective article description (MPAD) was explored and preliminarily tested. The model assumes that journal articles should be described for bibliographic purposes from three different perspectives, i.e. by using keywords extracted from: (1) article titles and abstracts, (2) titles of their cited references, and (3) titles and abstracts of articles citing them in the future. In order to explore the relationships among keyword types and to test the model preliminarily, a method labeled as Multistage Indexing of Subject Headings (MISH) based on the Keyphrase Extraction Algorithm (KEA) was employed to provide all three types of keywords for all articles from the sample. The articles were sampled from SCIndeks: The Serbian citation index. Three separate maps of (local, peripheral) science were constructed, each based on a different type of keywords. The Partitioning Around Medoid method (PAM) for cluster analyses, followed by multi-dimensional scaling for visual representations of extracted clusters, was employed. Re-

sults suggest that the three types of keywords generate relatively similar maps, encouraging keywords aggregation for practical purposes. Some differences among the maps are not fully consistent with the predictions derived from the model. They reveal some methodological deficiencies of the study and indicate the most promising directions for further research.

Keywords: citations; titles; models; science; maps; indexing; articles; abstracts; references; clusters

Introduction

Description of research papers by keywords extracted from their titles and abstracts is a rather standard procedure to ensure information retrieval ([Medelyan & Witten, 2008](#)). Although most widespread, this procedure is heavily criticized for numerous imperfections. Keywords generated by authors are repeatedly found to be too diverging, those generated by experts too expensive, while the ones extracted by non-human (intelligent) agents too simple-minded to reveal the implicit contents of papers.

The criticism in question stimulated research looking for alternative solutions. This resulted in the emergence of some innovative approaches, the most important being citations-based (related records), but efforts to improve text-based codification of research papers proceeded. Research along this line, that has continued to these days, was focused in its early phases on the automatic extraction of keywords from articles' references. The method was implemented in ISI (now Thomson) citation databases under the trademark KeyWords Plus® ([Garfield & Sher, 1993](#)). The interest was recently extended to include the utilization of metadata from citing papers. The primary interest in experiments using this approach varied from identifying papers relevant to a research topic of interest ([Qazvinian & Radev, 2008](#)) to the summarization of the results of cluster analysis ([Chen et al., 2010](#)). Studies on “reference direct indexing” ([Bradshaw, 2003](#); [Ritchie, Teufel & Robertson, 2008](#)), “citation context” ([Elkiss et al., 2008](#)), and “citation-based automatic indexing” ([Mahdi & Joorabchi, 2010](#)) showed that using information from citing documents can substantially improve the full-text indexing and, consequently, the searching of scientific literature.

By capitalizing on the above ideas and findings, we developed a model of multi-perspective articles description (MPAD; Šipka et al., 2011). MPAD is based on the assumption that a single piece of research has to be observed and described “in motion”, starting with the previous research from which it arose and ending with the subsequent research it triggered. Accordingly, keywords describing an article as a report of a single piece of research should be generated not only from this article’s metadata (title and abstract) or full text, but also (a) from the titles of all its references and (b) metadata belonging to all papers citing this particular article. MPAD can be illustrated by a logical framework matrix reflecting various phases of the research cycle (as different perspectives of viewing the targeted research) within various paradigms existing in the philosophy of science and research methodology, including their genuine parallel from the economy of industry (Table 1).

Table 1. Model of multi-perspective article description

		paradigms				
		economic concepts	bibliometric metaphors	substrate location	keywords extraction sources	keywords types
cycle phases	raw material		intellectual base	previous studies	titles of article references	KW-Rs
	value added		research front	present study	article titles and abstracts	KWs
	product		research heritage	subsequent studies	citing articles' titles and abstracts	KW-Cs

MPAD was implemented in a CEON/CEES web application supporting journal publishing (ASEESTANT: The SouthEast European Journal Production Assistant, <http://aseestant.ceon.rs>). ASEESTANT is a journal management system created primarily to serve the idea of “quality enforcement”, in order to bring journal publishing in the region to the level of world reputed periodicals. Enforcement of the quality is to be ensured, among else, by KwASS (Keywords Assignment Support System), a module designed to produce three kinds of keywords posited by the MPAD model. All three types of keywords are generated by the same algorithm.

Some restrictions make MPAD difficult to apply in practical solutions. A necessary condition for its implementation is the availability of titles of both references and citations. This practically limits the use of MPAD to journal papers indexed in the same citation index or citation-enhanced database. The additional problems come from the fact that the number of citations to an article is virtually infinite. As a consequence, an upper limit for the number of keywords isolated from citing articles had to be stated, leaving at the same time the possibility for the permanent accumulation of new KW-Cs in an “article profile”. This was to be managed through an automated software agent.

In building MPAD into ASEESTANT, we surmounted these and similar difficulties. Yet, many methodological problems remained. Within this study we focused on the relationship among the three types of MPAD keywords. Two interrelated questions were sought to be answered: (1) are different types of keywords summative, i.e. is it appropriate to aggregate them into a common profile for individual articles; and (2) are differences among the keyword types, if existent, interpretable, i.e. consistent with predictions drawn from MPAD.

Answers to the above questions were sought by comparing the bibliographic maps of science produced by the three different types of MPAD keywords. The observed basic similarity of maps was supposed to be an argument in favor of their summative nature, while their observed differences were expected to be in accordance with the MPAD description of the three stages of the research cycle: intellectual base, research front, and research heritage.

Method

Indexing method

Automatic generation of keywords in ASEESTANT is based on KEA: Keyphrase Extraction Algorithm ([Medelyan & Witten, 2008](#)). KEA initially identifies candidates for keyphrases by extracting n-grams from text. In the second phase, the algorithm uses machine learning techniques in order to select the most important candidates, by considering their attributes, such as frequency, inverse document frequency, position in the

text and number of words in a phrase. KEA offers possibilities for free keyphrase extraction and indexing using controlled vocabularies. However, several successive trials during the development process revealed the shortcomings of both methods. Free indexing was found to produce too large a number of divergent terms, many of which were grammatically incorrect phrases. On the other hand, assignment based on controlled vocabularies tended to isolate too small a number of terms, which in most cases turned out to be too general and imprecise.

In order to improve KEA performance, in the final stage of ASEESTANT development a new method was developed. The method, dubbed Multi-stage Indexing of Subject Headings (MISH), employs KEA's free indexing algorithm as a first step to extract candidates for descriptors. The extracted terms are then matched against terms from some of the pre-selected international domain-specific thesauri and from two home-made CEON/CEES vocabularies. Matched multi-word terms are finally stemmed in order to collapse those with the same meaning into a single entry to avoid duplication.

The above procedure was applied to produce all three types of MPAD keywords. Only papers having titles, abstracts, and at least one reference in English could be included in the analysis, since the thesauri and vocabularies employed are all monolingual.

Data sample

The sample of papers included 13,032 papers indexed with all three types of automatically generated descriptors. All papers were taken from SCIndeks: The Serbian citation Index ([Šipka, 2005](#); <http://ceon.rs>). Due to the above stated language restrictions, a majority of them came from journals covering chemistry (19%), agriculture (18%), and medicine. The sample of terms included 96,766 KWs (7.4 per paper on average), 96,537 KW-Rs (7.4 per paper), and 154,305 KW-Cs (11.8 per paper). The analysis included only KWs having a frequency higher than 17 (1,006 terms), KW-Rs with a frequency higher than 14 (1,279 terms) and KW-Cs with a frequency higher than 27 (1,030 terms).

Data analysis

Maps were generated using a combination of techniques which have proved to be appropriate for the representation of a large amount of entities ([Moya-Anegón, Herrero-Solana & Jiménez-Contreras, 2006](#)). Distance matrices calculated using the Jaccard index were created for each sample of words separately. Cluster analyses were applied on these distance matrices, with the Partitioning Around Medoids method (PAM). Silhouette and connectivity measures were employed to estimate the number of clusters ([Kaufman & Rousseeuw, 1990](#)). Clusters of terms were visually represented in two-dimensional space using multidimensional scaling (MDS). Distances among clusters on the map reflect the average distances of terms within clusters which are their best representatives. The size of the circles representing clusters is proportional to the number of terms in each of them.

Results

Number of clusters

The validity of cluster solutions estimated by the silhouette and connectivity measures is shown in Figures [1](#) and [2](#). Better solutions are indicated by higher silhouette and lower connectivity values. The results suggest that the clusters are of low quality and unclear structure. The solution giving 9 clusters in mapping KWs and KW-Cs seems to be the most appropriate. For KW-Rs, solutions resulting with 9 to 11 clusters appeared to be more interpretable than others. Among these, the ten-clusters solution was chosen for the main analysis.

Maps of science

The bibliographic maps are presented in chronological order, from the map of cited papers' keywords (KW-Rs; [Figure 3](#)) to the map based on keywords extracted from citing papers (KW-Cs; [Figure 5](#)). The maps are generally similar both in terms of content and the structure of clusters. Natural sciences are basically reduced to two areas most heavily represented in the sample: agriculture and chemistry. Judged by the number of clusters, agriculture seems to be the most differentiated research area. Similar configurations with the predominance of agriculture and chemis-

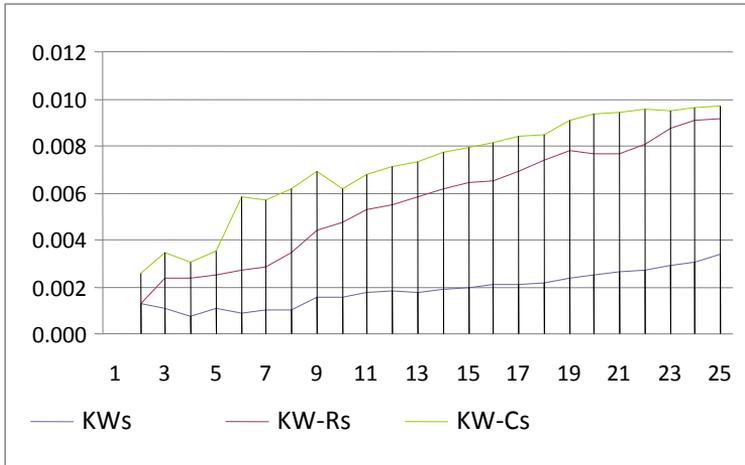


Figure 1. Silhouette for diff. number of clusters

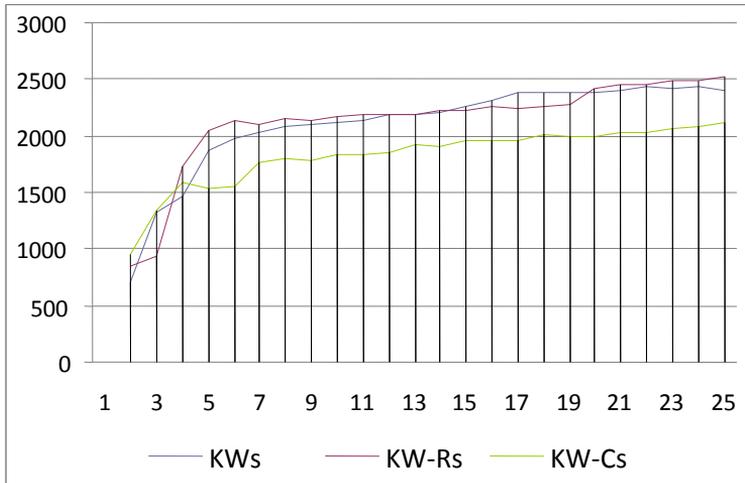


Figure 2. Connectivity for diff. number of clusters

try were already observed in the profiles of other developing countries ([Schultz & Manganote, 2012](#)). Another similarity comes from the close positions of social sciences and medicine in all three maps. Finally, all maps have failed to show some disciplines, such as mathematics, computer sciences, and earth sciences, which can normally be found on large

scale world-science maps (Boyack, Klavans & Börner, 2005). It should be noted, however, that the KW-Rs map corresponds slightly more closely to global maps of science, thanks to the better differentiation of basic natural sciences.

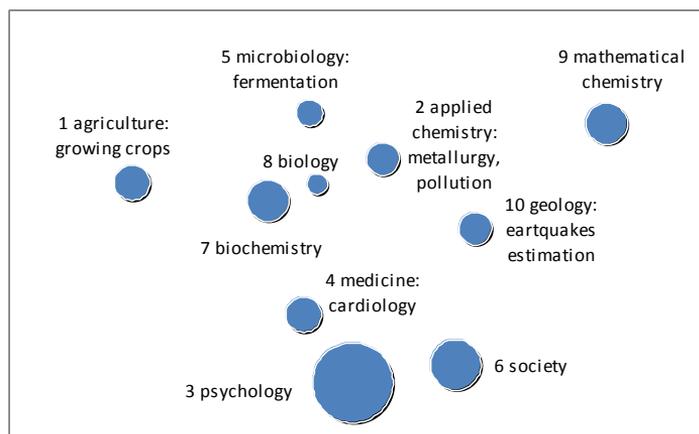


Figure 3. Map of science described by KW-Rs

Regarding the content of clusters, maps based on keywords describing the papers themselves (KW; [Figure 4](#)) and citing papers' keywords (KW-Cs; [Figure 5](#)) look more similar. Both these types of keywords were isolated from SCIndeks papers, published mainly by domestic, Serbian authors who share the same orientation towards findings application. In such papers more uniform and less technical terminology, typical of peripheral science, is predominant. On the other hand, the titles of cited papers condensed into KW-Rs come mainly from international publications offering a greater diversity of topics and richer vocabulary. Apart from better outlining some applied disciplines like metallurgy and earthquake estimation, the cited references (KW-Rs) map is more convincing in featuring basic disciplines, especially biology and microbiology. As for social sciences, clusters in this map have a more general and fundamental profile instead of an applied one.

The expectation based on previous studies showing that cited references reflect intellectual base seems to be at least partially confirmed. However,

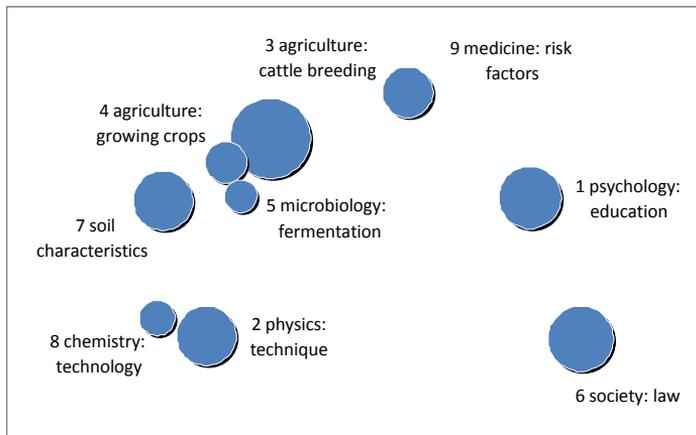


Figure 4. Map of science described by KWs

the results failed to fully support the hypothesis about keywords (KW) being good descriptors of research presence (“state-of-the art”), and KW-Cs being descriptors of research results’ consequences (“research heritage”). The two cluster analyses resulted in structures too similar to reveal interpretable differences between these two stages of the research cycle. This can be attributed to the rather imitative character of research presented in SCIndeks, as well as to the lack of interdisciplinarity of research in science periphery. Apparently, the findings are too rarely applied outside their narrow disciplinary area to be recognized, when transformed into KW-Cs, as something forming a new, discrete segment of research space.

Differences among the maps in terms of the interpretability of MDS dimensions are more or less in line with the results of the cluster analyses. The KWs map somehow provided more interpretable dimensions than the other two maps. One of the dimensions, plotted on the x-axis, apparently differentiates social sciences from sciences, while the second one contrasts life sciences with the group consisting of chemistry, physics and, interestingly enough, research on society dominated by legal issues.

The basic nature of the dimensions in the other two maps seems to be quite similar. At the same time, some congruent deviations from the KWs map can be observed. The main dimension from this map, reflecting the

science-social sciences dichotomy, appeared as the y-axis on the other two maps, accounting for less variance than previously stated. Also, the second dimension is obviously shaped not only by the genuine differences among the research areas, but also by their bibliometric specificities. Polarization on this dimension seems to have occurred as a result of the activity of two dominant research groups, one of which is extremely large, open, and highly productive (agricultural researchers), while the other is rather closed and self-sufficient (mathematical chemists). The latter group is small and characterized by an extremely high self- and inter-citation rate. In the analysis based on KWs, it was not recognized as a cluster at all. More research is needed to see if its isolation, position on the map, and influence on the whole constellation of research areas is an artifact or the structural characteristic of the research space under study.

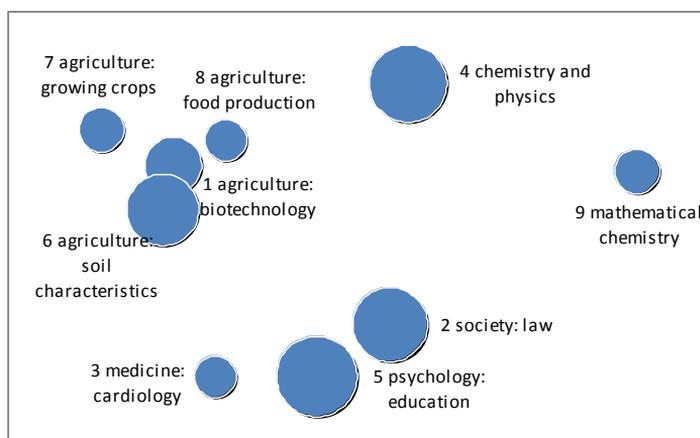


Figure 5. Map of science described by KW-Cs

Conclusions

The three bibliographic maps obtained seem to be sufficiently congruent to allow for the routine aggregation of different types of MPAD keywords, i.e. combining them into unique article profiles. Some observed differences among the maps are not clearly attributable to differences among the three stages of the research cycle, as hypothesized by MPAD and expected on the basis of previous studies about the influence of “his-

torical variation” on the research front (e.g. [Lucio-Arias & Leydesdorff, 2009](#)).

The reasons for the lack of support for MPAD might be quite different. Some hints suggest that they are methodological in nature and most likely related to sampling issues. Future studies should be concentrated on overcoming some obvious deficiencies of the present study:

- (1) Papers published in reputed international journals were not covered for practical reasons. This may have affected the resulting picture of science, which is reduced to its “import-oriented” part, hiding an important share capable of “findings re-export”.
- (2) Papers not having English titles and/or abstracts, most of which lacking English references as well, were not covered for technical reasons. This may have affected results on clustering, and consequently mapping.
- (3) The differences in the number of papers among various disciplines were not controlled. This probably affected the nature of isolated MDA dimensions. In a study aimed at testing a model such as MPAD, the area sampling of papers would have probably been more appropriate.
- (4) The differences in information pregnancy of sources used for keyword extraction was not possible to control. KW-Rs were extracted only from reference titles, while in producing KW-s and KW-Cs article abstracts were used as well.

These and other deficiencies suggest caution in accepting the findings of this study, calling for further research. They, however, do not make the results invalid, and certainly do not make MPAD unfeasible.

References

- ASEESTANT: The SouthEast European Journal Production Assistant [Computer software]. Belgrade: Centre for Evaluation in Education and Science. Retrieved from <http://aseestant.ceon.rs>.
- Bradshaw, S. (2003). Reference directed indexing: Redeeming relevance for subject search in citation indexes. In: *Proceedings of the 7th European Conference on Research and Advanced Technology for Digital Libraries*

- (ECDL) (pp. 499-510). Trondheim, Norway: Springer. Retrieved from http://link.springer.com/chapter/10.1007%2F978-3-540-45175-4_45
- Boyack, K. W., Klavans, R. & Börner, K. (2005). Mapping the backbone of science. *Scientometrics*, 64(3), 351-374. doi:10.1007/s11192-005-0255-6.
- Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010). The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. *Journal of the American Society for Information Science and Technology*, 61(7), 1386–1409. doi:10.1002/asi.21309.
- Elkiss, A., Shen, S., Fader, A., Erkan, G., States, D. J., & Radev, D. R. (2008). Blind men and elephants: What do citation summaries tell us about a research article? *Journal of the American Society for Information Science and Technology*, 59(1), 51–62. doi:10.1002/asi.20707.
- Garfield, E. & Sher, I. H. (1993). KeyWords Plus – Algorithmic derivative indexing. *Journal of the American Society for Information Science*, 44(5), 298-299. doi:10.1002/(SICI)1097-4571(199306)44:5<298::AID-ASI5>3.0.CO;2-A.
- Kaufman, L. & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. New York: Wiley.
- Lucio-Arias, D., & Leydesdorff, L. (2009a). An indicator of research front activity: Measuring intellectual organization as uncertainty reduction in document sets. *Journal of the American Society for information Science and Technology*, 60(12), 2488-2498. doi:10.1002/asi.21199.
- Mahdi, A. E. & Joorabchi, A. (2010). A Citation-based approach to automatic topical indexing of scientific literature. *Journal of Information Science*, 36(6), 798–811. doi:10.1177/0165551510388080.
- Medelyan, O. & Witten, I. H. (2008). Domain independent keyphrase indexing with small training sets. *Journal of the American Society for Information Science and Technology*, 59(7), pp.1026-1040. doi:10.1002/asi.20790
- Moya-Anegón, F., Herrero-Solana, V. & Jiménez-Contreras, E. (2006). A connectionist and multivariate approach to science maps: the SOM, clustering and MDS applied to library and information science research. *Journal of Information Science*, 32(1), 63-77. doi:10.1177/0165551506059226.
- Qazvinian, V. & Radev, D. R. (2008). Scientific paper summarization using citation summary networks. In: *Proceedings of the 22nd International Conference on Computational Linguistics COLING 2008*, Manchester, August 2008 (pp. 689–696). Stroudsburg, PA, USA: Association for Com-

- putational Linguistics. Retrieved from <http://www-personal.umich.edu/~vahed/papers/citsum.pdf>
- Ritchie, A., Teufel, S. & Robertson, S. (2008). Using terms from citations for IR: some first results. In: C. Macdonald, I. Ounis, V. Plachouras, I. Ruthven & R.W. White (Eds.), *Advances in Information Retrieval* (pp. 211–221). Heidelberg: Springer. Retrieved from http://link.springer.com/chapter/10.1007%2F978-3-540-78646-7_21
- Schultz, P.A., & Manganote, E.J.T. (2012). Revisiting country research profiles: Learning about the scientific cultures. *Scientometrics*, online 13 March 2012, doi:10.1007/s11192-012-0696-7.
- Šipka, P. (2005). The Serbian Citation Index: Context and content. In: *Proc. ISSI 2005 – 10th International Conference of the Society for Scientometrics and Informetrics* (pp.710-711). Stockholm, Sweden: Karolinska Univ. Press. Retrieved from http://ceon.rs/pdf/Sipka_SCIndeks_proceedings.pdf
- Šipka, P., Sotirović, M., Pajić, D., Jevremov, T. & Stanić, N. (2011). *Četvorodimenziona vizualizacija [Four-dimensional visualization]*, Unpublished application for the "Best technological innovations 2011", Belgrade: Centre for Evaluation in Education and Science.

Supporting Open Access Nationwide

Jadranka Stojanovski

jadranka.stojanovski@irb.hr

University of Zadar, LIS Dept, Ruđer Bošković Institute, Library, F. Tuđmana 24i,
HR-23000 Zadar, Bijenička c. 54, HR-10000 Zagreb (Croatia)

Abstract

To support the Croatian scholarly publishing environment, and inspired by the global open access movement, the portal of Croatian scientific journals HRČAK (<http://hrcak.srce.hr>) was introduced in 2006 offering an open access publishing platform for Croatian journals. Today, HRČAK gathers about 290 scholarly and professional Croatian journals. This paper is focused on the currency and visibility of the journals included in HRČAK, giving accurate statistical data about the HRČAK repository, its growth and development. Collaboration with Croatian publishers – namely, they are mainly academic and research institutions or professional societies – on the continuous work of raising the quality of Croatian scientific journals is presented in this paper. HRČAK journals are available for harvesting using the OAI-PMH protocol and papers are distributed through many different repositories, archives, databases and search engines. Future plans include work on full-text documents, inclusion of the additional types of publications and formats, harvesting process improvements, additional functionalities and standardization.

Keywords: Croatia; Open Access; publishing; scientific journals; Gold Open Access

Introduction

In the era of ACTA¹, PIPA², SOPA³, RWA⁴ and other threats to the open access movement, it is important to recognize specific practices which are present in the scholarly publishing environment, and to be aware of the present changes and future trends. Today's research is interdisciplinary and collaborative, requiring fast and efficient information and data ex-

¹ *Anti-Counterfeiting Trade Agreement*

² *Protect IP Act*

³ *Stop Online Piracy Act*

⁴ *Research Works Act*

change, as well as full and open access to published research. However, the predominant part of scholarly publications is still in the hands of publishers with profit oriented business models. Although today the average researcher at the university or research institute has access to a much broader range of scholarly literature than before, in the print era, the open access (OA) movement promoting access to articles without any restrictions posed by subscriptions is more important than ever ([Björk et al., 2010](#)). OA journals are more visible and readable and that is why a comparison of OA and NOA articles showed that OA articles consistently have more citations, varying from 36%-172% by discipline and year ([Hajjem, Harnad, & Gingras, 2005](#)).

The most important step in bringing research from small and developing scientific communities to mainstream science was the inclusion of journals in international bibliographic databases and making journals available in free full text on the Internet ([Marušić, Sambunjak, & Marušić, 2006](#), page 372). It is also a widely accepted opinion that publicly funded research data should be openly available. This statement is supported in the official science & technology policy issued by the Croatian Ministry of Science, Education and Sports (MSES), where it is specified that the results of research “...financed by public resources must be accessible to the public in the form of open publications or databases” ([Stojanovski, Petrak, & Macan, 2009](#)), and it is taken as the base for the promotion of the OA movement in Croatia.

The transformation of Croatian scientific journals from print to digital started in the late nineties and at a very slow pace. In 2002 there were only 21 Croatian scientific journal with online versions and full-articles available ([Konjević, 2003](#)). Publishers were not ready for evolutionary changes in the publication process for many reasons: lack of awareness about existing trends, poor ICT equipment, low level of computer literacy, lack of technical support, etc. To support the Croatian scholarly publishing environment, and inspired by the global open access movement, the portal of Croatian scientific journals HRČAK (<http://hrcak.srce.hr>) was introduced in 2006 offering an OA publishing platform for Croatian journals. The abbreviation HRČAK (in English: hamster) comes from the phrase “HRvatski ČAsopisi” (in English: Croatian Journals). HRČAK

was conceptualized and designed by a small group of librarians, researchers and information specialists, and created in cooperation with programmers at the University Computing Centre SRCE, with MSES support. The primary intention of the HRČAK team was to offer a free building tool for digital versions of the Croatian journals, and from the very beginning open access was promoted as the most appropriate model for scholarly publishing in Croatia.

There are two widely accepted routes to OA: digital repositories or archives (“green”) and OA journals (“gold”), often considered as two parallel tracks. “Gold” journals are available to their readers free of charge, and “green” uses online repositories to store digital duplicates of published research output and make them freely and openly available across the Internet ([Hubbard, 2011](#)). “Gold” applies only to publication through publishers, while “green” relates to self-archiving. Although this classification often caused confusion ([Murray-Rust, 2011](#)), it is widely adopted in the OA community. However, it has recently become apparent that there is potential for repositories and journals to interact with each other on an ongoing basis and to form a coherent OA scholarly communication system between them ([Pinfield, 2009](#)). In Croatia, the majority of journals followed the “golden” open access road in providing digital journals, including freely available peer-reviewed articles, and the HRČAK version of an online journal is the only digital version of the journal. However, some journals have their online version independent from the HRČAK platform, and editors submit articles in HRČAK as additional copies, so this part can be considered as “green”, even though it is not submitted by the author.

The main goal of HRČAK was to provide a simple and free interface for creating digital online versions of Croatian journals, technical support for publishers, maintenance and long-term data protection and management. With time, the visibility of Croatian scientific journals was improved, and research in Croatia in general became one of the priorities. Visibility can be measured by different criteria: number of visits, usage of the Open Access Initiative – Protocol for Metadata Harvesting (OAI-PMH) for sharing data, data traffic, replication of bibliographic content by online databases, number of citations, etc. The research question in this paper is to what ex-

tent the analysis of different data sets can reflect the visibility of the HRČAK portal, including journals and their content.

Data and methods

In this paper the visibility of the journals included in HRČAK is analysed and discussed, giving accurate statistical data about the HRČAK repository, its growth and development. The descriptive data analysis and interpretation method was used in this study, implying a simple quantitative summary of data sets that have been collected from different sources. The data about journals, their size, and the number of full-text articles stored were collected from the HRČAK database. The data about traffic, visits, and harvesters were extracted from Web server log files using Webalizer, a program that looks at Web server logs and produces detailed reports accessible online. The data about harvesters using the OAI-PMH protocol were analysed separately from the usage statistics representing visits to regular Web pages at <http://hrcak.srce.hr>, from which usage of the editors and staff was excluded. Scopus (Elsevier), Web of Science – WoS (Thomson Reuters) and Journal Citation Reports – JCR (Thomson Reuters) databases were consulted to get the data about indexed journals. All data were gathered and analysed in the period March-May 2012.

Results and discussion

There are about 220 scientific and professional journals in Croatia that are publicly funded and mostly published by academic and research institutions or professional societies, half of which come from social sciences and humanities. According to a 2007 study, the median grant for journals by MSES was 46.000 HRK, e.g. approximately 6.200 EUR ([Macan & Stojanovski, 2008](#)). To get a grant from MSES, publishers are evaluated by several criteria: publishers should be not-for-profit organisations, the journal should be issued regularly, the articles should pass a peer review process, the journal content should be more focused on original scientific papers, a certain number of journal papers should have a corresponding author from a foreign country, the journal should be in OA, the journal should be indexed in relevant international bibliographic databases, publishers should use a journal management and publishing system, the journal should be edited according to the *Guidelines for editing and*

*formatting journals*⁵, the journals should have cultural, national and local importance for the scientific community, etc.⁶

The number of journals included in HRČAK increased steadily from 2006 till 2012, and in May 2012 HRČAK gathered 292 scholarly and professional Croatian journals. The distribution of the journals by discipline shows that the majority of journals come from humanities and social sciences (33 percent and 28 percent accordingly), which corresponds to the number of journals supported by MSES in these disciplines (Stojanovski et al., 2009). Social sciences and humanities journals are closely oriented to subjects related to Croatian language, culture and history. Other disciplines are represented by 7 to 11 percent (Figure 1), also following the proportion of the total number of journals published in these disciplines. Croatian researchers working in sciences, biomedicine and health, technical sciences and biotechnical sciences also publish in international journals, and that is why there is no need for a larger number of domestic journals in those scientific fields. Also, the advancement and evaluation criteria particularly motivate science researchers to publish in more prominent international journals.

HRČAK's growth was impressive, going far beyond the team's expectations. The majority of publishers immediately accepted the HRČAK portal after the introduction of the OA approach, and also the visible data about journal/issue/paper visits and downloads by HRČAK played a significant role in their further self-promotion. Very soon after the first release, without any systematic promotional activities, the HRČAK portal became the most popular place for accessing scholarly content. There are many reasons for the success of HRČAK, and some of them are (not ordered by relevance):

- (1) readers are more likely to access scholarly content through a single portal instead of trying to find and select the content of interest from many different websites;

⁵ <http://public.mzos.hr/Default.aspx?art=9110&sec=3202>

⁶ Guidelines about financial support of scholarly publishing for 2010 issued by the Croatian Ministry of Science, Education and Sport.

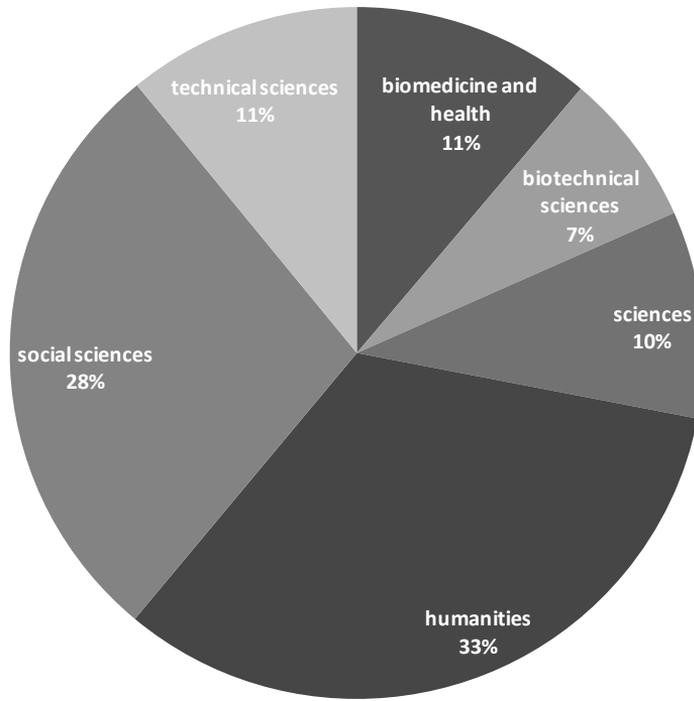


Figure 1. Distribution of journals included in the HRČAK portal by discipline

- (2) free infrastructure and support from the HRČAK team were encouraging for many Croatian publishers unable to publish digital version of their journals using their own resources;
- (3) credibility and competences of the HRČAK team and the University Computing Centre SRCE helped form a trustworthy relationship between Croatian publishers and the HRČAK team from the very beginning;
- (4) MSES, which supports a majority of the journals, shared the HRČAK's main goals, such as online environment, open access, increased visibility, user-centred approach, and advancements in editorial work;

- (5) HRČAK platform is used for collaboration and cooperation, establishing a dialog and empowering small publishers' positions;
- (6) regular statistical reports with data about repository growth, number of visits, data traffic, etc.

In May 2011, for 292 included journals there were 74,404 full-text articles stored in the HRČAK repository. The size of the repository is 89.2 GB: 85 GB for the file system and 4.2 GB for the database. Some publishers provided the content of their journals since volume 1, using HRČAK as the place to archive their complete collections.

Data traffic in 2011 according to Webalizer was 5.94 TB in total, with an average traffic of 465 GB per month ([Table 1](#)), and traffic growth is following content growth. HRČAK is in the top ten world repositories, according to the Ranking of Web of World Repositories⁷. Data about traffic done by OAI harvesters are excluded from the statistics.

Table 1. Traffic in the period 2006-2011

<i>year</i>	<i>traffic (GB)</i>
2006	116.48
2007	1259.89
2008	2307.23
2009	3733.34
2010	5589.74
2011	5953.98

In the time period 2006-2011, the top number of visits was in 2010 ([Figure 2](#)). During 2011, 5,421,200 visits were recorded, with an average of 451,767 visits per month, where a visit is defined as any number of requests from the same IP address processed during a period of 30 minutes. The server processed a total of 51,011,881 requests (hits) during 2011, and the busiest months were October and November ([Figure 3](#)).

⁷ <http://repositories.webometrics.info/>

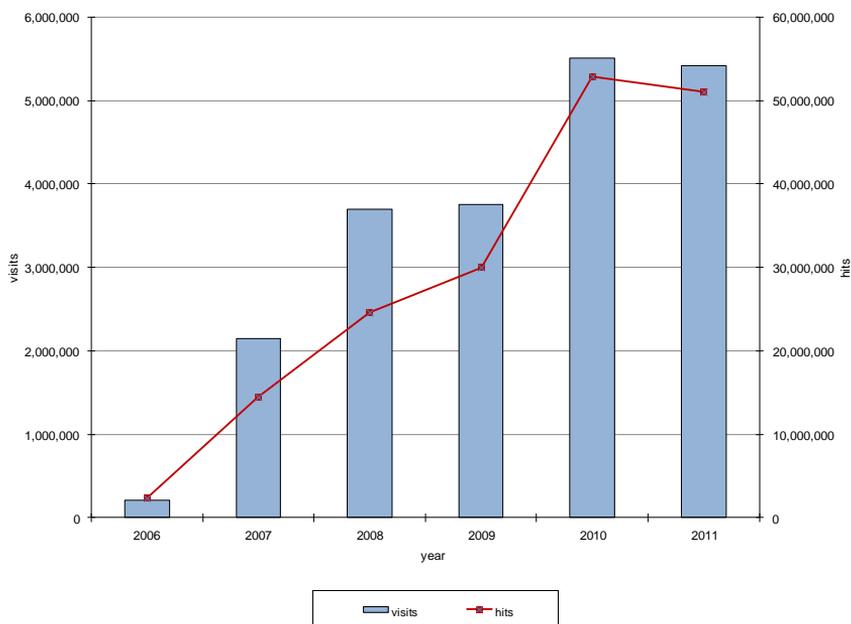


Figure 2. Visits and hits by year 2006-2011

HRČAK journals are available for harvesting using the OAI-PMH protocol⁸ (RepositoryIdentifier: hrcak.hr; SampleIdentifier: oai:hrcak.hr:25), and papers are distributed through many different repositories, archives, databases and search engines. In 2011 the most active harvesters with resolved IP addresses were: Google Scholar, IFREMER, Yahoo! Research, Sakura Internet (Japan), Long Island University, OCLC, Bielefeld University, Scientific Commons Project, QSensei, Purdue University Libraries, Università degli Studi di Firenze, Physikalisch-Technische Bundesanstalt, University of Southampton, School of Electronics and Computer Science, and Staatsbibliothek zu Berlin.

Harvesters visited HRČAK on average 2508 times per month in 2011, and 3543 times per month during the first four months in 2012, which is a significant increase. The majority of harvesters are located inside the US commercial domain, Japan, France, US, Croatia, South Africa, Switzerland,

⁸ <http://hrcak.srce.hr/oai>

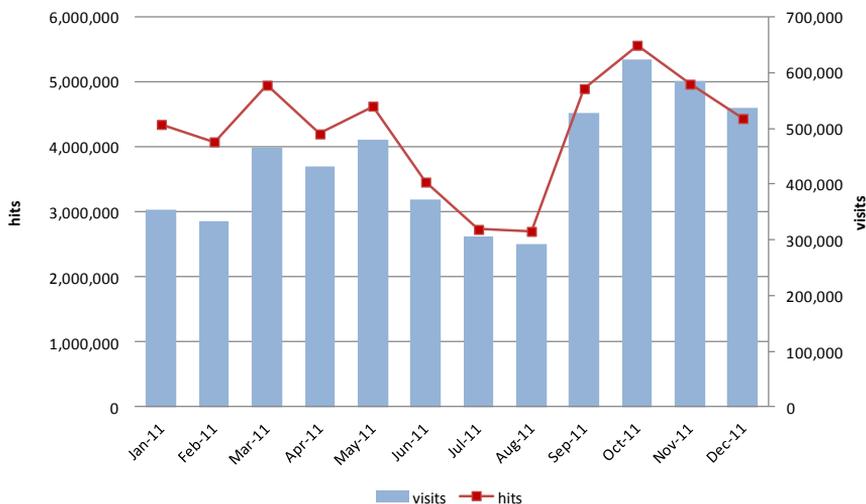


Figure 3. Visits and hits by month for 2011

Bosnia and Herzegovina and Germany. A lot of IP addresses remain unresolved, and even for the known domains it was difficult to reveal in which repository, database, archive or catalogue the harvested data were included. Important information about a crawler, shared during “crawler introduction”, can improve the data exchange and should become a standard.

In addition to disseminating data about Croatian scientific and professional journals and published papers through OAI-PMH, the HRČAK team is involved in the process of the inclusion of Croatian journals in the most prominent fee and free databases like Web of Science (WoS), Scopus, Google Scholar, EBSCO databases, DOAJ, Google Scholar, OAIster, BASE etc. Although HRČAK is an OA portal, exchanging its data with OA databases, archives, repositories and other digital collections primarily, and inclusion in popular, even commercial products, significantly improves the visibility of published papers. For articles indexed in citation databases the number of citations can be recorded, which can be used for evaluation and advancement purposes. In 2007 there were only 16 Croatian journals included in WoS ([Macan & Stojanovski, 2008](#)). To satisfy the inclusion criteria, publishers usually need to improve some elements of the publishing process and to follow professional standards for editing

more strictly. HRČAK as a collaboration tool supports publishers during this process in different ways. The HRČAK team provides Croatian publishers with contacts, links to selection/inclusion criteria, and help with their applications. These efforts, together with WoS's expansion of their content coverage, resulted in more than 60 Croatian journals (54 HRČAK) being indexed in the Web of Science in 2011. The Scopus database, launched in 2004, systematically included Croatian journals from the beginning, and is currently indexing 105 Croatian journals (90 HRČAK). From 292 journals included in HRČAK, 31.5 percent (92 titles) are indexed in WoS or Scopus databases: 48 journals are indexed in both databases, 6 journals are indexed only in WoS, and 42 journals are indexed only in Scopus. HRČAK journals included in WoS are distributed evenly across disciplines: natural sciences (11), biomedicine (8), technical sciences (11), biotechnology (5), social sciences (9), and humanities (10). According to JCR, there are 35 Croatian journals with an impact factor (IF), and only 4 of them are non-HRČAK journals. The best Croatian journals according to IF are included in HRČAK, and the first on the ranked list is the *Croatian Medical Journal* (IF 1.455). Two non-HRČAK journals with a relatively high IF are *Operators and Matrices* (IF 0.542) and *Mathematical Inequalities and Applications* (IF 0.524). Both journals have their web sites and articles can be accessed by pay-per-view/download (an advance payment of 10 EUR per article is charged) or by login for journal subscribers. The other two non-HRČAK journals, *Transactions of FAMENA* (IF 0.143) and *Neurologia Croatica* (IF 0.125), have web pages with general information about the journal and published articles, but full-text articles are available only by print subscription. Other Croatian journals included in WoS are still not listed in JCR, since a minimum of two year of presence in WoS is needed for IF calculation. In cooperation with database publishers and aggregators, the HRČAK team is working together with their technical teams on data harvesting using the OAI-PMH protocol instead of the traditional ways of data gathering. Croatian publishers are using HRČAK's mailing list to exchange their experiences, which was recognized as a good motivation tool. The popular search engines and databases that use OAI-PMH on a regular basis are Google Scholar (GS), where all HRČAK journals and papers are indexed and searchable, OAIster (OCLC) and BASE (Bielefeld).

To support the whole publishing process, an open source journal management system OJS – Open Journals System⁹, was introduced in 2008 to the Croatian publishers, and data synchronisation with the HRČAK repository has been established. For publishers without their own servers, there is a central version of OJS ver. 2.2.2.0 installed at SRCE¹⁰, and also local OJS installations are supported. At present, about twenty publishers are using OJS for the automation of their publishing process, from paper submission and peer review to the selection of papers and online journal publishing. Although OJS has been designed to help enhance the high quality standard for journal publishing, reduce cost, and provide efficient editorial processes (Willinski J. according [Kwaku Smit Esseh, 2006](#)), for a majority of Croatian publishers it encompasses the redefinition of the existing editorial process. This could be the reason for the relatively small number of publisher who accepted OJS as a tool for journal management. To educate, share opinions and to meet editors' requirements, workshops and round tables for discussions about current topics are regularly organized.

Conclusions

The analysis in this paper revealed growth in all aspects of the HRČAK portal: number of included journals, content, usage statistics, and exchange of data, proving that the OA movement has found fertile ground inside the small Croatian publishing environment. One of the reasons for HRČAK's big success is the coherence with MSES criteria for journal grants, including the promotion of OA, the usage of journal management systems and the improved visibility of scholarly publishing in Croatia through the exchange of data and journal indexing in popular databases. Editors and authors can also track data about the number of visits and downloads – available on the journal, issue and article level, thus measuring their popularity.

There is still a lot of space for improvements, especially regarding interoperability and data exchange with other national and international information services. Future plans include work on existing full-text documents which are at the present only partially searchable, inclusion of

⁹ <http://pkp.sfu.ca/ojs>

¹⁰ <http://hrcak.srce.hr/ojs>

the additional types of publications, formats and research data gathering, harvesting process improvements, additional search functionalities, articles in press, website redesign and more standardization on different levels. Metadata description can also be improved. Although scientific journals in Croatia are recently suffering from serious budget cuts, this could also be considered as a new challenge leading to further modernisation and improvements, and leaving less space for unnecessary expenditures related to traditional publishing processes. The collaboration between publishers could still be improved, and more educational workshops and meetings should be organized. There is also a strong will to broaden HRČAK's coverage with conference proceedings and to include OA journals from other, neighbouring countries.

Acknowledgements

I wish to thank colleagues at the University Computing Centre SRCE and their leader Miroslav Milinović for providing the data about traffic and usage statistics.

References

- Björk, B., Welling, P., Laakso, M., Majlender, P., Hedlund, T., Guðnason, G., & Scalas, E. (2010). Open Access to the Scientific Journal Literature: Situation 2009. *PLoS One*, 5(6), [doi:10.1371/journal.pone.0011273](https://doi.org/10.1371/journal.pone.0011273).
- Hajjem, C., Harnad, S., & Gingras, Y. (2005). Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact. *IEEE Data Engineering Bulletin*, 28(4), 39-47. Retrieved from <http://eprints.soton.ac.uk/261688/1/ArticleIEEE.pdf>
- Hubbard, B. (2011). Green, Blue, Yellow, White & Gold. Retrieved from <http://www.sherpa.ac.uk/documents/sherpaplusdocs/Nottingham-colour-guide.pdf>.
- Konjević, S. (2003). Hrvatski znanstveni časopisi na Internetu [Croatian scientific journals on the Internet]. *Vijesnik bibliotekara Hrvatske*, 4, 111-118. Retrieved from <http://eprints.rclis.org/bitstream/10760/5598/1/Konjevic.pdf>.
- Smit, E. K. S. (2006). The open access (Open Journal Systems) paradigm and the production of scholarly journals in developing countries. Retrieved from <http://summit.sfu.ca/system/files/iritems1/2371/etd2183.pdf>.

- Macan, B., & Stojanovski, J. (2008). Analiza novčane potpore Ministarstva znanosti, obrazovanja i športa hrvatskim znanstvenim časopisima [Analysis of the financial support of the Ministry of Science, Education and Sports to the Croatian scientific journals]. *Kemija u industriji*, 57(3), 115-122. Retrieved from <http://hrcak.srce.hr/file/32964>.
- Marušić, M., Sambunjak, D., & Marušić, A. (2006). Life of small medical journal: How bibliographical indexing and international visibility affected editorial work in Croatian Medical Journal. *Croat Med J*, 47(3), 372-5. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2080425&tool=pmcentrez&rendertype=abstract>.
- Murray-Rust, P. (2011). Green and gold open access?, Libre and gratis. Reasons why readers and re-users matter. Retrieved from <http://blogs.ch.cam.ac.uk/pmr/2011/07/19/green-and-gold-open-access-libre-and-gratis-reasons-why-readers-and-re-users-matter/>, 2012 Apr 12.
- Pinfield, S. (2009). Journals and repositories: An evolving relationship. *Learned Publishing*, 22(3), 165-175. doi:10.1087/2009302.
- Stojanovski, J., Petrak, J., & Macan, B. (2009). The Croatian national open access journal platform. *Learned Publishing*, 22(4), 263-273. doi:10.1087/20090402.

Does the Open Access Business Model Have a Significant Impact on the Citation of Publications? Case Study in the Field of Civil Engineering

Teja Koler-Povh¹, Goran Turk² and Primož Južnič³

¹ teja.povh@fgg.uni-lj.si

University of Ljubljana, Faculty of Civil and Geodetic Engineering – Library, Jamova c. 2, SI-1000 Ljubljana (Slovenia)

² goran.turk@fgg.uni-lj.si

University of Ljubljana, Faculty of Civil and Geodetic Engineering, Jamova c. 2, SI-1000 Ljubljana (Slovenia)

³ primoz.juznic@ff.uni-lj.si

Faculty of Arts, Department of Library and Information Science and Book Science, Aškerčeva c. 2, SI-1000 Ljubljana (Slovenia)

Abstract

One of the possible benefits of open access (OA) might be the better visibility of articles, which is usually measured by the number of citations of the article. In order to realistically estimate the effect of OA on citation, it is not enough to compare OA and non-OA ISI journals. Thus, as [Harnad and Brody \(2004\)](#) suggested, the numbers of citations of OA and non-OA articles from the same journals were compared. Therefore, we have chosen to analyze the publications in three international journals in the field of civil engineering. All of them have an ISI impact factor in the Civil engineering subject category in the ISI/Web of science database (WOS). The articles were classified into two groups – the OA publications and the non-OA publications. We analyzed all the articles published in the same year and the number of their citations until the end of February 2012, seeking to find out if these two groups differ from each other.

Keywords: Open Access; impact factor; citations; civil engineering; articles; databases; funding; libraries; scientists; WoS; Google

Introduction

Open Access (OA) was developed as a reaction to the increasing prices of scholarly and scientific journals. This was first emphasized by librarians, as libraries were the first to be hit by this development. The movement evolved when scientists themselves started to doubt the existing economic model of journals as the main means of scientific communication and information. The most important boost of OA was when a number of funding bodies, research councils and governmental bodies began transforming their view on the publication of the research they were financing. The most well-known example is a 2008 mandate of the US National Institutes of Health, which requires researchers in receipt of federal funding to deposit their final, peer-reviewed manuscripts in PubMed Central ([NIH Public Access Policy Details, 2008, 2009](#)).

There are two types of OA: the “golden” road (to publish your article in an OA journal) and the “green” road (to publish your article in a non-OA journal, but to also self-archive it in an OA archive). According to a 2008 research ([Harnad et al., 2008](#)), about 10% of journals were gold, but over 90% were already green (i.e. they have given the authors the right to self-archive); however, only a fraction of the articles have been self-archived. Most of the traditional publishers have started to offer a special “open access” option, which involves paying a publication fee for the right to archive the material online. This fee, which is actually an additional one, as subscription is paid as well, can be quite substantial – as far as 3000 € per article. Publication fees or article-processing charges (APC) have become the predominant means for funding professional OA journals and recent findings are suggesting that OA publishing funded through APC is likely to increase furthermore ([Solomon and Björk, 2012](#)).

The main point – as journals are a part of the scientific information and communication system – is how a typical scientist seeks information. Researchers submit their research results in the form of publication after peer-review, with the main idea that other researchers can use and apply their findings. Other researchers must first find the published articles of other researchers and be able to access them. The second step is to find these research results useful, which is exhibited by using and citing them.

Typically, a scientist will begin by using a bibliographic tool, e.g. the Web of Science. In Slovenia, the Slovenian Research Agency evaluates scientific work through publications and citations included in WOS. In Slovenia, where WOS was bought by the Slovenian Research Agency in the early nineties of the last century, all members of Slovenian universities can easily access the WOS database and look up the publications and their citations. In other countries, this may often not be possible, since their libraries might not have subscriptions ([Guédon, 2004](#); [Guédon, 2008](#)). Libraries simply cannot afford to pay a subscription for all valuable information resources, such as journals. The principal two arguments in support of open access publishing rest on the belief that the subscription-based publishing model has produced a crisis of accessibility to scientific literature and that research article access is a problem ([Harnad et al., 2008](#)). Availability of the article is not a sufficient condition for citation, but a necessary one. OA increases the number of potential users of any given article. The researchers who would otherwise have been unable to access articles because their institution could not afford the access-tolls of the journal can easily use OA articles. Therefore, OA can only increase both usage and impact.

There are different opinions about the extent of the problem. A recent review article ([Davis and Walters, 2011](#)) did not support the notion that there is a problem, as it stated that recent studies provide little evidence to support the idea that there is a crisis in access to scientific literature, at least for scientists themselves. Scientists do not perceive access to scientific literature as an especially important problem ([Davis and Walters, 2011](#)). Access to scientific literature is not a serious concern for most scientists in developed nations, and they actually feel that their access to the literature is steadily improving. A survey on information-seeking behavior of Slovenian researchers has confirmed that OA has obviously not gained a lot of scientists' attention. One of the reasons might also be the traditionally well-organized access to academic journals in Slovenia ([Vilar, Juznic, Bartol, 2012](#)).

Citation analysis and impact of OA publishing

Surveys are one way to estimate the effect of OA on the impact of the article but may be biased due to subjective views. This is the main reason

why OA impact on the availability of information sources for researchers was also a topic which was researched by bibliometrical methods, namely citation analysis.

The citation analysis is based on the information seeking behavior of the researchers as they cite their information resources. Citations identify earlier research results which have been used by the authors in the process of performing and representing their own research. A citation can also be understood as the manifestation of scientific influence. As stated by [Moed \(2005\)](#), “Outcomes of citation analysis must be valued in terms of a qualitative, evaluative framework that takes into account the substantive contents of the works under evaluation”.

Open accessibility is also a very important political issue, since it can improve the awareness of researchers and their funding bodies for OA. The published results tend to differ, although most find positive correlations between OA and the impact/citation advantage. The comparison of the citation counts of individual OA and non-OA articles published in the same non-OA journals was suggested as the correct method ([Harnad and Brody, 2004](#)). The authors claimed that the advantage of OA was very obvious in fields such as computer science and physics, and suggested that self-archiving increases citations by 50% or more ([Harnad, 2006](#)). Many authors have reported that online availability considerably increases a paper's impact ([Harnad and Brody, 2004](#); [Antelman, 2004](#); [Eysenbach, 2006](#)). There are also some analysis which do not confirm the link between OA and the increasing number of citations ([Craig et al., 2007](#); [Norris et al., 2008](#); [Miguel et al., 2011](#)). Some previous studies have indicated that self-archiving ([Gargouri et al., 2010](#)) and open accessibility ([Lin, 2007](#); [Lin, 2009](#)) substantially increase citation impact, but the citation advantage effect of OA publications varies between disciplines ([Norris et al., 2008](#)). The conclusions of [Moed \(2007\)](#) for articles in solid state physics are the same as Kurtz's ([Kurtz et al., 2005](#)) for articles in astronomy. The quality effect (authors post their most quality articles as openly accessible) and an early view effect (preprints are visible earlier and are cited more often) impacts the number of citations, and therefore the OA articles receive more citations than non-OA articles. [Eysenbach \(2006\)](#) published the results of an analysis of articles in a hybrid OA journal, the Proceed-

ings of the National Academy of Sciences (PNAS). While OA status was found to remain a significant predictor, there were also other factors – among these were the number of authors of the paper and funding from competitive grants. One very important conclusion of this study is the appeal to funders. They might provide financial support to authors for paying the publishers for immediate OA to their publication.

Research

In our study on the influence of open-access on articles' citations we relied on objective bibliometric evidence. We compared OA articles with non-OA articles in the same journal and hypothesized that OA would have a positive impact on the number of citations. We analyzed articles in three research journals published in 2007 in the field of civil engineering. These journals were chosen on the basis of two criteria. Firstly, the journals in which the researchers from the University of Ljubljana, Faculty of Civil and Geodetic Engineering (UL FGG) have published were chosen. Secondly, all three journals have an impact factor (IF) and are categorized in the same subject category "engineering, civil". This subject category included 88 journals in 2007. Their IFs were different, but generally ranked higher than average in this category. Two of them were in the second quarter (*Automation in Construction* (hereinafter AutCon), IF 2007 = 0.61 and *The Journal of Computing in Civil Engineering* (hereinafter JCCE), IF 2007 = 0.71), and one (*Computers & Structures* (hereinafter Com&S), IF 2007 = 0.93) in the first quarter. In a previous study ([Koler-Povh et al., 2011](#)), it was found that researchers of UL FGG published most of their scientific articles (72%) in journals that are ranked in the first or second quarter. Two databases were used as a source of citation data: Google Scholar (GS), which is available to common users and allows them the open or free access to full texts of journal articles without the condition of toll-access, and the Web of Science (WOS), as a world known system for bibliometric evaluation of scientific work that is relevant in Slovenia too. By choosing GS instead of Google, we assumed that scholarly information is sought by users who typically use the specialized GS database rather than Google. As [Jacso \(2005\)](#) found, the GS database includes many scientific publications, like thesis, research papers, reports of scientific projects etc., which are not included in the Google search en-

gine, which includes many commercial publications that are not useful for scientific work. Comparing the databases of WOS, Scopus and GS, [Bar-Ilan \(2010\)](#) found that GS gives the highest number of citations. However, the fact that GS often includes repeated citations of the same article should not be neglected, i.e. in bilingual articles/journals, or articles with the same title published in conference proceedings and in the journal, or the publishing of the summary and the publishing of full text at the same time. When counting journal citation only, WOS gives the most relevant results.

We studied articles published in 2007 in the three journals which are present in both systems, with an emphasis on the citation of each of them with citation data until February 29th 2012.

Methodology

Firstly, the GS database was used. The search was conducted outside the domain of the University of Ljubljana, as we emulated a user who uses the GS tool for Internet access to scholarly information and has no access to the journals subscribed to by the libraries of the University of Ljubljana. We collected data in the period between February 26th and 28th 2012. The data were sorted by journal and categorized into two groups. The first group included OA articles, whereas the second comprised non-OA articles. For each article we collected the number of cumulative citations until the end of February 2012. The data obtained were compared to those obtained with the same analysis in WOS. The data for each journal are presented in [Table 1](#).

Table 1. The number of articles by journals and by the open access business model

	No. of OA articles	Share %	No. of non-OA articles	Share %	Total	Share %
AutCon	30	34	59	66	89	100
JCCE	11	23	37	77	48	100
Com&S	40	27	109	73	149	100

The share of OA in the three journals, published in 2007, was substantial. The lowest share was 23% in the JCCE-ASCE, and the highest was in the AutCon (34%). Mainly, the articles are made OA mostly via different institutional repositories, and only a few of them are openly accessible through the author's websites. The articles are usually in the form of a peer-reviewed author's final version in pdf format. Rarely are they identical to the printed version of the publisher.

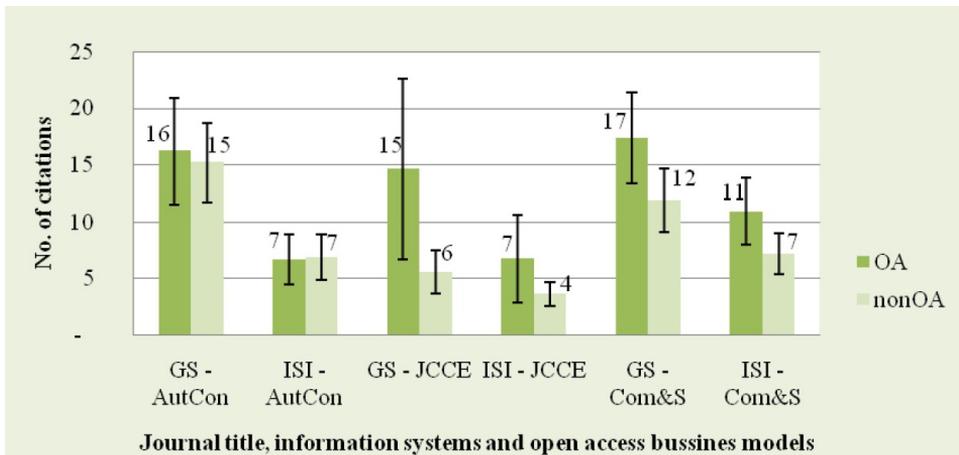


Figure 1. The arithmetic mean of citations and 95% confidence intervals

As shown in [Figure 1](#), open-access articles received more citations than those which were not openly accessible. However, the t-test reveals that this applies to the Com&S journal only (the significance level is $\alpha=5\%$). This is the journal with the highest IF ranked into the first quarter. In the case of JCCE, open access significantly (with the significance level being 5 percent or less) affects the number of citations only in the GS database. The effect of OA on the citation number is not significant for AutCon in neither the GS nor the WOS database.

In order to conclude if the business model of OA has a significant effect on the citation number, we will perform an analysis with a larger sample. Certainly, the large variability of data influenced the results. Namely, the numbers of citations are very different for either OA or non-OA articles in the same journal. The data are given in [Table 2](#).

Interestingly, the most cited article is a non-OA article (from Perez and Behdinan, published in Com&S). This article gets the maximum of citations in both information systems, GS and WOS. That might support other findings that, indeed, the quality of the article is the most important factor which influences the authors' decision to deposit the articles of highest quality into repositories. This is also observed by [Kurtz et al. \(2005\)](#), [Davis and Fromerth \(2007\)](#), [Craig et al. \(2007\)](#) and [Gargouri et al. \(2010\)](#).

Table 2. Number of citations

			OA	non-OA
AutCon	GS	max	41	65
		min	1	1
		median	13	10
	WOS	max	23	40
		min	0	0
		median	5	4
JCCE	GS	max	35	22
		min	3	0
		median	16	3
	WOS	max	15	11
		min	0	0
		median	8	3
Com&S	GS	max	47	103
		min	1	0
		median	13	8
	WOS	max	37	64
		min	1	0
		median	10	5

5. Conclusion

The analysis of articles' citations was the main scope of the presented paper. The articles published in 2007 in three civil engineering non-OA journals were considered. Google Scholar and Web of Science databases were utilized. It was found that OA significantly influenced the citation counts for the articles published in the *Computers & Structures* journal, which is ranked in the first quarter – according to both databases. Only the GS database showed a significant effect of OA on citations for the articles published in the *Journal of Computing in Civil Engineering*. Neither GS nor WOS databases indicate a significant effect of OA on the citation counts of articles in the *Automation in Construction* journal. These two journals are ranked in the second quarter among 88 journals in the same subject category, civil engineering. The present results indicate that more research is needed to give a final answer to the principle question of the paper: does open access have a significant impact on citations in the field of civil engineering. Some other potentially influential factors will be tested as well.

References

- Antelman, K. (2004). Do open-access articles have a greater citation impact? *College & Research Libraries*, 65(5), 372-382. Retrieved from http://eprints.rclis.org/5463/1/do_open_access_CRL.pdf
- Bar-Ilan, J. (2010). Citations to the “Introduction to informetrics” indexed by WOS, Scopus and Google Scholar. *Scientometrics*, 82, 495-506. doi:10.1007/s11192-010-0185-9.
- Craig, I.D., Plume, A.M., McVeigh, M.E., Pringle, M., & Amin, M. (2007). Do Open Access Articles Have Greater Citation Impact? *Journal of Informetrics*, 1, 239-248. Retrieved from http://www.publishingresearch.net/Citations-SummaryPaper3_000.pdf.pdf
- Davis, P.M., & Fromerth, M.J. (2007). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? *Scientometrics*, 71(2), 203-215. Retrieved from <http://arxiv.org/ftp/cs/papers/0603/0603056.pdf>
- Davis, P.M., & Walters, W.H. (2011). The impact of free access to the scientific literature: A review of recent research. *Journal of the Medical Library Association*, 99(3), 208-217. doi:10.3163/1536-5050.99.3.008.

- Eysenbach, G. (2006). Citation Advantage of Open Access Articles. *PLoS Biology*, 4(5), e157. [doi:10.1371/journal.pbio.0040157](https://doi.org/10.1371/journal.pbio.0040157).
- Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., Harnad, S. (2010). Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research. *PLoS One*, 5(10), e13636, [doi:10.1371/journal.pone.0013636](https://doi.org/10.1371/journal.pone.0013636).
- Guédon, J. (2004). The "Green" and "Gold" Roads to Open Access: The Case for Mixing and Matching. *Serials Review*, 30(4), 315-328. [doi:10.1016/j.serrev.2004.09.005](https://doi.org/10.1016/j.serrev.2004.09.005).
- Guédon, J. (2008). Mixing and Matching the Green and Gold Roads to Open Access: Take 2. *Serials Review*, 34(1), 41-51. [doi:10.1016/j.serrev.2007.12.008](https://doi.org/10.1016/j.serrev.2007.12.008).
- Harnad, S., & Brody, T. (2004). Comparing the impact of open access (OA) vs. non-OA articles in the same journals. *D-Lib Magazine*, 10(6), 1–6. Retrieved from <http://www.dlib.org/dlib/june04/harnad/06harnad.html>, 2012 Feb 29.
- Harnad, S. (2006). Publish or perish - self-archive to flourish: The green route to open access. *ERCIM News*, 64, Retrieved from <http://eprints.soton.ac.uk/261715/1/harnad-ercim.pdf>, 2012 Feb 29.
- Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., et al. (2008). The Access/Impact Problem and the Green and Gold Roads to Open Access: An Update. *Serials Review*, 34(1), 36-40. [doi:10.1016/j.serrev.2007.12.005](https://doi.org/10.1016/j.serrev.2007.12.005).
- Jacso, P. (2005). As we may search: Comparison of major features of Web of Science, Scopus and Google Scholar citation-based and citation-enhanced databases. *Current Science*, 89(9), 1537-1547. Retrieved from <http://www.iisc.ernet.in/currsci/nov102005/1537.pdf>
- Koler-Povh, T., Južnič, P., Turk, Ž., & Turk, G. (2011). Analysis of scientific publications in civil and geodetic engineering in Slovenia, in the case of the Faculty of civil and geodetic engineering in University of Ljubljana. *Geodetski vestnik*, 55(4), 764-779. Retrieved from http://www.geodetski-vestnik.com/55/4/gv55-4_764-780.pdf, 2012 Jan 7.
- Kurtz, M.J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Henneken, E., & Murray, S.S. (2005). The effect of use and access on citations. *Information Processing & Management*, 41(6), 1395-1402. [doi:10.1016/j.ipm.2005.03.010](https://doi.org/10.1016/j.ipm.2005.03.010).
- Lin, S. (2007). Non-Open Access and Its Adverse Impact on Molecules. *Molecules*, 12(7), 1436-1437. [doi:10.3390/12071436](https://doi.org/10.3390/12071436).

- Lin, S. (2009). Full Open Access Journals Have Increased Impact Factors. *Molecules*, 14(6), 2254-2255. [doi:10.3390/molecules14062254](https://doi.org/10.3390/molecules14062254).
- Miguel, S., Chinchilla-Rodriguez, Z., & De, M.F. (2011). Open access and Scopus: A new approach to scientific visibility from the standpoint of access. *Journal of the American Society for Information Science and Technology*, 62(6), 1130-1145. [doi:10.1002/asi.21532](https://doi.org/10.1002/asi.21532).
- Moed, H.F. (2007). The effect of “open access” on citation impact: An analysis of ArXiv's condensed matter section. *Journal of the American Society for Information Science and Technology*, 58(13), 2047-2054. [doi:10.1002/asi.20663](https://doi.org/10.1002/asi.20663).
- Moed, H.F. (2005). Citation analysis in Research Evaluation. In: *Information Science and Knowledge Management*, 9, 346. Retrieved from <http://www.springer.com/computer/book/978-1-4020-3713-9>
- Norris, M., Oppenheim, C., & Rowland, F. (2008). The citation advantage of open-access articles. *Journal of the American Society for Information Science and Technology*, 59(12), 1963-1972. [doi:10.1002/asi.20898](https://doi.org/10.1002/asi.20898).
- NIH Public Access Policy Details. Retrieved from <http://publicaccess.nih.gov/policy.htm>, 2012 Feb 29.
- Solomon, D.J., & Björk, B.C. (2012). Publication fees in open access publishing: Sources of funding and factors influencing choice of journal. *Journal of the American Society for Information Science and Technology*, 63(1), 98-107. [doi:10.1002/asi.21660](https://doi.org/10.1002/asi.21660)
- Vilar, P., Juznic, P., & Bartol, T. (2012). Information-Seeking Behaviour of Slovenian Researchers: Implications for information services. *The Grey Journal, International Journal on Grey Literature*, 8(1), 43-53. Retrieved from <http://connection.ebscohost.com/c/articles/73934397/information-seeking-behaviour-slovenian-researchers-implications-information-services>

Computer Science and Information Systems: Publishing an International Open Access Journal in a Developing Country

Mirjana Ivanović, Miloš Radovanović, Vladimir Kurbalija and
Jovana Vidaković

comsis@uns.ac.rs

University of Novi Sad, Department of Mathematics and Informatics, Trg D. Obradovića
4, 21000 Novi Sad (Serbia)

Abstract

Computer Science and Information Systems (ComSIS) is an open-access, peer-reviewed scientific journal founded in Serbia. Based on our experience in managing ComSIS, in this paper we analyze the impact of the open-access publication model on the success of journals from developing and transition countries (DTCs). We discuss alternatives for financial sustainability, the importance of coverage in citation databases through observed effects on citation and article submission patterns, and the challenges that potential success can bring with respect to the increased volume of submitted articles, plagiarism, and the perception of the journal. Our experiences suggest that: (1) open access and financial sustainability can coexist in a DTC, (2) coverage in WoS is the single most determining factor for the success of an up-and-coming journal, (3) domestic and international perception a journal from a DTC, newly covered in WoS, can lead to an increased volume of low-quality and even plagiarized submissions, placing additional strain on the editorial staff. We hope that the findings and arguments presented in this paper will assist current and potential journal editors and publishers, especially from developing countries, in finding suitable models for ensuring the success of their respective journals.

Keywords: Open Access; journal publishing; developing and transition countries; computer science; sustainability; coverage; perception; citation; Web of Science; Google Scholar

Introduction

As in many other areas dependent on financial sustainability, scientific journals from developing and transition countries (DTCs) may find it difficult to adopt the open-access model of publishing. Based on our experi-

ence in editing and managing Computer Science and Information Systems (ComSIS), an open-access, peer-reviewed scientific journal founded in Serbia, in this paper we analyze the feasibility and impact of open access on journal management, in the context of the business model, patterns of article submission and citation, coverage in major indexing databases (in this case, WoS), and the perception of the journal by the scientific community. Our primary goal is to provide current and potential journal editors and publishers, especially from DTCs, with useful guidelines for establishing suitable models of publication for their respective journals, rather than advertise ComSIS (although the latter may be difficult to avoid).

The rest of the paper is organized as follows. After describing issues pertinent to open-access publishing models, we provide a basic description of the ComSIS journal, its business model, patterns of article submission, and patterns of citation. This is followed by a discussion of the implications and recommendations relevant to journal management.

Open-access publishing

The open-access policy of distribution of published manuscripts ensures free availability on the public Internet, permitting any user to read, download, copy, distribute, print, search, or link to the full texts of published articles. From the standpoint of readers and authors this may be an ideal scenario for content dissemination; however the journals deciding to implement such a policy are faced with difficulties in generating revenue. [Björk \(2004\)](#) discusses business-model problems for OA journals, emphasizing that scaling-up from a few papers a year to larger publication volumes is difficult (employing staff, copyediting). The models considered by [Björk \(2004\)](#) include advertisements, subsidies from learned societies or research funding agencies, author charges, and “umbrella agreements” between universities and publishers.

On the upside, there is ample evidence supporting a positive correlation between journal accessibility and impact ([Wren, 2005](#)). In the cases where only some articles in a journal are published as open-access, the OA articles tend to be more cited than non-OA articles in the same journal ([Harnad & Brody, 2004](#); [Eysenbach, 2006](#)). [Norris et al. \(2008\)](#) also

support this observation, noting the existence of variety in this behavior across different disciplines (ecology, applied mathematics, sociology, and economics). [Davis et al. \(2008\)](#), on the other hand, argue that OA articles are no more likely to be cited than subscription-access articles in the first year after publication. Finally, [McVeigh \(2004\)](#) compares the impact factors of OA and non-OA journals indexed by ISI, observing that many OA journals are ranked low in their respective fields, with the existence of notable journals with very high rank, producing a mean impact factor for OA journals that is slightly above average. In addition, [McVeigh \(2004\)](#) notes that OA journals excel in the immediacy index.

Author perceptions of open-access journals are studied by [Schroter et al. \(2005\)](#). They conclude that authors generally do not object to author charges, however concern is raised regarding authors from developing and transition countries. In addition, OA journals may need to do more than their non-OA counterparts to reassure authors of their quality, as perceived quality is more important than the property of being OA or non-OA.

From the authors' viewpoint, "roads to open access" refers to the options for selecting a journal with the goal of making their article freely accessible: the "gold road" – publishing in a pure OA journal, and the "green road" – publishing in a non-OA journal that allows self-archiving of articles. [Harnad et al. \(2004\)](#), [\(2008\)](#) and [Guédon \(2004\)](#), [\(2008\)](#) discuss many issues and concerns regarding the different roads. In this article, we will highlight the importance of *journals* from DTCs taking the "gold road", as this provided our journal with the visibility of published articles instrumental in the struggle for acquiring international recognition.

The Computer Science and Information Systems (ComSIS) journal

The Computer Science and Information System (ComSIS) journal is an international open-access journal published in Serbia by the ComSIS Consortium. The ComSIS Consortium is a group of leading academic institutions from universities in Serbia and Montenegro, including the Serbian Academy of Sciences and Arts, who jointly publish the ComSIS journal. The first volume was printed in February 2004. Since then, nine volumes (21 issues) were published. In the 2010 release of Journal Citation Reports by Thomson Reuters, the ComSIS journal received its first

two-year impact factor of 0.324. In addition, ComSIS is covered or selected for coverage in various indexing services, including Scopus, EM-BASE, EBSCOhost, Inspec, DOAJ, DBLP and SCIndex.

ComSIS publishes original papers on both the theoretical foundations of computer science and commercial, industrial or educational aspects that provide new insights into the design and implementation of software and information systems. It also welcomes survey articles, and features invited papers by eminent researchers that contribute to the understanding of emerging and important fields of computer science. Plans are underway to establish regular columns covering reviews of newly published books, presentations of selected PhD and master thesis, as well as information on forthcoming professional meetings. In addition to wide-scope regular issues, ComSIS also includes special issues covering specific topics in all areas of computer science and information systems. All papers are published in English. Papers that pass a strict reviewing process are accepted for publishing.

Business model

From its inception, ComSIS has relied on three main sources of income: (1) an annual grant from the Serbian Ministry of Education and Science, (2) yearly institutional fees for consortium membership, and (3) publication of special issues. Special issues usually consist of selected and expanded versions of papers presented at international conferences and workshops, where event organizers are asked to cover the expenses of issue preparation and publication. Advertisements and sponsorship by companies were utilized in the first several years, but not recently. There are no author charges, and ComSIS charges an annual subscription fee for printed volumes only. Generally, publication expenses are on the rise because of the recent rapid increase of the number of article submissions. Taking everything into account, ComSIS is self-sustainable, but not-for-profit.

Submission of articles

All articles submitted to ComSIS need to pass a strict reviewing procedure before being accepted for publication. The reviewing procedure consists of one or more reviewing cycles depending on the reviewers' judgments. Several criteria influence the final decision: originality of the

paper, matching the scope of the journal, academic merit, quality of presentation, technical correctness, etc.

One of the most important steps in the whole process of publishing a paper is the initial examination of the submitted manuscript. Upon receiving a manuscript, the Editor-In-Chief and three members of the Editorial Board will examine it, and if the paper is outside the scope of the journal or does not meet the basic criteria for a journal article, it will be immediately rejected without further review. This step recently became particularly important because of the dramatically increased volume of submissions in the previous period. The number of submitted articles by month from January 2010 is shown in [Figure 1](#). The boost of submissions in the second half of 2011 is the direct consequence of acquiring the two-year impact factor (which was announced in June 2011). It is interesting to observe the fact that the number of preliminarily rejected papers followed this increasing trend, while the number of quality papers retained in reviewing stayed approximately the same (5-10 per month).

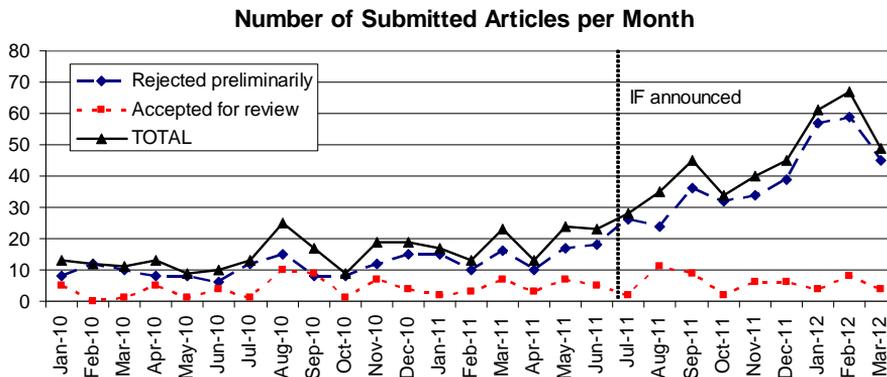


Figure 1. Number of articles submitted to ComSIS from January 2010 to March 2012.

Since the number of the preliminarily rejected papers dramatically increased from July 2011, it is interesting to examine the reasons for their rejection. The distribution of rejected papers according to the reason for rejection is shown in [Figure 2\(a\)](#). It is important to stress here that the Ed-

itor-In-Chief and the members of the Editorial Board make their decision solely on the basis of the quality of the paper and whether the subject of the paper fits the scope of the journal. Other aspects of paper assessment are left for later examination by reviewers. At this point, almost 50% of the papers are rejected because of their low quality and more than 80% because their topic falls outside the scope, while 32.56% of the papers satisfy both reasons. In addition, we have also noted (unfortunately, mainly during the review stage) a significant number of absolutely plagiarized and self-plagiarized submissions. In [Figure 2\(b\)](#) it can be seen that the majority of preliminarily rejected articles originate from Asian countries (with the category “Other,” which accumulates countries with less than 20 preliminarily rejected papers for the period, dominated by Asian and African countries). We do not imply that this observation is a general indicator of the quality of research in the given countries (after all, many articles published in ComSIS are authored by Asian researchers), but it is rather an artifact of formal criteria for academic advancement in these countries.

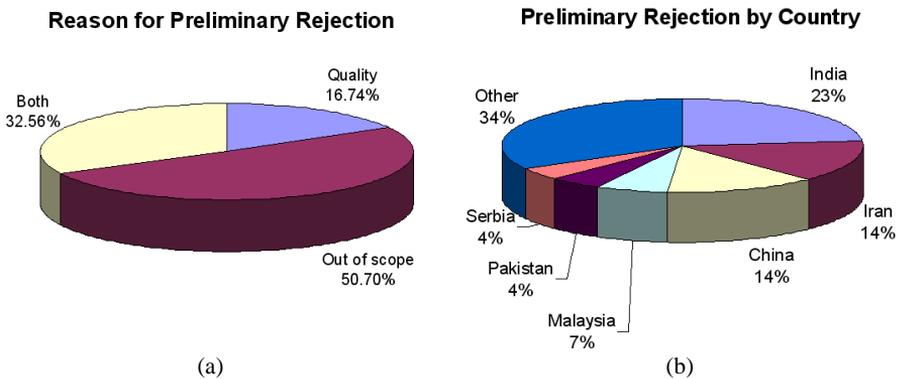


Figure 2. Distribution of preliminarily rejected papers by reason (a) and by country (b), from January 2010 to March 2012.

Article citation

The number of published articles and citations of ComSIS articles registered by WoS is shown in [Figure 3](#). Since the coverage of ComSIS in the Web of Science started in 2008, there was a general increase of interest in

the journal – reflected not so much in the number of regular submissions, but in proposals of third parties to publish special issues. Hence the increase in the number of published articles starting from 2010. Also, a large increase in the number of citations to ComSIS articles can be observed in 2011.

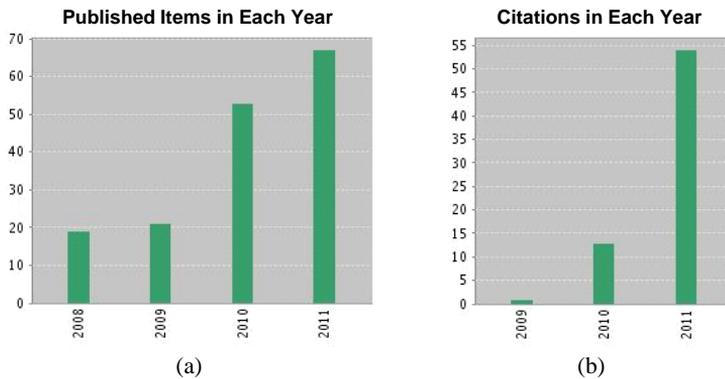


Figure 3. Number of ComSIS articles published each year (a) and number of citations in each year (b), according to the Web of Science (May 8, 2012).

To complement the citation data from WoS, we stretched Google Scholar’s functionality for creating author profiles and made a profile for our journal by carefully inserting all articles published in ComSIS into the profile. [Figure 4](#) shows citation indices and the number of citations of ComSIS articles per month, as detected by Google Scholar. Although the data is evidently not 100% accurate (e.g. citations in 2003 could not have existed since the journal is published from 2004), through manual inspection we have determined it is accurate enough to correctly express increasing/decreasing trends. In addition, Google Scholar’s data includes citations by conference and workshop papers that are mostly missing from WoS, which is significant since these venues play very important roles in the field of computer science, with papers usually needing to pass strict review procedures. A two-fold surge in the number of citations can be observed in 2009, which we believe can be associated with the announcement of the inclusion of ComSIS on the SCI-Expanded journal list by Thomson Reuters made in the previous year. The increase of the num-

ber of citations in 2011 can also be observed in this data, which can be explained partly by the increased volume of published articles (as opposed to 2009), and partly by efforts of the editorial staff that started in 2009 to make the journal more visible (covered in more indexing databases). The one database where ComSIS has been continually covered since its inception in 2004 is the Serbian Citation Index (SCIndex), whose support was crucial for the visibility of ComSIS articles, especially in the early years.



Figure 4. Citation indices (a) and number of citations to ComSIS articles in each year (b), according to Google Scholar (May 8, 2012).

Discussion

Based on our experiences editing and managing the ComSIS journal, described in the preceding section, we can arrive at the following conclusions and recommendations for current and prospective journal editors in DTCs.

Firstly, there are many options beside reader subscription for generating revenue, therefore open access and financial sustainability can coexist in a DTC. In addition, sacrificing short-term financial benefits of a closed-access model can lead to important long-term gains in terms of journal visibility, which ultimately translates to increased opportunity for attracting citations. Although ComSIS is backed by leading Serbian academic institutions, the influence of these institutions on the international stage is not strong enough to ensure a sufficient number of reader subscriptions were we to insist on a closed-access model. On the other hand, ComSIS is financially supported by the public sector through the annual grant from the Ministry of Education and Science, which is a financial option many journals may not have.

Secondly, although ComSIS is not supported by an internationally recognized and influential institution or publisher, such backing was effectively established through coverage in WoS, which we regard as the single most determining factor for the success of our up-and-coming journal: before acquiring an impact factor, ComSIS barely existed on the international stage; afterwards, the interest in the journal was dramatically boosted, with open access certainly contributing to increased citation and continued coverage.

Finally, based on the article submission patterns given in the previous section, it can be deduced that the perception of ComSIS as a journal from a DTC, newly covered in WoS, is that of an “easy” place to publish an article in a journal with an impact factor (which constitutes required minima for PhD theses, academic positions, etc., in a number of countries, some of which are shown in [Figure 2\(b\)](#)). This is most likely what lead to the increased volume of low-quality and even plagiarized submissions, placing additional strain on the editorial staff. Here it is important to emphasize that our editorial staff was not only faced with an increased volume in a “physical” sense, but was also challenged to retain a rigorous acceptance criterion and not succumb to the temptation of dramatically increasing the number of published articles, which may on one hand lead to short-term (financial) gains, but on the other hand is very likely to damage the reputation of the journal in the long run.

Conclusion

In this paper we believe we have shown that although a journal may be based in a developing country and might not have support from an internationally influential institution or publisher, it is still possible for the journal to acquire world-wide recognition and be financially sustainable. Although the journal we manage and use as a case study in this paper, Computer Science and Information Systems (ComSIS), still has a long way to go in terms of absolute success in its field, key factors for achieving the above goals can be identified: a suitable business model that incorporates open access, coverage in WoS, and preparedness to handle an increased workload and retain a rigorous acceptance criterion. As [Schrotter et al. \(2005\)](#) noted that open-access journals may need to achieve more in order to reassure authors of their quality (to which we would add that

DTC journals need to do even more), at the end of the day it is the perceived quality that is most important, and in our view quite achievable in a DTC with an appropriate journal management strategy, great effort, and enthusiasm.

Acknowledgements

The authors would like to thank the Serbian Ministry of Education, Science and Technological Development for their support through the “Intelligent techniques and their integration into wide-spectrum decision support” (OI174023) project.

References

- Björk, B.C. (2004). Open access to scientific publications: An analysis of the barriers to change. *Information Research*, 9(2), paper 170. Retrieved from <http://informationr.net/ir/9-2/paper170.html>
- Davis, P.M., Lewenstein, B.V., Simon, D.H., Booth, J.G., & Connolly, M.J.L. (2008). Open access publishing, article downloads, and citations: randomised controlled trial. *BMJ (Clinical research ed.)*, 337. [pmid:18669565](https://pubmed.ncbi.nlm.nih.gov/18669565/), [doi: 10.1136/bmj.a568](https://doi.org/10.1136/bmj.a568).
- Eysenbach, G. (2006). Citation Advantage of Open Access Articles. *PLoS Biology*, 4(5). [doi:10.1371/journal.pbio.0040157](https://doi.org/10.1371/journal.pbio.0040157).
- Guédon, J. (2004). The “Green” and “Gold” Roads to Open Access: The Case for Mixing and Matching. *Serials Review*, 30(4), 315-328. [doi:10.1016/j.serrev.2004.09.005](https://doi.org/10.1016/j.serrev.2004.09.005).
- Guédon, J. (2008). Mixing and Matching the Green and Gold Roads to Open Access: Take 2. *Serials Review*, 34(1), 41-51. [doi:10.1016/j.serrev.2007.12.008](https://doi.org/10.1016/j.serrev.2007.12.008).
- Harnad, S., & Brody, T. (2004). Comparing the impact of open access (OA) vs. non-OA articles in the same journals. *D-Lib Magazine*, 10(6). Retrieved from <http://www.dlib.org/dlib/june04/harnad/06harnad.html>
- Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., et al. (2004). The Access/Impact Problem and the Green and Gold Roads to Open Access. *Serials Review*, 30(4), 310-314. [doi:10.1016/j.serrev.2004.09.013](https://doi.org/10.1016/j.serrev.2004.09.013).
- Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., et al. (2008). The Access/Impact Problem and the Green and Gold Roads to Open

Access: An Update. *Serials Review*, 34(1), 36-40. [doi:10.1016/j.serrev.2007.12.005](https://doi.org/10.1016/j.serrev.2007.12.005).

- Mcveigh, M.E. (2004). *Open access journals in the ISI citation databases: Analysis of impact factors and citation patterns*. Thomson Corporation. Technical Report. Retrieved from <http://ip-science.thomsonreuters.com/m/pdfs/open-accesscitations2.pdf>
- Norris, M., Oppenheim, C., & Rowland, F. (2008). The citation advantage of open-access articles. *Journal of the American Society for Information Science and Technology*, 59(12), 1963-1972. [doi:10.1002/asi.20898](https://doi.org/10.1002/asi.20898).
- Schroter, S., Tite, L., & Smith, R. (2005). Perceptions of open access publishing: interviews with journal authors. *BMJ*, [doi:10.1136/bmj.38359.695220.82](https://doi.org/10.1136/bmj.38359.695220.82).
- Wren, J.D. (2005). Open access and openly accessible: a study of scientific publications shared via the internet. *BMJ*, [doi:10.1136/bmj.38422.611736.E0](https://doi.org/10.1136/bmj.38422.611736.E0).

Applicability of the Nano-publication Concept for Fostering Open Access in Developing and Transition Countries

Adam Sofronijević¹ and Aleksandra Pavlović²

¹sofronijevic@unilib.bg.ac.rs

University of Belgrade, University library “Svetozar Marković”, Bulevar Kralja Aleksandra 71, 11000 Belgrade (Serbia)

²pavlovic@unilib.bg.ac.rs

University of Belgrade, University library “Svetozar Marković”, Bulevar Kralja Aleksandra 71, 11000 Belgrade (Serbia)

Abstract

The paper presents nano-publications as a publishing alternative primarily developed for exposing OA (Open Access) datasets. The possible role of nano-publications for fostering usage of OA in developing and transitional countries is analysed. The paper provides insight into the theoretical background of nano-publications and describes the steps necessary to develop this format of publication. In this framework we present a more detailed analysis of the possibilities of involving the scientific communities of developing and transitional countries in operations on OA datasets exposed via nano-publications. Incentives and ethical issues are discussed in order to provide the initial motivation for further research of this novel idea. The role of nano-publications in the area of humanities is briefly described, along with suggestions of the possibilities for involving humanity scientific communities of developing and transitional countries in nanopublishing. The advantages of fostering OA initiatives in developing and transitional countries by implementing activities regarding nanopublishing are presented in order to encourage contemplation on this format of publishing in the OA communities of these countries.

Keywords: nano-publication; Open Access; developing and transition countries; dataset publishing; ethical issues; humanities; metadata; scientific communication

Introduction

A nano-publication is a very short declaration connecting two concepts by means of a third and providing metadata about this relation (conditions

under which the relation is viable, author, timestamp, etc.). An example of such a declaration would be: Trojans (concept 1) are transmitted by (concept 3) exe files (concept 2), with metadata describing the conditions (in PC computers), author, date of publication etc. Nano-publishing is a concept in the making that appeared as a response to the growing inadequacy of various aspects of existing tools for scientific communication. So far, the publishing industry and OA (Open Access) initiatives have not been very successful in using all the advantages of digital technology for the dissemination of scientific results, especially datasets. Therefore novel ideas and solutions are needed in order to expose the huge datasets that now exist, but are still invisible for those outside the labs producing them ([Bourne & de Waard, 2011](#)). Originating in life sciences, nano-publications seem to be envisioned and increasingly shaped as a tool for the efficient publishing of datasets. The example provided at the beginning of this section is general and not related to datasets, but an example of a nano-publication exposing small parts of datasets is given in the following section. The abundance of datasets is a relatively novel development in science. Not long ago quality datasets were strictly guarded and unavailable to outside researchers. Nowadays the gap between available datasets and the resources to even curate them let alone analyse them is widening each day. Contemporary tools for publishing datasets seem to be insufficient in comparison to the amount of data available and the possibilities that more effective data dissemination tools might offer. In this regard, one must observe that there are no sufficient incentives for researchers to publish datasets apart from those being used to support the publishing of datasets closely related to findings published in traditional scientific journals. The nano-publication concept has the potential to successfully face the challenge of providing a novel method of evaluating datasets and scientific work based on them, while at the same time preserving the values of the traditional means of scientific communication ([Mons et al., 2011](#)). When considering the ethical issues that need to be addressed in regards to nanopublications, it is good to keep in mind the ethical basis of the existing publishing paradigm ([Axton, 2011](#)). With nano-publications a new publishing paradigm is rising and all of its aspects need to be addressed – incentives to use it, operational details, but also the ethical issues. Nano-articles provide the basis for contemplation on the improvement of scientific communication in other areas such as

humanities. The nano-publication concept which we will discuss in the next section – though derived from life sciences – is nevertheless general enough to allow for contemplating the use of the concept in other scientific areas and for the purposes of publishing data originating in other scientific areas.

Defining nano-publications

In order to define nano-publications clearly, an example available at ([Nanopublication Downloads section, 2012](#)) is provided, with only the most important parts presented for the sake of clarity:

```
<nanopublication id="0">
  <assertion>
    <subject>NG_000007.3:g.70628G>A</subject>
    <predicate>has variant frequency</predicate>
    <object>0.25%</object>
  </assertion>
  <condition>Sardinian</condition>
  <provenance>
    <dateofcreation>March 24, 2011</dateofcreation>
    ...
    <evidenceType>empirical</evidenceType>
    <authorID>Giardine et. al.</authorID>
    <curatorID>unresolved</curatorID>
    ...
  </provenance>
</nanopublication>
```

A nano-publication is identified uniquely by its id. The assertion part of the nano-publication offers the relation of the concepts and is followed by the condition under which the assertion holds and by the metadata describing the context of the relation. The nano-publication is presented by means of XML which enables machine readability.

Nano-publications will be presented in more detail by providing a short description of the steps necessary to perform in order to create a nano-publication as described in ([Mons & Velterop, 2009](#)).

The first step addresses the problem of ambiguity coming from the usage of different terms and identifiers that refer to a concept. A concept, which

is a unit of thought, may be related to an object in the physical world, but may also be related to a completely abstract object. The natural flow of any language, when used for communication by humans, is to describe the same concept by numerous different terms and identifiers. This creates ambiguity and causes problems for machine interpretation of such statements. Because of this, the first step in the creation of nano-publications is mapping the terms and identifiers to concepts, and by this the elimination of ambiguity.

The second step comprises the creation of statements from concepts. Statements are considered to be the smallest insights in exact science, consisting of three concepts. These three concepts are usually called subject, object and predicate, where the subject and object are associated by means of predicate. While the relationship between these three concepts is clear enough, the statement itself needs additional metadata in order to ensure clarity in regards to the conditions under which the statement holds, the authority, the creation date and other annotated metadata deemed necessary by the community using nano-publications. Adding these metadata is considered to be step three in nano-publication creation. We need to emphasise the simultaneous unity and difference of steps two and three. The completeness and consistency of the statement with the metadata describing it is imperative, but it is very important for us to notice that the process of the creation of a statement is not necessarily done by the same entity and in the same timeframe as the process of the annotation of that statement.

The third step is about making boundaries for the statements i.e. providing the range of validity, specifying authority and adding other important metadata about the statement so that it makes sense for the community using it. If conditional statements are used to describe the conditions under which the said statement is to hold, then it is imperative that the describing statement also comprises of unambiguous concepts. This is to ensure that operations on such a statement can be seamlessly performed along with operations on the original statement. In regards to authority, the importance of factors defined as status by ([Mons & Velterop, 2009](#)) is exceptional. Status may describe any of the following characteristics of the

annotated statement: peer-reviewed, curated, disputed, retracted, hypothetical, observational, repetitive, etc.

The fourth step brings annotated statements to the level of proper publications that can be cited and credited by treating them as nano-publications. By making this crucial step, the incentives to publish the short statements in the form of nano-publications will appear and their number should increase over time. ([Mons & Velterop, 2009](#)) distinguish three types of statements suitable for publishing in the format of nano-publications. These are curated statements, observational statements and hypothetical statements. Curated statements are the building blocks of ontologies and are intrinsically statements that are “as true as it gets in science”. They should be described by metadata containing data on the originator of the triple statement and its curator, in order to facilitate incentives for those two categories. Observational statements include factual statements originating in the literature or in datasets. The more observational statements become annotated and published in the format of nano-publications, the more knowledge will become efficiently accountable. Hypothetical statements represent triplets of yet undiscovered statements that by virtue of reasoning or data mining or inference from existing texts have a high probability of being true. The incentive for publishing such statements is that – when confirmed or “discovered ” by other statements that originate in experimental results – the authors of hypothetical statements may get credit for their prediction along with authors confirming it.

Step five is about making nano-publications really useful. To achieve this, the redundancy in the present body of published scientific knowledge needs to be addressed. In order to facilitate understanding and following what is natural for language usage as a communication tool among humans, repetitions are very common in scientific publishing. This may be useful for us, but for machines this redundancy makes searching and retrieval inefficient. When converted to annotated triplets most of the data becomes redundant, thus making the remaining triplets really useful for researchers since they are now easily discoverable by machine aided search and retrieval.

As stated by ([“NanoPub: A beginner's guide to data publishing”, 2012](#)) nano-publications are based on open standards such as RDF and XML.

Adhering to standards should provide for machine readability of nano-publications, which has enormous importance for raising the efficiency of scientific communication. The lack of standards as of April 2012 should be noted, along with the existence of guidelines for the creation of nano-publications ([“Open PHACTS Nanopublication Guidelines v1.81”](#), 2012).

Some general principles of nanopublishing are presented at ([“NanoPub: A beginner's guide to data publishing”](#), 2012). The principle of added value addresses the novelty and quality of nano-publications by presenting examples of the origin of a predicate that associates an object with its subject (mathematical model, new experimental dataset, exposure of an existing dataset, etc.). The principle of transparency establishes the assurance of quality assessment of a nano-publication. The metadata describing statement should provide enough data on the predicate to ensure that others can assess its validity by author, date and time, spatial data and intrinsic quality of the predicate. The principle of ambiguity avoidance describes the need for all concepts in the statement and the metadata describing it to be unambiguously resolved to certain concepts. The principle of global reference addresses the issue of the referencing of entities in nano-publications. If an entity is established on the Web the accompanying URI should be used, and if not a UUID (Universal and Unique Identifier) standardized by the Open Software Foundation (OSF) should be derived from the available underlying database of concepts.

Nano-publications in developing and transition countries

Nano-publications have the potential to be an important driver of scientific research in first world countries which are ample in resources that allow for the production of datasets in areas like life sciences. Could nano-publications also play a significant role in fostering communication and usage of OA resources in scientific communities of developing and transitional countries? In general these countries lack the resources and the infrastructure for conducting research that produces large datasets. Consequently entire scientific disciplines in these countries are left behind or become obsolete. In order to prevent this, the scientific communities of these countries should be focusing their efforts on the usage of freely available scientifically relevant datasets. This was not the case in the past,

because datasets were not freely exposed. All of this may change with the rise of nano-publications.

Scientific communities and librarians of developing and transitional countries can engage in several activities and operations based on freely available datasets that may start to emerge with the wider usage of nano-publications. In order to have the basis for participation in operations over datasets, they may engage in creating and editing concepts regarding the members of their own community in databases such as ConceptWiki. Furthermore they can add to the mapping efforts that lead to unambiguously resolving the terms and identifiers into unique concepts. The need for *a posteriori* mapping of terms and identifiers in already published texts is vast, and text and data mining combined with human curation input can provide significant results in this area. If we further follow the steps in the creation of nano-publications described in the previous section, steps two and three provide a rich source of activities that may be of interest to scientific communities of developing and transitional countries. The exposure of triplet statements from datasets that are freely available is one such activity. The annotation of these triplet statements that were either produced by those creating annotation or by someone else is the next possibility. Various statements regarding the context and provenance of the triplet provide many opportunities if an incentive system is created – primarily within scientific communities of developing and transitional countries, but also worldwide. The creation of hypothetical statements building on available datasets may be especially interesting. This comes from the high probability that a worldwide incentives system may emerge for these types of predictions. Redundancy removal described as step five in the previous section also offers some possibilities for the activities of members of scientific communities of developing and transitional countries. The analysis of statements and the removal of redundancies along with the clustering and linking of statements that are not redundant may present important work that should upgrade the existing semantic Web and provide for fair rewards for those involved with it.

So far we have explained in some detail the usage and purpose of nano-publications in life sciences. In developing and transitional countries humanities and social sciences scholars are still numerous when compared

to statistics of the first world countries. Besides this, humanities and social sciences in developing and transitional countries are especially hard-hit by the widening gap in skills and knowledge when compared to their peers in first world countries, and also by budget cuts due to the ongoing financial crisis. Therefore, in order to discuss the role of nano-publications in developing and transitional countries we need to touch upon their possible role in humanities. Very few mentions of this role are available as of April 2012. ([Mons & Velterop, 2009](#)) mention in passing the applicability of nano-publications in humanities and conclude that it may have an important role in this area as well, and that most of the characteristics described on the basis of examples originating in life sciences may also be applicable for humanities. ([Chambers, 2011](#)) provides a more detailed example of a nano-publication originating from humanities. Based on this example it seems that nanopublishing should be embraced by the humanities scientific community, especially in regards to recent developments regarding the increasing interest in digital humanities. This may be an important field of activity for scientific communities in the humanities from developing and transitional countries, since resources needed to employ nanopublishing are mostly available to them thanks to the fact that they are mostly open sourced. Basic ICT equipment should suffice to employ these open sourced technologies and it is up to the mentioned scientific communities to demonstrate their innovativeness and employ the idea originating from other scientific fields to their own with merit and ethical stance. The creation of a database comprising concepts in humanities may be followed by nanopublishing initiatives similar to the one presented in ([Chambers, 2011](#)).

All mentioned activities are very promising if fair use is applied i.e. if ethical standards of scientific conduct are respected. In order to ensure this, especially in scientific communities of developing and transitional countries that may still be struggling to achieve proper levels of scientific ethics, a system of regulations and control need to be established. On the other hand, imposing too much control and regulation in an area that builds on OA principles and high ideals of international scientific community collaboration may damage the overall effort and stall the process to a standstill. Therefore we propose a balanced approach and employment of self-regulatory mechanisms. One such example may be Concept-

Wiki, where quality is intrinsically ensured by the properties of Wiki technology that enables seamless internal control of those involved in the process of creation and the editing of concepts. In general we propose usage of Web 2.0 technologies for the dissemination of results in the area of nanopublishing because by their own virtue these technologies provide a transparent environment that is likely to expose any misbehaviour. Mandatory blogging of results for all those involved in this line of scientific work is one example of activities that add to the transparency of the nanopublishing process. The importance of the ethical issues regarding scientific work in the area of nano-publications is underlined in ([Groth, Gibson & Velterop, 2010](#)), with an emphasis on citation and the *asserted by* mechanism that allows for the evaluation of the source grounding the base for nano-publications. Existing guidelines for the creation of nano-publications ([“Open PHACTS Nanopublication Guidelines v1.81”, 2012](#)) do encompass an Integrity key that should ensure the authority of nano-publications and a nano-publication ID that should ensure its unique identification among other nano-publications. As of April 2012 these two parameters are not yet technically described in these guidelines, but their mere presence suggests the importance of ethical issues in nanopublishing.

Finally, does nano-publishing have the potential of fostering OA in developing and transitional countries? We have shown that strong incentives may exist for the usage of nano-publications in developing countries. Since all concepts and technologies involved with nano-publications are open sourced and compliant with the OA initiative, the sheer usage of nano-publications should be considered as an important part of OA usage. Further, by becoming familiar with different open source technologies and ideas that are intrinsic for nano-publications, they may become a specific driver of OA in developing and transitional countries. Therefore we propose special support for nano-publication initiatives from the OA community in order to foster the usage of OA materials in conjunction with nanopublishing. Also, the sheer usage of nano-publications in developing and transitional countries will ensure the dissemination of novel technologies and ideas such as RDF, XML, URI, UUID, ConceptWiki, etc. that ground a basis for the usage and flourishing of OA initiatives.

Conclusions

Nano-publications have always been related to OA (Open Access). This mostly comes from the opposition of authors contemplating it to the socially inefficient system prevailing the contemporary publishing industry. Nano-publications were related to OA either directly – by referencing the concept as such by authors – or indirectly – by means of relating the background of authors contemplating it. In all cases nano-publications were considered as OA publications only. Nanopublishing may also be considered as an important driver of OA promotion and usage in general. This especially stands for developing and transitional countries, where the specific circumstances of scientific activities predominantly determined by lack of resources promotes the usage of existing datasets via nano-publishing. A possibility for nano-publishing usage in the area of humanities is also an important aspect of their overall applicability in these countries.

Nano-publication theory and practice are both in a very early phase. A lot of additional contemplation and implementation experience is needed in order to further develop this idea and achieve its wider usage. In spite of this, it is obvious even in this early phase that nano-publications are inherently efficient in scientific knowledge dissemination. This important fact may be used as a basis for further contemplation and practical implementation, but one also has to have in mind that this too could lead to the commandeering of this publication form by the for-profit publishing industry. Nano-publications, especially some of their elements like the open availability of data (full text / raw data) and the drop of emphasis on the “publish or perish” attitude ([Laika's Medliblog, 2010](#)), have been perceived in a broader context as a new form of scientific dissemination and access provisioning that can contribute in reducing the tensions caused by the rigidity of the publishing industry. Because of this, theoretical and practical provisions may be needed in order to enable equally qualitative usage in different business models employing this novel concept.

References

- Axton, M. (2011). Crowdsourcing human mutations. *Nature Genetics*, 43(4), 279-279. [doi:10.1038/ng0411-279](https://doi.org/10.1038/ng0411-279).

- Bourne, P., & de Waard, A. (2011). Notes from the breakout sessions. In: *Workshop Beyond the PDF* held on January 19-21, 2011 at University of California San Diego. Retrieved from <http://sites.google.com/site/beyondthepdf/home/program-draft/notes-from-the-breakout>, 2012 Jan 25.
- Chambers, S. (2011). Nano-publications in the arts and humanities. Retrieved from <http://www.slideshare.net/schambers3/nanopublications-in-the-arts-and-humanities>, 2012 Feb 15.
- Groth, P., Gibson, A., & Velterop, J. (2010). The anatomy of a nanopublication. *Information Services and Use*, 30(1), 51-56. Retrieved from [http://www.w3.org/wiki/images/c/c0/HCLSIG\\$\\$SWANSIOC\\$\\$Actions\\$\\$RhetoricalStructure\\$\\$meetings\\$\\$20100215\\$cwa-anatomy-nanopub-v3.pdf](http://www.w3.org/wiki/images/c/c0/HCLSIG$$SWANSIOC$$Actions$$RhetoricalStructure$$meetings$$20100215$cwa-anatomy-nanopub-v3.pdf)
- Laika's, M. (2010). Will Nano-Publications & Triplets Replace The Classic Journal Articles? Retrieved from <http://laikaspotnik.wordpress.com/2010/06/23/will-nano-publications-triplets-replace-the-classic-journal-articles/>, 2012 Jan 24.
- Mons, B., & Velterop, J. (2009). Nano-Publication in the e-Science Era. In: *Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009)*. Retrieved from [http://www.w3.org/wiki/images/4/4a/HCLSIG\\$\\$ISWC2009\\$\\$Workshop\\$\\$Mons.pdf](http://www.w3.org/wiki/images/4/4a/HCLSIG$$ISWC2009$$Workshop$$Mons.pdf)
- Mons, B., Van, H.H., Chichester, C., Hoen, P., Den, D.J.T., Van, O.G., et al. (2011). The value of data. *Nature Genetics*, 43(4), 281-283. [pmid:21445068](https://pubmed.ncbi.nlm.nih.gov/21445068/), [doi: 10.1038/ng0411-281](https://doi.org/10.1038/ng0411-281).
- NanoPub: A beginner's guide to data publishing (2012). Retrieved from <http://nanopub.org/original/index.html>, 2012 Apr 10.
- Nanopublication Downloads section, Human Hemoglobin Genetic Variation (2012). Retrieved from http://nanopub.org/original/hemoglobin_nano-pub_download.html#, 2012 Apr 10.
- Open PHACTS Nanopublication Guidelines v1. 81 (2012). Retrieved from http://www.openphacts.org/documents/publications/OpenPHACTS_Nanopublication_Guidelines_v1.8.1.pdf, 2012 Apr 10.

Output in WoS vs. Representation in JCR of SEE Nations: Does Mother Thomson Cherish All Her Children Equally

Biljana Kosanović¹ and Pero Šipka²

¹ biljana@nb.rs

National Library of Serbia, Skerlićeva 1, 11000 Belgrade (Serbia)

² sipka@ceon.rs

Centre for Evaluation in Education and Science, Kneza Miloša 17, 11000 Belgrade (Serbia)

Abstract

Within the 2005-2010 campaign aimed at extending the coverage of quality regional journals, Thomson Reuters started indexing a large number of journals published among else in SouthEast Europe (SEE). In this study we looked at the range and effects of this expansion. The number of national journals indexed in JCR was related to the national performance in WoS calculated in two ways: as it is and as corrected for the contribution of national journals. The study included 14 countries. All data were downloaded from WoS. The period from 2007 to 2010 was covered. It was found that, as a result of the campaign, coverage of various SEE countries became highly uneven. It was also found that the increment in coverage poorly corresponds with the potentials of various national research communities, as measured by the Relative Citation Index (RCI). Corrected performance, calculated by omitting national journals from each country's output, did influence the publication and citation rate of SEE countries, but not to the degree of changing country rankings on the two measures. However, it produced substantial changes in ranking on RCI, suggesting that overrepresentation has serious practical consequences on the bibliometric status of countries from the region. Decline in RCI rankings was observed among those overrepresented countries that published journals of lower internationality and Impact Factor. The results generally suggest that overrepresentation in WoS/JCR is not in the long-term interest of both awarded countries and Thomson Reuters itself.

Keywords: countries; scientific journals; coverage; performance; citations; Web

Introduction

The coverage of early Thomson Reuters (then ISI) citation databases has been widely criticized for its "North American centrism" and "English language favouritism". Partly in response to this criticism, the share of journals and papers published outside the USA started growing in the 1980s, with an acceleration in the 1990s ([Shelton et al., 2009](#)). The trend received a boost in 2005-2010, when Thomson Reuters accepted for indexing as many as 1600 journals in order "to complement its coverage of top tier, highly cited international journals with an appropriately broad representation of the world's best regional literature" ([Testa, 2011](#)). Testa pointed out that, following Thomson Reuters' editorial mission, all journals added to Web of Science coverage were evaluated and selected individually according to the principles articulated in the Journal Selection Process ([Testa, 2006](#)).

One of the regions that profited from the Thomson campaign the most was SouthEast Europe (SEE), where research communities had traditionally been unhappy with the representation of their journals in international databases. Inclusion of a large number (108 in 2009, and another 88 in 2010) of SEE journals in the JCR was accepted in the region with unshared enthusiasm. The admission was seen as a way out of the oft-repeated "vicious circle of inadequacy" of small journals ([Marušić & Marušić, 1999](#)), expected to result in what journals urged for – "increasing their visibility, raising the number of submissions, enabling their editors to publish better science, and eventually improving the journal's impact factor" ([Sambunjak et al., 2008](#)). At the same time, massive admission brought some concerns about the possible controversial effects on the position of individual journals in the well established system of their hierarchy in regional science. Many instantly became aware that the number of newly accepted journals was not evenly distributed among SEE countries of about the same scientific potential and publishing tradition. This, on the other hand, brought concerns about the possible effects of a sudden change on the cross-national comparison in performance, as measured by WoS publications and citations. [Basu \(2010\)](#) demonstrated that

the Thomson coverage of journals published in various countries, for which he coined the term “journal packing density”, can have a serious impact on these countries’ research performance. By using correlation design on a sample of 86 countries, he found that as much as 87% of the variance in the number of papers published in *Scopus* by a country can be explained by the number of journals published in that country.

Since this finding is open to various interpretations, we decided to study the same problem by using a different approach in a different context. More specifically, the purpose of our study was to check if the new Thomson extension of JCR coverage in SEE countries has any effect on their research performance as measured by productivity and impact in WoS. We sought answers to the following specific questions:

- (1) How do SEE countries compare among themselves in terms of WoS publication rate, WoS citation rate, and Relative Citation Index (RCI), as a synthetic indicator combining the two measures?
- (2) To what extent are these countries represented in WoS and JCR journal lists?
- (3) Does relative performance of SEE countries change if their coverage by WoS/JCR is made virtually constant, i.e. when papers published and citations received in regional journals are removed from the calculation of performance?
- (4) Are these changes, if existent, related to the basic characteristics of “national” journals, such as Impact Factor and internationality?

Method

All data were obtained by searching the Web of Science, using the 2007-2011 publication window, and by extracting information from the Journal Citation Report, 2003-2010 edition. The time span was intentionally selected to cover the period of the large extension of Thomson’s coverage, as well as the period preceding this critical change. Downloading was performed during the last week of February, 2012.

The list of countries to be studied was adopted from SEESAmE: South-East European Science Amended through Evaluation (sesame.ceon.rs). In SEESAmE, the SEE region was delineated by using both geographical and historico-political criteria.

Representation of SEE countries in JCR was expressed by the number of journals in the Journal Summary List of the JCR 2010 edition ([Thomson Reuters, 2011](#)), as the last published at the time of gathering data for the study. Data for years 2003-2009 were taken from JCR editions for respective years.

Representation of SEE countries in WoS was expressed by the number of journals on the Source Publication List, March 2012 edition ([Thomson Reuters, 2012](#)).

Country productivity in WoS was expressed as the number of papers published in WoS from 2007 to 2011. Only papers classified as Articles or Reviews were counted. A paper was accredited to a country if any of the authors was affiliated with that country. Papers by multiple authors from different countries were assigned to all affiliated countries, regardless of the number of authors from each listed country.

Country citation rate in WoS was expressed as the number of citations received in WoS journals only for articles published in WoS. Assignment of citations to countries was done by using the same formula as for publications.

Country Relative Citation Index was calculated as the ratio of the number of papers and the number of citations in the observed period.

Corrected countries' performance, as a measure of their output without the contribution of their own ("national") journals, was determined by omitting all journals published in an analysed country in repeated country-wise search. While performing this operation, journals from other SEE countries were not excluded.

Average Journal Impact Factor was calculated as the mean score of values extracted from the 2003-2010 editions of the Journal Summary List of JCR.

Journals Internationality was expressed as the share (%) of articles by foreign authors, excluding articles co-authored with national authors.

Results

It is evident from [Figure 1](#) that SEE as a region benefited from the Thomson extension campaign heavily. However, the gain was not evenly spread across the countries, which produced marked changes in their rankings, relatively stable for decades. The largest number of accepted journals came from Turkey and Romania. They practically became the top represented countries in the region overnight. The increment of Croatian journals is also huge. Hungary belongs to this group too, but was granted to a lesser extent, which made it lose its long-term primacy. All in all, the changes can be characterized as drastic, both in terms of ranking order and differences among countries.

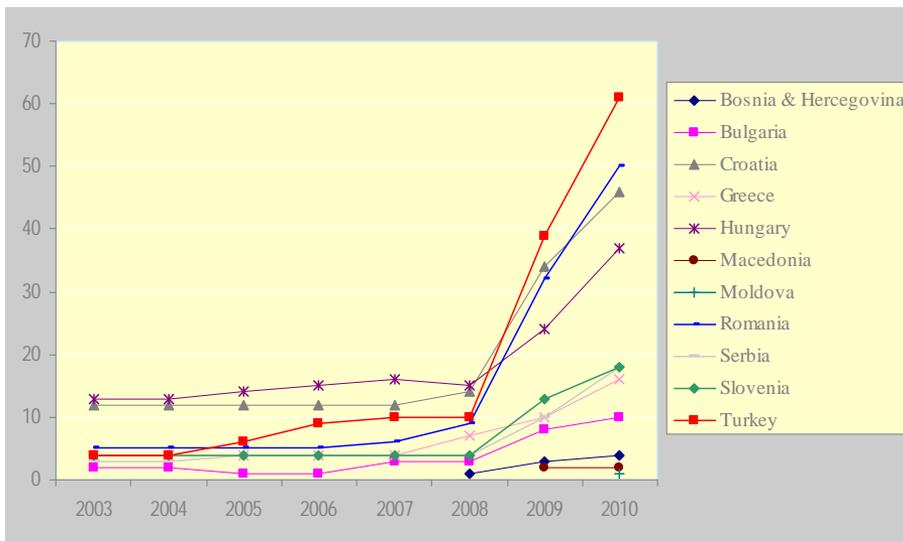


Figure 1. Number of JCR-indexed journals published in SEE countries

One can see from [Table 1](#) that the changes poorly coincide with the WoS performance of SEE countries. Some countries became visibly overrepresented. Rank correlations between increments expressed in the number of newly added (in 2009 and 2010) journals on one side, and, on the other side, performance expressed as the number of papers, number of citations, and RCIs, are $\rho=0.797$, $p=0.003$; $\rho=0.774$, $p=0.005$; and $\rho=-0.132$, $p=0.698$ respectively. We can see that the correlation with countries' RCI was even negative (although non-significant), suggesting that Thomson's choice of journals was unrelated with the quality of research output of the countries these journals came from. One can argue that the quality of research publishing in a country is not necessarily a direct function of the quality of national research outputs. The average IF of journals already present in WoS could thus be regarded as a better criterion for judging the quality of Thomson's selection decisions. We tested this provisionally by

Table 1. Performance of SEE countries in WoS in comparison with the number of JCR-indexed national journals

	Number of papers 2007-2011	Number of citations 2007-2011	Relative Citation Index	Number of JCR-indexed journals 2008	Number of JCR-indexed journals 2010
Turkey	105,915	301,655	2.85	10	61
Croatia	16,193	55,191	3.41	14	46
Romania	29,568	81,820	2.77	9	50
Hungary	27,792	169,026	6.08	15	37
Slovenia	16,288	73,283	4.50	4	18
Serbia	16,962	51,308	3.02	4	18
Greece	51,445	290,671	5.65	7	16
Bulgaria	11,176	47,998	4.29	3	10
Bosnia	1,906	3,745	1.96	1	4
Macedonia	1,111	3,751	3.38		2
Moldova	1,161	4,761	4.10		1
Albania	410	972	2.37		
Cyprus	3,085	14,448	4.68		
Montenegro	559	1,165	2.08		

calculating rank correlation between the size of increment and the average IF of national journals in 2008, i.e. at the time preceding the journal's admission in JCR. The resulting value of $\rho=0.214$ ($p>0.10$), although obtained for only 7 countries under study (Šipka, 2012), reassured us that factors other than countries' research and publishing potential played a decisive role in Thomson's selection process.

In [Table 2](#) results showing the effects of correction for the contribution of national journals to the countries' performance are presented. The data

Table 2. WoS results and rankings (in brackets) of SEE countries before and after correction for contribution of national journals to the countries' output

	Number of Papers			Citations			RCI		
	in WoS	after correction	Δ	in WoS	after correction	Δ	in WoS	after correction	Δ
Albania	410 (14)	410 (14)		972 (14)	972 (14)		2.37 (12)	2.37 (13)	-1
Bosnia	1,906 (10)	1,244 (10)		3,745 (12)	3,407 (12)		1.96 (14)	2.74 (12)	2
Bulgaria	11,176 (8)	9,136 (8)		47,998 (8)	47,040 (8)		4.29 (5)	5.15 (4)	1
Croatia	16,193 (7)	10,362 (7)		55,191 (6)	50,712 (6)		3.41 (7)	4.89 (5)	2
Cyprus	3,085 (9)	3,085 (9)		14,448 (9)	14,448 (9)		4.68 (3)	4.68 (6)	-3
Greece	51,445 (2)	50,087 (2)		290,671 (2)	287,732 (2)		5.65 (2)	5.74 (2)	
Hungary	27,792 (4)	24,667 (3)	1	169,026 (3)	164,979 (3)		6.08 (1)	6.69 (1)	
Macedonia	1,111 (12)	1,048 (12)		3,751 (11)	3,694 (11)		3.38 (8)	3.52 (10)	-2
Moldova	1,161 (11)	1,086 (11)		4,761 (10)	4,718 (10)		4.10 (6)	4.34 (8)	-2
Montenegro	559 (13)	559 (13)		1,165 (13)	1,165 (13)		2.08 (13)	2.08 (14)	-1
Romania	29,496 (3)	15,988 (4)	-1	81,820 (4)	70,229 (5)	-1	2.77 (11)	4.39 (7)	4
Serbia	16,962 (5)	13,672 (6)	-1	51,308 (7)	48,449 (7)		3.02 (9)	3.54 (9)	
Slovenia	16,288 (6)	13,738 (5)	1	73,283 (5)	71,717 (4)	1	4.50 (4)	5.22 (3)	1
Turkey	105,380 (1)	91,233 (1)		301,655 (1)	289,208 (1)		2.86 (10)	3.17 (11)	-1

Δ = rank changes

suggest that the effect of the exclusion of national journals from the analysis on the national publication rate is noticeable, but did not reflect in the country rankings, apparently due to the large differences among countries. The correction operation resulted in only two pairs of countries exchanging their positions on the ranking list. The effects of the correction on the country ranking in the citation rate are even smaller. In this case, only Romania and Slovenia exchanged their positions on the list. The lack of changes can again be attributed to the distribution of the variable under observation.

Changes, however, became obvious, and even huge, when rankings on RCI were observed. This is not surprising since RCI, as a normalized variable, is much more stable and more convenient for the comparisons of entities varying in broad range on the raw bibliometric indicators. What is more important here is that shifts in RCI rankings are interpretable. Countries that profited the most from the Thomson campaign of WoS enlargement, such as Croatia and Bosnia, and especially Romania, suffered the largest drop in their rankings in the period succeeding enlargement.

Apparently, some countries would have had higher rankings if they had not been so well represented in WoS. This is particularly true for Romania, which is ranked 11th, with its RCI being hardly better than the one calculated for countries of a slight tradition in science and minor R&D capacity. Romania's rank calculated to present the hypothetical situation where it is not represented in WoS at all (7th) is in much better agreement with the results of previous studies comparing the WoS output of SEE countries ([Šipka, 2009](#)).

Results shown in [Table 3](#) shed additional light on the consequences of enhancing WoS coverage in SEE. One can see that the biggest deficit in performance quality (RCI) has been suffered by those overrepresented countries that publish journals of lower internationality and impact (IF). The only exception from this rule is Turkey. Turkey benefits from overrepresentation in spite of its journals showing the lowest internationality

and average IF. This can be attributed to the high self-citation rate of Turkish journals.

Table 3. Changes in countries' ranks due to correction in three indicators of performance in comparison with national journals' internationality and IF

	Rank changes in number of papers	Rank changes in number of citations	Rank changes in RCI	Inter-nationality of national journals	Ranking in internationality	Average IF of national journals	Ranking in average IF
Romania (60)	-1	-1	4	40.25	10	0.647	3
Croatia (61)			2	47.28	8	0.354	7
Bosnia (4)			2	54.16	7	0.224	9
Slovenia (26)	1	1	1	43.17	9	0.376	6
Bulgaria (11)			1	58.55	5	0.274	8
Greece (17)			0	87.35	1	1.109	1
Serbia (20)	-1		0	57.99	6	0.821	2
Hungary (41)	1		0	77.01	2	0.567	4
Turkey (75)			-1	20.65	11	0.382	5
Macedonia (2)			-2	67.19	4	0.202	10
Moldova (1)			-2	70.36	3		11
Albania			-1	-	-	-	-
Montenegro			-1	-	-	-	-
Cyprus			-3	-	-	-	-

Discussion

Some limitations of this study have to be raised straightforwardly. Some factors that could have additionally shaped the relationships under study were not controlled. One such factor is definitely the self-citation rate of journals. Also, journals' internationality should have been measured more rigorously, by more indicators, as for example in a study by [Calver et al. \(2010\)](#). In future studies the field representation of national journals in WoS should be taken into account, since Impact Factor and RCI are field-dependant. Mapping instrumental citing, plagiarism, and other forms of

misconduct in regional WoS journals, both newly- and previously-accepted, would also shed more light and help in the evaluation of the 2005-10 Thomson extension campaign. The widespread denial of misconduct prevents the full understanding of the complex relationships of journal publishing and national R&D strategies of small and developing countries. Here, important aspects of the context surrounding publishing in such environments remained in the shadow and should be brought to light in future studies.

Still, we can equally straightforwardly state that at least some findings are convincing and well worth noticing. First of all, it is clear that the choices Thomson Reuters made while extending the share of SEE journals in WoS/JCR are, to say the least, questionable: (1) journals from different countries are not evenly represented in the contingent, and (2) the differences are huge and do not correspond with the differences in national WoS performance. There seem to be two plausible explanations for these findings: (1) some countries are given an advantage on grounds other than their R&D potential (“Mother Thomson has some pets”), and (2) journal publishers and/or editors from some countries are more skilful in struggling with WoS admission criteria, or are more motivated to win them, i.e. more persistent in applying (“Mother Thomson’s babies who cry more, get more food”).

Whatever be true, over-representation brings problems to local evaluators, both researchers and decision-makers. Evaluators at the science periphery benchmark their national STI outputs not against outputs of highly developed nations, nor against world averages, but first and foremost against the outputs of their neighbours. Knowing the “situation on the ground”, they are reluctant to accept the results of benchmarking tools such as the Essential Science Indicator if they are aware that the performance of some countries they compare themselves with is inflated, thanks to the enormous number of journals covered by WoS/JCR. These journals are treated by ESI as international, but are practically national, since they predominantly publish papers by authors of national affiliations. The results of this study should be taken as a warning to all evaluators. They are in full accordance with [Basu’s \(2010\)](#) findings. The results are also consistent with the findings of [Zitt et al.’s](#) seminal study providing evidence of the

drastic changes in national output as a function of the Thomson journal set (“perimeter”) extension, which led them to state that “benchmarking indicators are sensitive to the delineation of perimeters, so which database (or version thereof) is more appropriate for the construction of international benchmarking indicators?” ([Zitt et al., 2003, p. 279](#)).

Secondly, the results indicate that the over-representation of some countries in WoS/JCR may have paradoxical effects. Ironically, it is against the long-term interest of the over-represented countries themselves. Gains in national productivity, achieved through the contribution of domestic journals that reached WoS, are nullified by losses on an indicator which is more important for national STI development, and that is the Relative Citation Index. It seems that over-representation as a sort of reward is not rewarding at all (“Mother Thomson feeds her pet children with fast food”). The results warn against the pressure which is made by editors from developing countries on the Thomson Reuters admission process. WoS/JCR can indeed be understood as the final destination of national journals’ development and a reason for the self-satisfaction of their editors. However, WoS should not be perceived as the Olympic podium or even a social club for the privileged, but rather a world league of teams able to stand the challenges of high professionalism and constant confirmation (“Mother Thomson’s children should learn for life not for school”). The results directly imply that making real use of being indexed by WoS/JCR asks for a certain capacity of all players, from the journal’s publisher to the so called journal’s author base. Funding institutions of SEE countries, in their policies of supporting national journals to reach WoS, should be very careful not to make them bite off more than they can chew.

Finally, the over-representation of some countries presumably puts Thomson Reuters’ interests at risk. The role of Thomson Reuters’ products in developing countries is more reformatory than regulatory. On a social terrain lacking evaluation culture, the emancipation role is of essential importance. This is where the special responsibilities of Thomson’s decision makers come from. At risk is also the validity of ESI and JCR as evaluation tools. Both are already insensitive to differences among entities of low performance. Additional deterrence of evaluators and decision mak-

ers in developing countries from using such products is maybe not a threat to Thomson Reuters' operating income, but it definitely is for its mission in developing countries.

Under-representation of some parts of the world in Thomson databases can be made up for by developing regional or national citation databases. Building such databases to serve as add-ons to (then) ISI indexes is an old idea of Garfield, coming fully true only these past years with a dozen of such products, from less successful European trials to, going eastward, Turkish, Iranian, Indian, Thai, Chinese, and Taiwanese national citation databases. Experience with SCIndeks, the Serbian Citation Index ([Šipka, 2005](#)), which has been in use as an evaluation tool for over a decade now, clearly shows that this approach is complementary to WoS, as well as feasible and sustainable.

On the other hand, the damage from the over-representation of some regions or countries cannot easily be made up for. Opening the WoS door too wide eventually leads to the further contamination of the international research area, already polluted by redundant and low-quality papers published increasingly in prestigious international journals under ever-growing pressure to publish. Hopes that journals of dubious quality will, after being "let in", exploit their WoS status to catch-up with international quality and standards may prove to be justified, but also elusive. Counting on the more favourable among the two outcomes does not seem to be more than a game of coin tossing.

References

- Basu, A. (2010). Does a country's scientific 'productivity' depend critically on the number of country journals indexed? *Scientometrics*, 82(3), 507-516. [doi: 10.1007/s11192-010-0186-8](#).
- Calver, M., Wardell-Johnson, G., Bradley, S., & Taplin, R. (2010). What makes a journal international? A case study using conservation biology journals. *Scientometrics*, 85(2), 387-400. [doi:10.1007/s11192-010-0273-x](#).
- Marušić, A., & Marušić, M. (1999). Small scientific journals from small countries: Breaking from a vicious circle of inadequacy. *Croat Med J*, 40(4), 508-14. [pmid:10554353](#).

- Sambunjak, D., Ivaniš, A., Marušić, A., & Marušić, M. (2008). Representation of journals from five neighboring European countries in the Journal Citation Reports. *Scientometrics*, 76(2), 261-271. doi:10.1007/s11192-007-1915-5.
- Shelton, R.D., Foland, P., & Gorelsky, R. (2009). Do new SCI journals have a different national bias. *Scientometrics*, 79(2), 351-363. doi:10.1007/s11192-009-0423-1 .
- Testa, J. (2006). The Thomson Scientific Journal Selection Process. *International Microbiology*, 9, 135-138. Retrieved from <http://revistes.iec.cat/index.php/IM/article/viewFile/4c457c8c230e1.002/9558>
- Testa, J. (2011). The globalization of Web of Science 2005-2010. Retrieved from <http://wokinfo.com/media/pdf/globalwos-essay.pdf>, 2012 Apr 1.
- Thomson Reuters (2011). 2010 Journal Citation Reports. Retrieved from <http://admin-apps.webofknowledge.com/JCR/JCR?SID=Z2DL9JlI8gEMBgnFKLm>, 2012 Apr 1.
- Thomson Reuters (2012). Source publication list for Web Of Science, Science Citation Index Expanded, Social Science Citation Index, Arts and Humanities Citation Index, March 2012. Retrieved from http://ip-science.thomsonreuters.com/mjl/publist_sciex.pdf, 2012 Apr 1.
- Šipka, P. (2005). The Serbian citation index: Context and content. In: *Proceedings of the 10th Conference of the International Society for Scientometrics and Informetrics*, Stockholm, Sweden, pp. 710-711. Retrieved from http://ceon.rs/pdf/Sipka_SCIndeks_proceedings.pdf
- Šipka, P. et al. (2009, June). *Naučni učinak Srbije u 2008 godini: Uporedna analiza na osnovu bibliometrijskih pokazatelja, [Research output of Serbia in 2008: A comparative analysis based on bibliometric indicators]*. Belgrade: Centre for Evaluation in Education and Science for the Ministry of science and technology development.
- Šipka, P. (in press). Bibliometric quality of Serbian journals 2002-2011: More than just a dress for success, In: P. Šipka (Ed.) *Journal Publishing in Developing, Transition and Emerging Countries: Proceedings of the 5th Belgrade International Open Access Conference 2012*, Belgrade, Serbia, May 18-19. Retrieved from <http://boac.ceon.rs/index.php/BOAC/12/paper/view/110/7>
- Zitt, M., Ramanana-Rahary, S., & Bassecoulard, E. (2003). Correcting glasses help fair comparisons in international landscape: Country indicators as a function of ISI database delineation. *Scientometrics*, 56(2), 259-282. doi:10.1023/A:1021923329277.

Towards Automated Enrichment of Scientific Publications with Complex Meta-data

Aleksandar Kovačević¹, Zora Konjović² and Goran Nenadić³

¹ kocha78@uns.ac.rs

University of Novi Sad, Faculty of Technical Sciences, Trg Dositeja Obradovića 6,
21000 Novi Sad (Serbia)

² ftn_zora@uns.ac.rs

University of Novi Sad, Faculty of Technical Sciences, Trg Dositeja Obradovića 6,
21000 Novi Sad (Serbia)

³ g.nenadic@manchester.ac.uk

University of Manchester, School of Computer Science, Oxford Road, Manchester, M13
9PL (United Kingdom)

Keywords: semantically enriched indexing; methodology mining; metadata extraction; complex-metadata; machine learning; conditional random fields; support vector machines; automatic terminology mining

Introduction

An ever-growing amount of digital scientific resources requiring indexing represents a major challenge for libraries and institutional repositories. The ultimate goal is to enable efficient access and retrieval of scientific papers and other digital content in order to alleviate the extremely time-consuming task of keeping up with the literature in a particular domain. There are two major drawbacks in the process of indexing. Firstly, typically only the manuscript's metadata is stored in structured form and thus easy to access. This implies that in order to gain semantic information about the article (e.g. methodology or data used by authors), the publication itself (often full text or at least an abstract) must be read.

The second drawback is that metadata is typically produced manually by curators, a scarce resource in comparison to the overwhelming quantity of scientific publications.

In this paper we present a system that employs machine learning to extract both metadata and semantic information from scientific publications. We first perform automatic extraction of traditional metadata from the full-text of scientific articles. This information is categorised in eight pre-defined categories: *Title*, *Authors*, *Affiliation*, *Address*, *Email*, *Abstract*, *Keywords* and *Publication Note*. Eight separate *Support Vector Machines* (SVM) models were developed for this task. We then automatically identify the elements of the methodology that was

followed by the authors as presented in their paper. Each methodological mention is categorised in four semantic categories: *Task*, *Method*, *Resource/Feature* and *Implementation*.

Here we present the system overview and the results of applying the proposed methodology on a corpus of publications in the field of Automated Terminology Recognition (ATR).

System overview

The system comprises two parts. The first part recognises traditional metadata, while the second part captures methodological information. Both parts were developed using machine learning.

The task of automatic metadata extraction was formalised as an information extraction (IE) problem and solved by classification. Extraction was performed from full text documents available in the PDF format, which was first converted to HTML in order to capture the five types of different features: formatting, layout, word-specific, named-entities and lexicon ([Kovačević et al., 2011](#)).

We have trained a suite of SVM classifiers to identify the eight categories. Empirical evidence showed that eight separate SVM models performed better at this task than a single 8-class model.

Methodology mining was done at three levels. At the sentence level, sentences were categorised – by following previous work by [Teufel and Moens \(2002\)](#), into seven categories describing *Background*

of the research, its *Aim*, considerations on the *Basis*, *Contrast* to other work, mentions of specific *Other* work, *Textual* structure of the article and *Own*. The *Own* category describes the main work presented in the paper, including method, results, and future work. It further drilled down with the addition of three new sub-categories: *Solution* (presenting methodological sentences), *Result* (results), *Own_Else* (cases that did not belong to the two previous categories). We used a two-level classification model: the first step identified sentences belonging to the *Own* category, whereas the second differentiated between the three sub-categories. Each sentence was represented by seven groups of features following ([Teufel & Moens, 2002](#)). The *naive Bayes* model provided best results for this step.

Identification of methodological segments was accomplished by treating each *Solution* sentence as a sequence of syntactic phrases and applying *Conditional Random Fields* (CRF) models. CRFs were used because they have been successful in sequence tagging. Each of the four semantic classes of methodology mentions (*Task*, *Method*, *Resource/Feature*, and *Implementation*) was recognised by a separate CRF model. The feature set representing each phrase consisted of five major groups ([Kovačević et al., 2012](#)), capturing various information such as: lexical, syntactic (both shallow and deep), semantic, citations, etc.

Gold-standard corpus

The system was evaluated on a manually annotated corpus comprising 100 publicly

available full text articles from the field of ATR obtained by searching the Web and various conferences. All papers were manually annotated with eight pre-defined metadata categories. A subset of 45 papers was chosen randomly and annotated with methodological information at three levels. At the sentence level, sentences were first categorised in one of the seven categories from (Teufel & Moens, 2002), and sentences belonging to the *Own* category were further categorised into three sub-categories: *Solution*, *Result*, *Own_Else*. Methodology mentions in *Solution* sentences have been annotated at the segment level with four semantic categories (*Task*, *Method*, *Resource/Feature*, or *Implementation*). Document-level annotations were created by manually collapsing all segment mentions that referred to the same methodological entity (e.g. the same *Task* or the same *Method*).

Results

We measured the performance of our system in terms of F-measure using standard five-fold cross-validation. The results for eight SVM models, used for metadata recognition, were generally good, with the F-measure between 81% (*Keywords*) and 99% (*Title*) (Kovačević et al., 2011b), indicating that this type of task can be successfully solved by applying machine learning. The results for methodological sentence recognition were similar to those presented by Teufel and Moens (2002), with an F-measure of 83% for the *Own* and 78% for the *Solution* category. The CRF models for methodologi-

cal segment categorisation achieved moderate results. The highest F-measure was achieved for the *Implementation* category with 75% (86% precision) and *Resource/Feature* with 61% (67% precision). The models for *Method* and *Task* mentions had a slightly lower F-measure of 53% (with 70% precision) (Kovačević et al., 2012). Document-level results revealed that terminological variability still presents a challenge for systems such as ours. When term variations were collapsed and methodological segments considered only at the document level, the results generally improved by 10%, with an F-measure of 72% (81% precision) for the *Task*, 60% (81% precision) for the *Method*, 74% (78% precision) for the *Resource/Feature* and 79% (81% precision) for the *Implementation* category.

In order to gain some insight about methodologies used in the ATR domain, we applied the system to all of the 100 ATR papers (Kovačević et al., 2012). The top ten most frequent methodology mentions are shown in Table 1.

With the goal of improving the understanding of the used methodologies, we have further analysed the co-occurrences between task, method, resource and implementation mentions and the contents of metadata. We only considered metadata categories that contained information relevant to the article's content i.e., *Title* and *Abstract*. The *Keywords* class, although relevant, was omitted since less than 1% of the corpus had this information (Kovačević et al., 2011).

Table 1. Top ten methodology mentions for the whole gold-standard corpus.

Task	Method
<i>term recognition</i>	<i>part-of-speech tagging</i>
<i>classification</i>	<i>morphological analysis</i>
<i>pattern matching</i>	<i>syntactic parsing</i>
<i>similarity calculation</i>	<i>genetic algorithm</i>
<i>frequency analysis</i>	<i>stemming</i>
<i>clustering</i>	<i>corpus analysis</i>
<i>dictionary construction</i>	<i>lexical lookup</i>
<i>rule learning</i>	<i>statistical method</i>
<i>disambiguation</i>	<i>suffix checking</i>
<i>ontology construction</i>	<i>manual annotation</i>
Resource/Feature	Implementation
<i>corpus frequency</i>	<i>Perl</i>
<i>contextual information</i>	<i>Access</i>
<i>syntactic patterns</i>	<i>Lucene</i>
<i>linguistic rule</i>	<i>Conexor parser</i>
<i>dictionary</i>	<i>Chasen system</i>
<i>termhood</i>	<i>Stanford parser</i>
<i>ontology</i>	<i>Atract</i>
<i>lexicon</i>	<i>Fastr</i>
<i>umls</i>	<i>Sylex</i>
<i>similarity measure</i>	<i>Xtract</i>

The results revealed that virtually in all cases the abstract of the paper contained at least one *Task* and *Resource/Feature* mention and in 86% of the articles one *Method* mention. *Implementations* are only mentioned in 13% of the abstracts. Based on this, it seems that if the goal is

to extract the main tasks, methods and resources it is enough to process the abstract. The title of the paper can also be very useful in determining the main task (70% of the corpus has one task mentioned in the title), and it is a good indicator of resources (56%). As expected, methods (16%) and implementations (3%) are seldom mentioned in the title. Our analysis also confirmed that processing just the abstract of the article is not enough to grasp its full methodology. On average, less than half of all of the tasks in one paper (43%) are mentioned in the abstract. This percentage is considerably lower for resources (26%) and methods (25%). Only 7% of the implementations that authors use are contained in the abstract.

Conclusion

We have developed and evaluated a method for automated enrichment of scientific publications with complex meta-data. The proposed method suggested that machine learning methods can be used to automatically enrich scientific publications with both traditional and more complex meta-data. Further work is needed on automated clustering of methodological mentions at the document level.

Acknowledgements

The work presented here was partially supported by the Serbian Ministry of Education and Science (projects III44010, III44006).

References

- Kovačević, A., Konjović, Z., Milosavljević, B., & Nenadić, G. (2012). Mining methodologies from NLP publications: A case study in automatic terminology recognition. *Computer Speech & Language*, 26(2), 105-126. [doi:10.1016/j.csl.2011.09.001](https://doi.org/10.1016/j.csl.2011.09.001).
- Kovačević, A., Ivanović, D., Milosavljević, B., Konjović, Z., & Surla, D. (2011). Automatic extraction of metadata from scientific publications for CRIS systems. *Program: Electronic library and information systems*, 45(4), 376-396. [doi:10.1108/00330331111182094](https://doi.org/10.1108/00330331111182094).
- Teufel, S., & Moens, M. (2002). Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status. *Computational Linguistics*, 28(4), 409-445. [doi:10.1162/089120102762671936](https://doi.org/10.1162/089120102762671936).

Origin of Plagiarized Manuscripts Submitted in the Croatian Medical Journal

Ksenija Baždarić¹, Lidija Bilić-Zulle² and Mladen Petrovečki³

¹ ksenija.bazdaric@medri.uniri.hr

Rijeka University School of Medicine, Department of Medical Informatics, Braće Branchetta 20, 51000 Rijeka (Croatia)

² lidija.bilic.zulle@medri.uniri.hr

Clinical Hospital Centre Rijeka, Clinical Department of Laboratory Diagnostics, Krešimirova 42, 51000 Rijeka (Croatia)

³ mladen.petroveckii@medri.uniri.hr

Dubrava Clinical Hospital, Department of Clinical Laboratory Diagnostics, Avenija Gojka Šuška 6, 10000 Zagreb (Croatia)

Keywords: detection; plagiarism; origin; similarity; software; text

Introduction

The introduction of plagiarism detection software began at the end of the 20th century with the simple off-line software that compares given documents and calculates text similarity rates ([Faidhi & Robinson, 1987](#); [Bilić-Zulle et al., 2005](#)). Over the years, plagiarism detection has become more sophisticated and software that compares a given document with texts found on the Internet has been developed ([Bilic-Zulle et al., 2005](#); [Garner, 2011](#)). Plagiarism detection software that compares texts to documents in large libraries of published scientific material was introduced a few years ago, with [eTBLAST](#) and [CrossCheck](#) being the most frequently used ([Garner, 2011](#); [Baždarić et al., 2012](#); [Baždarić, 2012](#)).

There are many reasons for plagiarizing, but those most mentioned in the literature are: pressure to publish, lenient attitudes towards plagiarism, limited English proficiency and low or no skill in scientific methodology ([Baždarić et al., 2012](#); [Bilic-Zulle, 2010](#); [Mavrinac et al., 2010](#); [Wager, 2011](#)). The aim of our study was to detect the origin of plagiarized manuscripts submitted for publishing in the Croatian Medical Journal during 2009-2010.

Method

Software

All of the manuscripts submitted in the Croatian Medical Journal (CMJ; www.cmj.hr; IF 1.46 for 2009-2010) from January 1, 2009 to December 31, 2010 were first analyzed using the plagia-

rism detection software [eTBLAST](#) and [CrossCheck](#) with the iThenticate algorithm. The text similarity was validated with the [WCOPYfind software](#).

Procedure

The submitted manuscripts with more than 10% of text derived from previously published paper(s) were suspected of plagiarism and analyzed in detail, following the previously published algorithm for the plagiarism detection procedure and text similarity analysis ([Bazdaric et al., 2012](#)). The results were manually verified as plagiarized or non-plagiarized by the investigators, following [CMJ's Guidelines to authors](#) and [COPE's flowcharts](#). The origin of the manuscripts was recorded as corresponding to the author's current affiliation. Countries were categorized as emerging and developed economies according to the International Monetary Fund's World Economic Outlook Report.

Statistical analysis

The results are presented as frequencies and relative frequencies, and distributions are compared with the Chi square test and the test of proportions. A p value <0.05 was considered to be statistically significant. The MedCalc statistical software was used to analyze the data (version 11.2.0.0, MedCalc Inc., Mariakerke, Belgium).

Results

There were 754 manuscripts submitted over a two year period, and 105 (14%) of the manuscripts were suspected of plagiarism through automatic software detec-

tion. Out of these, 85 (11%) were determined to be plagiarized after manual verification.

Most of the plagiarized manuscripts were submitted from emerging or developing economies (78 (92%) vs. 7 (8%), $p<0.001$; ([Table 1](#)). The proportion of plagiarized manuscripts is greater than the proportion of non-plagiarized manuscripts for manuscripts from China, India, Nigeria and Libya.

Discussion

The authors of plagiarized manuscripts were mostly from emerging economies. English is commonly their second language and, presumably, they have limited English proficiency, which is considered to be one of the main reasons for plagiarism ([Bazdaric, 2012](#); [Bilic-Zulle, 2010](#); [Wager, 2011](#)). The other reasons may lie in the fact that those authors may have more lenient attitudes towards plagiarism – they live in communities with cultural values that tolerate plagiarism more than the societies of developed economies ([Bazdaric et al., 2012](#); [Mavrinac et al., 2010](#); [Pupovac et al., 2010](#)). In order to prevent plagiarism, scientific editors should have strict and clear instructions about plagiarism in their journals, and educate authors about scientific methodology and research integrity.

Finally, all submitted manuscripts should be checked for plagiarism because plagiarism detection software with manual verification is a valuable procedure for assisting with the prevention of plagiarism and achieving high quality among

Table 1. Plagiarized and non-plagiarized submitted manuscripts according to the country of origin

Country	P	non-P	Statistics <i>p</i>
<i>Emerging economies</i>			
China	18 (21)	57 (9)	0.001
Turkey	16 (19)	91 (14)	0.285
Croatia	12 (14)	165 (25)	0.035
India	5 (6)	10 (1)	0.002
Iran	4 (5)	33 (5)	0.791
FYRM	3 (4)	16 (2)	0.431
Nigeria	3 (4)	3 (0)	<0.001
Serbia	3 (4)	33 (5)	0.892
BIH	2 (2)	14 (2)	0.680
Egypt	2 (2)	8 (1)	0.780
Hungary	2 (2)	9 (1)	0.780
Libya	2 (2)	0 (0)	0.014
Romania	2 (2)	9 (1)	0.780
Albania	1 (1)	5 (1)	0.563
Brasil	1 (1)	9 (1)	0.563
Jordan	1 (1)	2 (0)	0.383
Lebanon	1 (1)	0 (0)	0.383
<i>Developed economies</i>			
Italy	3 (4)	6 (1)	0.077
Slovenia	2 (2)	33 (5)	0.336
Spain	1 (1)	4 (1)	0.563
US	1 (1)	17 (3)	0.480
Other countries	0 (0)	145 (22)	-
Total	85 (100)	669 (100)	-

Legend: P – plagiarized, non-P – non-plagiarized

scientific journals. Prejudice towards authors from emerging countries should be avoided and thorough precautions are needed when evaluating papers for publishing.

Acknowledgements

This paper presents results from the research project “Prevalence and attitudes towards plagiarism” supported by the Croatian Ministry of Science Education and Sports and the Committee on Publication Ethics (COPE). Part of the project results have been previously published ([Bazdaric et al., 2012](#)).

References

- Baždarić, K. (2012). Plagiarism detection: Quality management tool for all scientific journals. *Croat Med J*, 53(1), 1-3. [doi:10.3325/cmj.2012.53.1](#).
- Baždarić, K., Bilić-Zulle, L., Brumini, G., & Petrovečki, M. (2012). Prevalence of Plagiarism in Recent Submissions to the Croatian Medical Journal. *Sci Eng Ethics*, 18(2):223–239. [doi:10.1007/s11948-011-9347-2](#).
- Bilić-Zulle, L., Frković, V., Turk, T., Ažman, J., & Petrovečki, M. (2005). Prevalence of plagiarism among medical students. *Croat Med J*, 45(1), 126-131. [pmid:15726686](#)
- Bilić-Zulle, L. (2010). Responsible writing in science. *Biochem Medica*, 20(3), 279-281. [doi:10.11613/BM.2010.034](#).
- CrossCheck software. Retrieved from <http://www.ithenticate.com/>, 2012 Apr 17.

- CMJ's Guidelines to authors. Retrieved from http://neuron.mefst.hr/docs/CMJ/issues/2012/53/1/CMJ_53%281%29_GUIDE-LINES.pdf, 2012 Apr 17.
- COPE's flowcharts. Retrieved from http://www.publicationethics.org/files/u2/All_flowcharts.pdf, 2012 Apr 17.
- eTBLAST software. Retrieved from <http://etest.vbi.vt.edu/etblast3/>, 2012 Apr 17.
- Faidhi, J.A.W., & Robinson, S.K. (1987). An empirical approach for detecting program similarity and plagiarism within a university programming environment. *Computers & Education*, 11(1), 11-19. doi: [10.1016/0360-1315\(87\)90042-X](https://doi.org/10.1016/0360-1315(87)90042-X).
- Garner, H.R. (2011). Combating unethical publications with plagiarism detection services. *Urologic Oncology: Seminars and Original Investigations*, 29(1), 95-99. doi: [10.1016/j.urolonc.2010.09.016](https://doi.org/10.1016/j.urolonc.2010.09.016).
- International Monetary Fund's World Economic Outlook Report (2012). Retrieved from <http://www.imf.org/external/pubs/ft/weo/2012/01/pdf/text.pdf>, 2012 Apr 17.
- Mavrinac, M., Brumini, G., Bilić-Zulle, L., & Petrovečki, M. (2010). Construction and Validation of Attitudes Toward Plagiarism Questionnaire. *Croat Med J*, 51(3), 195-201. doi: [10.3325/cmj.2010.51.195](https://doi.org/10.3325/cmj.2010.51.195).
- Pupovac, V., Bilic-Zulle, L., Mavrinac, M., & Petrovecki, M. (2010). Attitudes toward plagiarism among pharmacy and medical biochemistry students: Cross-sectional survey study. *Biochem Medica*, 20(3), 307-313. doi: [10.11613/BM.2010.039](https://doi.org/10.11613/BM.2010.039).
- Wager, E. (2011). How should editors respond to plagiarism?: COPE discussion paper. Retrieved from http://publicationethics.org/files/COPE_plagiarism_discussion%20doc_26%20Apr%2011.pdf, 2012 Mar 20.
- WCopyfind software. Retrieved from <http://plagiarism.bloomfieldmedia.com/z-wordpress/software/>, 2012 Apr 17.

Challenges for an Emerging New Scientific Journal in Romania

Marius Drugaș¹, Gabriel Roșeanu² and Elena Bonchiș³

¹ mariusdrugas@gmail.com

University of Oradea, 1 Universității St. Oradea (Romania)

² gabiroseanu@gmail.com

University of Oradea, 1 Universității St. Oradea (Romania)

³ ebonchis@uoradea.ro

University of Oradea, 1 Universității St. Oradea (Romania)

Keywords: education and research; Romania; scientific journals

Introduction

In 2006, the Romanian president Traian Băsescu stated in a speech held at the opening of the academic year in Brașov that the Romanian education system is mediocre (as cited in [Frangopol, 2006](#)). One year later, the Vice President of the National Council of Scientific Research, Professor Daniel David PhD, stated that the general performance of Romanian education and research is mediocre, arguing that none of the Romanian universities were ranked among the first 500 internationally (as cited in [Frangopol, 2007](#)). There were only 16 Romanian journals indexed in the ISI-Web of Science.

The international position of Romanian publications did not provide grounds for satisfaction, but for concern. Things slowly started to change. Recently Romania has evolved towards performance-based funding of university research and education; this process is a response to

the knowledge economy, new public management and a desire for research excellence ([Hicks, 2012](#)).

One of the highly rated criteria refers to articles published in journals indexed in the ISI-Web of Science. As a result, the pressure on academic researchers greatly increased, and the number of Romanian ISI indexed journals grew to 58 (“[Romanian ISI Journals](#)”, 2012). Most of them are focused on Mathematics and Chemistry, and this is consistent with the findings of Glanzel et al. 2006, as cited in ([Schultz & Manganote, 2011](#)), which identify a predominance in Chemistry and Physics and less activity in life sciences as a common pattern for former socialist countries.

Currently there is only one Romanian journal in the field of Psychology indexed in the ISI-Web of Science. The pressure from researchers in this field on emerging Psychology journals is very strong. Still,

a great number of factors determine the success of journals: the prestige of the editorial board, the publishing conditions, the quality of received articles, the quality of the review process and so on.

The focus of our study was the case of *The Romanian Journal of School Psychology* (RJSP), a semi-annual peer-reviewed journal that published its first issue in 2008. The journal is supported by The National Association of School Psychologists in Romania and it considers articles from the field of school/educational psychology or related fields for publishing. Articles that are accepted for publication may be included in one of the four sections of the journal: studies and research, interviews, reviews, scientific life.

Considering the presented situation of Romanian scientific journals, we aimed to point out some of the key issues that new scientific journals (in the domain of Psychology) encounter within the first years of publication.

Methodology

Sample

A total of 487 articles from four different journals were used. The focus of our study was the case of *The Romanian Journal of School Psychology* (RJSP), a total of 84 articles within 8 issues. These were compared to three other journals: *Journal of Cognitive and Behavioral Psychotherapies* (JCBP) (the only Romanian journal on Psychology indexed in the ISI-Web of Science: 100 articles within 7 issues), *Journal of Psychological and*

Educational Research (JPER) (182 articles within 20 issues), and *Psychology of Human Resources* (PHR) (121 articles within 16 issues). These journals were selected due to the fact that they are psychology journals that have been at an international level for some time and that they are also edited in Romania.

Procedure

The articles from the different journals were analyzed with regard to (a) type (theoretical or research), (b) origin of the author (e.g. Romanian –local, Romanian – different city, foreigner), (c) author affiliation (e.g. faculty member, non-faculty member, mixed), and (d) number of references.

Results

Comparisons between the RJSP and the other international level journals indicate that a new scientific journal faces challenges on multiple levels.

Table 1. Comparison between the analyzed journals with regard to article type.

Journal	Article type	
	Theoretical	Research
RJSP	53,6%	46,4%
JCBP	30,0%	70,0%
JPER	39,6%	60,4%
PHR	50,4%	49,6%

$$\chi^2(3)=14,32; p<0,01; \text{Cramer's } \phi=0,17$$

It is clearly seen from the results (Table 1) that most articles published in a scien-

tific journal that has recently been established will be of theoretical nature. Researchers prefer to publish their work in a more developed journal that is perceived as more sophisticated.

Table 2. Comparison between the analyzed journals with regard to the origin of the author.

Journal	Origin of the author			
	R-L	R-DC	F	C
RJSP	29,8%	42,9%	27,4%	,0%
JCBP	43,0%	12,0%	33,0%	12,0%
JPER	61,5%	21,4%	13,7%	3,3%
PHR	21,0%	47,1%	24,4%	7,6%

R-L = Romanian, local; R-DC = Romanian, different city, F = Foreigner, different nationality, C = Collaboration: Romanian - Foreigner

$$\chi^2(9)=90,5; p<0,01; \text{Cramer's } \phi=0,24$$

With regard to the origin of the author, it is observed (Table 2) that more developed journals publish articles that are written in collaboration, between Romanians and foreigners. Also, Romanians who are not local have a higher chance of publishing their work in a newer (recently established) journal.

In all cases the majority of articles are written by faculty members (Table 3). This is not surprising as research activity is an intrinsic part of being a faculty member. Still, it is observed that the more developed journals are also approached by researchers who are not faculty members.

Table 3. Comparison between the analyzed journals with regard to author affiliation.

Journal	Author affiliation		
	F	C	NF
RJSP	73,8%	26,2%	,0%
JCBP	81,0%	12,0%	7,0%
JPER	75,8%	18,1%	6,0%
PHR	68,1%	31,9%	,0%

F = Faculty member, C = Non faculty member in collaboration with a faculty member, NF = Non faculty member

$$\chi^2(6)=26,34; p<0,01; \text{Cramer's } \phi=0,16$$

Table 4. Comparison between the issues of the RJSP and the JPER with regard to the number of references.

Journal		Sum of Squares	df	Mean Square	F	sig.	η_p^2
RJSP	Between Groups	1827.98	7	261.14	1.44	.20	-
	Within Groups	11784.2	65	181.29			
JPER	Between Groups	5430.59	19	285.82	1.72	.03	0,16
	Within Groups	26638.8	161	165.45			

For our last analysis (Table 4), regarding the number of references, we only compared the articles from the RJSP and the ones from the JPER. It seems that in the case of the more developed journal (JPER) the number of references differs from issue to issue. This is probably due to the fact that this journal preferentially publishes research articles.

Conclusions

Our results clearly show that new scientific journals face difficulties on several levels: (a) attracting researchers for publishing research studies; (b) attracting foreign researchers who collaborate with those from Romania; and (c) attracting non-faculty member researchers. There is great pressure for increasing the quality of Romanian journals, which could insure indexing in the ISI-Web of Science.

Acknowledgements

The authors wish to thank the National Association of School Psychologists in Romania for the support for participating in the 5BIOAC 2012.

References

- Frangopol, P.T. (2006). The Romanian higher education between mediocrity and competitiveness. *Journal of Science Politics and Scientometrics*, 4, 195-204.
- Frangopol, P.T. (2007). The Romanian research at crossroads, the dangers of mediocrity. Solutions, challenges, questions. *Journal of Science Politics and Scientometrics*, 25, 65-71.
- Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, 41(2), 251-261. doi: [10.1016/j.respol.2011.09.007](https://doi.org/10.1016/j.respol.2011.09.007).
- Romanian ISI Journals (2012). Retrieved from http://www.cnscis.ro/userfiles/file/CENAPOSS/A_2011_AIS.pdf, 2012 Mar 20.
- Schultz, P.A., & Manganote, E.J.T. (2011). Revisiting country research profiles: Learning about the scientific cultures. *Scientometrics*, 1-15. doi: [10.1007/s11192-012-0696-7](https://doi.org/10.1007/s11192-012-0696-7).

Publishing Scientific Journals in the Conditions of a Transition Society

Živan D. Živković¹ and Ivan Mihajlović²

¹ zzivkovic@tf.bor.ac.rs

University of Belgrade, Technical faculty in Bor, Management Department, Vojske Jugoslavije 12, 19210 Bor (Serbia)

² imihajlovic@tf.bor.ac.rs

University of Belgrade, Technical faculty in Bor, Management Department, Vojske Jugoslavije 12, 19210 Bor (Serbia)

Keywords: citations; impact factor; knowledge; ranking; scientific journals; Serbia

Introduction

The real challenges in the process of publishing and managing a scientific journal in the Serbian society in transitional conditions are described in this paper. The subject of the paper is approached from a practical viewpoint, considering the fact that the authors of this paper are in the editorial teams of two successful scientific journals: the Journal of Mining and Metallurgy – Section B – Metallurgy (www.jmmab.com) and the Serbian Journal of Management (www.sjm06.com). Both journals are published by the Technical faculty in Bor, which is part of the University of Belgrade.

The Journal of Mining and Metallurgy (JMM) is an international journal indexed and registered in the ISI Web of Knowledge database (SCIE), with an impact factor $IF = 1.294$ for the year 2010. This journal was established in 1965, as an international journal for the theory and practice of mining and metallurgy.

The other journal published by the Technical faculty in Bor is the Serbian Journal of Management. This Journal is not yet indexed in the ISI Web of Knowledge, since it was established in 2006, and it can still be regarded as a new journal.

Criteria for scientific journals' ranking in domestic scientific bodies

The Journal of Mining and Metallurgy, Section B, Metallurgy was accepted by the ISI Web of Knowledge database in the year 2007. Until then, its IF was constantly increasing, as presented in [Figure 1 \(SCImago, 2007\)](#). The SJR indicator measures the scientific influence of the average article in a journal. Cites per Doc (2y) measures the scientific impact of an average article published in the journal, computed using the same formula as for the journal's impact factor TM (Thomson Reuters).

The impact factor of this Journal is 3.342, calculated according to the Serbian Cita-

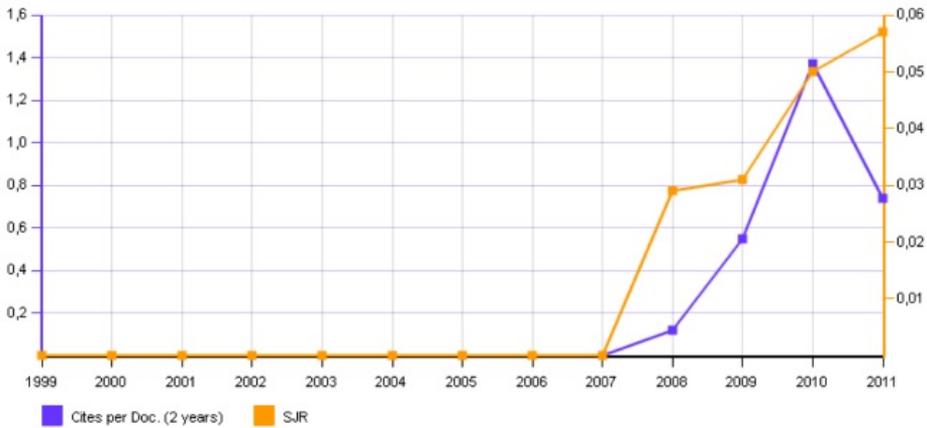


Figure 1. SJR indicator vs. cites per doc (for 2 years).

tion Index developed by the Centre for Evaluation in Education and Science (CEON/CEES), Serbia. This journal has always been highly ranked by CEON’s bibliometric report, since the year 2002 when this measurement was established for the first time.

However, according to the Master Scientific Board for Materials and Chemical Technologies of the Ministry of Science of Serbia, this journal was classified as an M51 journal (leading national journal), until obtaining the international impact factor in 2007. Only after being internationally recognized by the ISI Web of Knowledge, it obtained the national classification of M23 (international journal of a domestic publisher). Now, it is ranked as M21 (top-level international journal) for the year 2011.

The situation is not different for the Serbian Journal of Management, which is also published by the Technical faculty in

Bor. According to CEON’s bibliometric report, this journal should be ranked as M24 (journal of international importance) in the field of social sciences. The journal has an impact factor of 0.892 (for the year 2010), according to the Serbian Citation Index. According to the citations and the bibliometric quality, it is ranked 23rd out of 379 journals published in Serbia (in all scientific fields). Also, the SJM is third in the list of journals in the field of social sciences. However, this journal is ranked as M51 (leading national journal) by the Master Scientific Board for Social sciences of the Ministry of Science of Serbia, for the year 2011. The journal is placed on the 28th position in the list of journals in the field of social sciences, behind the journals which have 10 to 20 times less citations. Also, some of the journals placed in front of the Serbian Journal of Management, on this list, are highly irregular in regards to publishing periods.

The question that should be propounded here is what is the exact criterion for ranking the journals by the Master Scientific Boards in Serbia? On their report on journals' ranking there is no quantifier which could lead to the applied criterions.

The other question is why do the Master Scientific Boards not use the reports generated by CEON and the Serbian Citation Index? This is the only real measure of the journals' quality, because it ranks the journals according to the number of citations, the constancy of publication, and the quality of the publishing procedure. All of these criterions are assessed on an annual basis and readily available at CEON's web site.

Since this journal (Serbian Journal of Management) is also in the procedure of obtaining an international impact factor, which will happen for certain in the next 2 to 3 years, the authors of this paper believe that its adequate national ranking will not happen until then. This way, the pattern for the Journal of Mining and Metallurgy is again repeating. The authors of this paper can only express their regret for this sad situation. Namely, the fact is that the qualities of our journals are better and more quickly recognized by international than by national scientific bodies.

All over the world, the ranking of scientific journals is facilitated mainly by the citations and the impact factor. In our scientific society, the ranking still depends on the personal opinions of the members

of the Scientific Boards. We can only hope that once our society finishes its transition, the process of journal ranking will be the same as in other developed scientific societies.

Hyper production of manuscripts

Another issue that will be addressed here is the situation, also characteristic for transition societies, which can be presented as the hyperproduction of manuscripts after obtaining the IF in some journals. Namely, after becoming included in the SCI or SCIE list, the number of issues of some journals rapidly increases. Also, the number of papers per issue is usually larger than before obtaining the IF. The reason for such development is usually the fact that, after obtaining the IF, journals start to charge the authors for publication fees. Since it is an IF indexed journal, many authors are ready to pay for manuscript publication. The refereeing process of the manuscripts in such cases is usually not too demanding. This leads to a strong decrease in the journal's quality and usually such journals lose their position on the SCI or SCIE list in the next 2 to 3 years. Besides the obvious decrease in quality, the reason for being removed from the SCI list is in the fact that the increase of manuscripts published annually leads to the decrease of the impact factor of the journal, since the number of citations is not high enough to sustain the number of published papers. The mechanism behind the SCI

and SCIE lists is adequate to track such journals and to remove them from the list. However, since many journals from transitional societies behave according to such a pattern, Thompson Reuters and the ISI Web of Knowledge have now increased their criteria for including new journals in their lists.

Stop the “copy and paste” methodology

The process of accreditation of the Serbian universities introduced some new standards in rating the scientific institutions and the scientists. One of the obvious criteria was the number of publications in internationally recognized scientific journals. The accreditation process revealed that many Serbian faculties did not have any employees with such publications (Arsić et al., 2012). This is especially the case in the field of social sciences. Having in mind that the accreditation process will be repeated cyclically in 4 year long intervals, Serbian scientists are now constrained to writing and publishing papers in international journals. Considering the fact that it is not that easy to write a high quality paper for those who do not have previous experience, many authors have turned to the “copy and paste” methodology. This of course led to a large increase of plagiarism, especially in domestic scientific journals.

Luckily, the Centre for Evaluation in Education and Science (CEON/CEES), Serbia, developed an e-Ur service, which is a system for online journal editing and publishing, which includes almost all of the most important Serbian journals. One of the services which are included in this system is *iThenticate*. The most important aim of this service is to identify plagiarism in manuscripts submitted for publication. This service will for certain largely decrease the level of plagiarism in domestic scientific journals.

Conclusion

This paper presents a realistic overview of the present situation with domestic scientific journals, operating in the conditions of a transition society. The theses presented in this manuscript are based on the real experiences obtained during long years of editorial practice in two domestic scientific papers.

References

- Arsić, M., Milijić, N., Živković, D., Nikolić, Đ., & Živković, Ž. (2012). The analysis of scientific-research work of group of technical faculties of Belgrade University in the post-accreditation period. *Serbian Journal of Management*, 7(1), 9-24. doi: [10.5937/sjm1201009A](https://doi.org/10.5937/sjm1201009A).
- SCImago (2007). SJR - SCImago Journal & Country Rank. Retrieved from <http://www.scimagojr.com>, 2012 Feb 20.

Democratisation of Knowledge through Open Access Journals and Repositories in Transition Countries. Case study: Singidunum Journal of Applied Sciences and Singipedia

Gordana Dobrijević¹ and Svetlana M. Stanišić²

¹gdobrijevic@singidunum.ac.rs

Singidunum University, Danijelova 32, 11000 Belgrade (Serbia)

²sstanic@singidunum.ac.rs

Singidunum University, Danijelova 32, 11000 Belgrade (Serbia)

Keywords: articles; data; Google; information; Open Access; portals; publishing; quality; science

Introduction

Open access archives and journals are evidence that the world is moving in the direction of the democratisation of information and knowledge by removing access restrictions in the form of copyright protection or fee-based dissemination policies (Yiotis, 2005). This accelerates research, enriches education, and shares learning across rich and poor nations (Prosser, 2003).

The number of journals added to the Directory of Open Access Journals in the period from 2002-01-01 till 2012-01-01 is 7,363, containing 744,106 articles (DOAJ, 2012). Full text articles originate from 117 countries, with researchers from the USA, Brazil and the United Kingdom contributing the largest number of articles.

Pros and cons of Open Access

Solomon and Björk (2012) have surveyed 1,038 authors who recently published articles in order to determine the factors influencing their choice of a journal. The conclusion was that quality and speed of publication were the most important factors in the respondents' choice of a journal, while open access was a less important but still significant factor.

Clearly, authors want to be read and cited by as many people as possible. In this regard, freely available articles are more widely cited. The resulting success benefits the journal, the journal's prestige and the publisher (Morris, 2003). According to Willinsky (2006), journal publishers, including big corporate players like Elsevier, Springer, and Sage, recognized the contribution that open access archives are making and now grant authors permission to post their work on such sites.

Although the results of the Study of Open Access Publishing (SOAP) show that around 89% of the researchers consider open access publishing beneficial for their research field, most of them still do not consider it as a common practice. A possible reason for that is the lack of suitable quality journals in the particular discipline (Dallmeier-Tiessen et al., 2011). Besides, according to Mann et al., (2009), results of a similar study show that the majority of the survey participants fear that open access publishing might endanger their chances of promotion and decrease chances for access to research funds.

Understandably, many publishers are unwilling to provide free access to a repository that is expensive to create. In addition, preserving digital repositories presents a major challenge due to changes in technology (Dewatripont et al., 2006).

Singidunum Journal of Applied Sciences and Singipedia

The aim of our paper is to emphasize the importance of open access journals for knowledge sharing and transfer in a transition country such as Serbia, using the examples of the Singidunum Journal of Applied Sciences and Singipedia, a scientific online portal & repository, both open access. In June 2010 Singidunum University founded Singipedia, an open access scientific online portal & repository, based on vBulletin architecture. Its aim is to enable collaboration in writing research papers, forming research teams and acquiring necessary literature. Within the first year the portal won the IT

Globus award for its contribution in the area of informatics and computer science. It contains a wide range of materials, including university textbooks, working papers, published articles, teaching materials, students' master and PhD theses, etc. Among these, it also includes the Singidunum Journal of Applied Sciences (earlier Singidunum Scientific Review).

Data showing the total number of views, downloads, registered users, and visitors, were retrieved from Google Analytics. From June 2010 to December 2011 there were almost 2.7 million visits and almost 400,000 downloads on Singipedia, among them more than 82,000 visits to Singidunum Scientific Review. As shown in [Table 1](#), 136,799 visitors or 71.12% were from Serbia, while 28.12% were from abroad (109 countries), the majority of them from the region of former Yugoslavia (ex-Yu).

Table 1. Singipedia and Singidunum Scientific Review during 2010/2011.

Portal/Journal: type of access	Total
Singipedia: views total	2,689,457
Singipedia: downloads	394,777
Singipedia: registered users	3,929
Singipedia: visitors total	191,132
Singipedia: visitors from Serbia	136,799
Singidunum Scientific Review: views	82,290
Singidunum Scientific Review: downloaded articles	31,246

As the Singidunum Scientific Review was published in Serbian, with only a few articles published in English, the target group was limited to domestic and regional (ex-Yu) researchers and PhD students. However, as data showed that 30% of the visitors of Singipedia and the Singidunum Scientific Review were from abroad, and as it was noticed that the rate of downloading the articles written in English was higher compared to those written in Serbian, the Editorial Board of Singidunum Scientific Review decided to change their publishing policy. For this purpose the title was changed to the Singidunum Journal of Applied Sciences and a new web site was also registered in March 2012. Apart from the March issue, the Journal will be published exclusively in English. Also, the reviewing process has become faster and has improved considerably. In addition, the entire Singipedia portal has been optimised to be indexed on Google Scholar. As the next step, the focus will be on the quality and relevance of the articles.

Data retrieved from Google Analytics specifically for the purposes of this paper, during the period March 20th-April 20th 2012, show a considerable improvement regarding internationalisation. Namely, as shown in [Table 2](#), the vast majority of visitors come from English speaking countries. Also, during this first month, the total number of visitors to the Singidunum Journal website was 4,632, which is in line with the monthly average for the Singidunum Review on the Singipedia portal.

Table 2. Breakdown of visitors per native language

Visitors' native language	%
English (USA)	74.21
Serbian (Serbia)	12.32
English (international)	4.87
English (UK)	4.44
Spanish, German and French	1.3
Croatian	0.72
Other	2.14

Conclusion

Open access journals and repositories increase the visibility and contribution of authors and institutions, and they make knowledge and information available to scholars in developing and transition countries worldwide. Having in mind the requirements of contemporary publishing, the Singidunum Journal of Applied Sciences has introduced fundamental changes in its publishing policy, which lead to the Journal's greater impact and visibility. The most important among them was the change of the language of publication.

References

- Dallmeier-Tiessen, S., Darby, R., Goerner, B., Hyppoelae, J., Igo-Kemenes, P., Kahn, D., et al., (2011). Open Access journals-what publishers offer, what researchers want. *Information Services & Use*, 31, 85-91. [doi: 10.3233/ISU-2011-0624](https://doi.org/10.3233/ISU-2011-0624).

- Dewatripont, M., Ginsburgh, V., Legros, P., Walckiers, A., Devroey, J.P., Dujardin, M., et al., (2006). *Study on the economic and technical evolution of the scientific publication markets in Europe*. European Commission. Retrieved from http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf
- Directory of Open Access Journals, DOAJ (2012). Retrieved from http://www.doaj.org/doaj?func=newTitles&uiLanguage=en&fromDate=2002-01-01+00%3A00%3A00&untilDate=2012-01-01+16%3A20%3A25&orderBy=J.browse_title, 2012 Jan 25.
- Mann, F., von Walter, B., Hess, T., & Wigand, R.T. (2009). Open Access Publishing In Science. *Communication of the ACM*, 52, 135-139. Retrieved from http://openaccess-study.com/Mann_et_al_2008_Open_Access_Publishing_in_Science.pdf
- Morris, S. (2003). Open Publishing: How publishers are reacting. *Information Services & Use*, 23, 99-106. Retrieved from <http://iospress.metapress.com/content/blnkwkj0wvvyjbf8/>
- Prosser, D.C. (2003). Institutional repositories and Open Access: The future of scholarly communication. *Information Services and Use*, 23, 167-170. Retrieved from <http://iospress.metapress.com/content/b0t9y69afjhylw9y/>
- Solomon, D.J., & Björk, B. (2012). Publication fees in open access publishing: Sources of funding and factors influencing choice of journal. *Journal of the American Society for Information Science and Technology*, 63(1), 98-107. doi:10.1002/asi.21660.
- Willinsky, J. (2006). Access to Power. *Harvard International Review*, 28, 54-57. Retrieved from http://hir.harvard.edu/academy-and-policy/access-to-power?quicktabs_blogs=1&quicktabs_most_popular=1
- Yiotis, K. (2005). The Open Access Initiative: A New Paradigm for Scholarly Communications. *Information Technology & Libraries*, 24, 157-162. doi: 10.6017/ital.v24i4.3378.

Bibliometric Quality of Serbian Journals 2002-2011: More Than Just a Dress for Successⁱ

Pero Šipka

sipka@ceon.rs

Centre for Evaluation in Education and Science, Kneza Miloša 17, 11000 Belgrade
(Serbia)

Keywords: journals; bibliometrics; references; measuring; indicators; impact; citations; articles; Serbia

Introduction

Within a program sponsored by the government, Serbian journals are routinely evaluated, ranked and “categorized”. Results are published yearly in the Journal Bibliometric Report (<http://scindeks-bic.ceon.rs/eng>) by the Centre for Evaluation in Education and Science (CEON/CEES; <http://ceon.rs/>). Journal ranking is based primarily on impact, i.e. citations received in both the Web of Science (WoS) and the Serbian Citation Index (SCIndeks; www.scindeks.ceon.rs). Impact indicators are supplemented with a group of measures reflecting the formal characteristics of journals. These are derived by quantifying ISI (now Thomson Reuters) qualitative admission criteria, such as the share of references in English, which can be generated with low-cost from the SCIndeks data. These indicators are supplemented with a few classical bibliometric measures, such as the journal-to-monograph ratio, selected mostly on the basis of findings about the quality of Yugoslav journals, later reviewed and systemized by [Maričić \(2002\)](#) and opera-

tionalized by [Šipka \(2004\)](#). Maričić referred to such indicators as “communicability”, which was not broadly accepted. For lack of consensus, in CEON/CEES products they are dubbed “indicators of bibliometric quality” (IBQs), while the term “indicators of impact” was kept for all measures based on the citation rate. Authors studying the quality of journals published in developing countries (e.g. [Faria et al., 2007](#); [Sheibaninia et al., 2010](#)) find such indicators valid and useful. On the contrary, those dealing with reputed indexed journals quite expectedly do not find them useful and sometimes avoid them explicitly (e.g. [Lokker et al., 2012](#)).

The purpose of the original introduction of IBQs in Serbia was to stimulate the process of standardization and internationalization of national research publishing. At a practical level, the purpose was to provide editors with feedback on how close their journals are to the critical level of quality, sufficient for admission to WoS/JCR.

Unlike the JBR Impact Factor, IBQs have never been fully accepted by either decision-makers or journal editors. Informally, most of them referred to such indicators as outward, formal journal features, not having much to do with the real value of journals, or simply as “the suit that doesn't make the man”.

In this study, the results on the IBQs of Serbian WoS-indexed journals gathered during the whole 10-year period of following journals were observed in order to recognize the overall trends of their changes. The trends were compared with those of non-WoS national journals. To complete the picture, they were related to the movement of journals' WoS Impact Factor in the same period. The general purpose was to re-evaluate, after a decade of their continuous use, the face validity and benefituality of IBQs.

Method

Data for all journals (N=417), including newly established and discontinued ones, were extracted from the JBR 2002-2011 editions. All journals covered by WoS in 2010 were treated as WoS-indexed journals (N=20), while the rest were allocated to the group of non-WoS journals.

The two groups of journals were compared by using the following IBQs:

- (1) NumRef: average number of references per article;
- (2) AgeRef: average age of references (citing half-life);

- (3) IntAuth: share (%) of articles by international authors;
- (4) IntLang: share (%) of articles published in English and other “world languages”;
- (5) IntRefs share (%) of international references; and
- (6) JourRef: share (%) of journal references (“journal-to-monograph ratio”);

Repeated measures two-group ANOVA was used to test the differences in IBQs between WoS- and non-WoS-indexed journals. Only 17 WoS-indexed and 111 non-indexed journals with no missing data were included in this part of the analysis.

The Impact Factor of the Serbian WoS-indexed journals was compared with journals of the same status published in other SEE countries. The comparison covered the period from 2002 to 2011. All the data were retrieved from JCR. In order to include as many journals as possible, the Impact Factor calculated for a two year period (IF2) was used. IF results were normalized by calculating the z-values against the mean and S.D. of the group to which a journal was assigned in JCR. For journals classified in more than one JCR group, the z-value was calculated within each group and each year separately. The annual score for each country was expressed as the mean of all z-values of all national journals. Due to the low coverage of journals from most countries, big variations of coverage within countries, and relatively large number of countries

in the sample, using statistical tests in this case was not feasible.

Results and Discussion

Figure 1, displaying the results of Serbian WoS-indexed journals, indicates a slow but steady growth of IBQs during the observed period. The same tendency was registered for non-WoS journals (Figure 2). Results given in Table 1 suggest that the only exception is the Share of international references (IntRefs), which did not change significantly. This could not have been expected, since this indicator already reached the sub-maximal value in WoS-indexed journals almost at the very beginning of the period followed.

The two groups of journals differ significantly on all IBQs, except for Age of references (AgeRefs), which is also the only

variable showing a significant interaction effect. Major differences seem to exist in the Share of journal references (JourRef), as a measure of scientific seriosity, and Share of international references (IntRef), as a measure of internationality. To summarize, WoS-indexed journals are generally superior to non-WoS journals, especially on the measures of internationality. On most IBQs WoS-indexed journals did not reach a plateau, clearly indicating that there is still some room for their improvement.

In Figure 3 all the y-axis values are negative, revealing that regional journals as a group are well below the world average. No sign of them nearing the world IF means can be observed. All that can be seen as a general trend is that the impact level of national journals became more

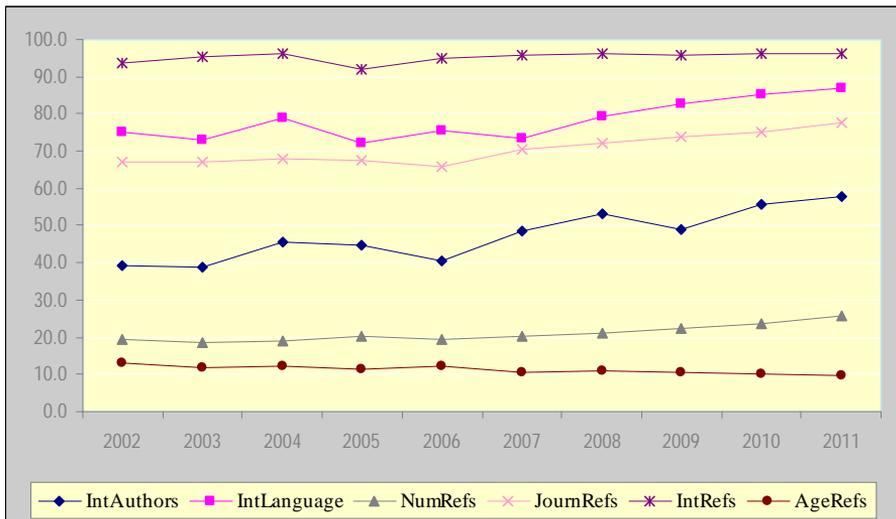


Figure 1. IBQ results for WoS-indexed Serbian journals

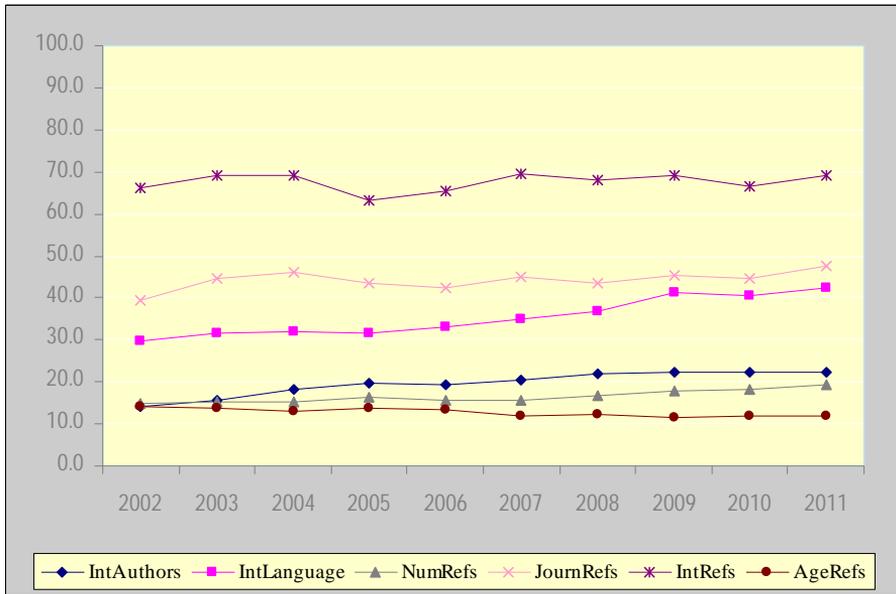


Figure 2. IBQ results for non-WoS Serbian journals

homogenous. Also, rankings changed from year to year. The most remarkable change is the elevation of IF of Serbian journals. We see that during the period under study the previously discussed stable rise of the bibliometric quality of these journals was paralleled by the rise of their Impact Factor. Admittedly, this trend was abruptly terminated in 2008, but this happened under the influence of an external factor, i.e. the extension of WoS coverage to the region, which is explained in detail in another study (Kosonović & Šipka, in press). It should be pointed out that this trend was only broken, not reversed, as was the case with most other SEE countries.

Table 1. ANOVA results for IBQs of Serbian journals

	WoS		years		WoS * years	
	F	p	F	p	F	p
NumRef	9.30	.00*	11.6	.00*	0.56	.83
AgeRef	.692	.41	3.87	.00*	1.97	.04*
IntAuth	18.22	.00*	5.72	.00*	1.45	.17
IntLang	18.76	.00*	3.88	.00*	1.03	.41
IntRefs	25.78	.00*	1.57	.12	0.30	.98
JourRef	28.88	.00*	9.64	.00*	1.50	.14

The increase of the impact of Serbian journals was so rapid that it can even be described as explosive. As a result, a few years ago they surprisingly reached the

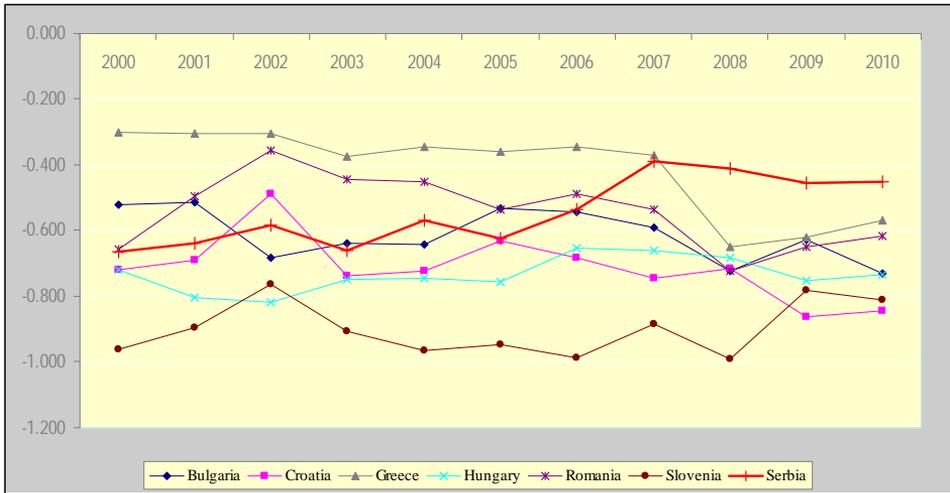


Figure 3. Normalized IF of SEE journals by country

top of the regional ranking list. This finding is in total discrepancy with all other known bibliometric indicators of (absolute) performance of Serbian science. It can of course, at least partially, be attributed to the relative under-representation of Serbian journals in JCR, which makes it less intriguing. Equally plausible, however, would be an explanation calling on the better support these journals were having during the recorded period.

The amount of financial support, almost exclusively provided from the skinny budget of the responsible government ministry, can hardly be recognized as a factor. All that the ministry can be credited for is that its regulations were systematically encouraging the internationalization of Serbian research publishing in the observed period. More credit should be given to the enthusiasm of editors and ed-

itorial boards of the journals in question. While not denying the contribution of such personal and socio-political factors, one can be confident that the program of evaluation and international promotion of Serbian journals, based firmly among else on IBQs' feedback, is the major immediate cause of the fast development of Serbian journals. This of course can be clearly demonstrated only through a complex and rather expensive study. However, it does not seem to be necessary, since coincidental evidences reported here show that the indicators of bibliometric quality are interrelated with the measures of impact in a way that could be understood as synergistic. What IBQs measure seems to be a structural component of the journals' communication potential, and even something that can be treated as a necessary condition for small journals' international excellence, rather than merely "the dress

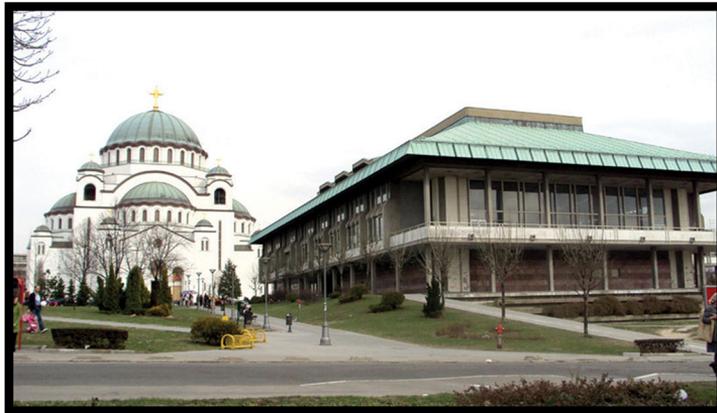
for success” or “the suit that doesn't make the man”. This does not mean that some IBQs cannot be further refined and that the whole battery cannot be strengthened by more up-to-date indicators, such as the Usage (downloading) factor or the Manuscript acceptance rate, which are measured more reliably now than a decade earlier when this battery was composed. Thus, the old suit called IBQs is still comfortable and should be maintained. It can be remade, so why throw it away?

References

- Faria, J.R., Araújo Jr., A., Shikida, C.D. (2007) The citation pattern of Brazilian economists. *Estudos Econômicos*, 37, 151-166. Retrieved from <http://www.scielo.br/pdf/ee/v37n1/06.pdf>
- Kosanović, B. & Šipka, P. (in press) Output in WoS vs. representation in JCR of SEE nations: Does mother Thomson cherish all her children equally? In: P. Šipka (Ed.) *Journal Publishing in Developing, Transition and Emerging Countries: Proceedings of the 5th Belgrade International Open Access Conference 2012*, Belgrade, Serbia, May 18-19. Retrieved from <http://boac.ceon.rs/index.php/BOAC/12/paper/view/111/10>
- Lokker, C., Haynes, R.B., Chu, R., McKibbin, K.A., Wilczynski, N., & Walter, S.D. (2012) How well are journal and clinical article characteristics associated with the Journal Impact Factor? A retrospective cohort study. *J Med Libr Assoc*, 100(1): 28–33. [pmid:22272156](https://pubmed.ncbi.nlm.nih.gov/22272156/). [doi:10.3163/1536-5050.100.1.006](https://doi.org/10.3163/1536-5050.100.1.006).
- Maričić, S. (2002). Scholarly journals from science periphery: Towards a common methodology for evaluating their scientific communicability? *Croatian Information and Documentation Society*, 50(1-2). Retrieved from <http://hdl.handle.net/10150/105820>.
- Sheibaninia, A., Valaiee, N., Tavakoli, B., Azizi, F., Mohamadsadeghi, Sh., & Nourbakhsh, M. (2010) Quality of research articles published in Iranian Dental Journals. *Pak J Med Sci*, 26(4):847-851. Retrieved from <http://pjms.com.pk/issues/octdec2010/pdf/article21.pdf>
- Šipka, P. (2004) Vrednovanje časopisa na osnovu bibliometrijskih pokazatelja - opis modela i implikacije za urednički rad [The evaluation of journals based on bibliometric indicators: Description of a model and implications for editorial work], *Srpski arhiv za celokupno lekarstvo*, 38-49 (special issue). Retrieved from http://ceon.rs/pdf/vrednovanje_casopisa_na_osnovu_bibliometrijskih_pokazatelja.pdf

ⁱ This paper is partly based on the results reported in CEON/CEES Working Paper ERD No 0107/11 available in Serbian at http://ceon.rs/pdf/IF_SEE_casopisa_prema_JCR.pdf

These Proceedings are a record of the 5th Belgrade International Open Access Conference, organized by the Centre for Evaluation in Education and Science.



Centre for Evaluation in Education and Science (CEON/CEES) is a virtual institute gathering experts from various R&D organizations to work together on issues of evaluation in science and higher education. Our primary activities are focused around adjusting, developing, and implementing scientometric indicators suitable for the use in small, developing, transition, and emerging countries. An important part of CEON/CEES's program is reserved for the development of information systems to support decision making in STI sector, as well as quality research publishing based on quality control, monitoring, and evaluation.