A modo de prólogo

The role of information professionals in interpreting the results of journal league lists for research performance evaluation

Péter Jacsó

Nota: Este artículo puede leerse traducido al español en: *http://www.thinkepi.net/Jacso*

Jacsó, Péter (2013). "The role of information professionals in interpreting the results of journal league lists for research performance evaluation". *Anuario ThinkEPI*, v. 7, pp. 11-15.

Abstract: This paper discusses some of the key content and software features which must be analyzed by skilled librarians/information professionals to deliver reliable, accurate, and appropriately comprehensive set of bibliographic data and bibliometric indicators about the status of the journals listed in the promotion, tenure and grant applications, as well as in journal desiderata for purposes of collection development in academic and special libraries.

Keywords: Journal evaluation, Journal league lists, Journal rankings, Citation databases, Inconsistencies, *WoS*, *Web of Science*, *Scopus*, *Google Scholar*, *GS*, *Google Scholar Metrics for Publications*, *GSMP*, *Microsoft Academic Search*, *MAS*, E*l profesional de la información*, *EPI*.

Título: Papel de los profesionales de la información en la interpretación de rankings de revistas para evaluar resultados de investigación

Resumen: Se analizan algunas de las novedades más importantes y las características del software que deberían ser analizadas por los expertos bibliotecarios y profesionales de la información para ofrecer conjuntos de datos bibliográficos e indicadores bibliométricos fiables, precisos y adecuadmente exhaustivos sobre las revistas que aparecen en los listados a tener en cuenta para la promoción académica, la acreditación y solicitudes de becas y ayudas, así como en la atención de las desideratas para el desarrollo de las colecciones de las bibliotecas universitarias y especializadas.

Palabras clave: Evaluación de revistas, Listas de revistas acreditadas, Rankings de revistas, Bases de datos de citas, Incoherencias, *WoS*, *Web of Science*, *Scopus*, *Google Scholar*, *GS*, *Google Scholar Metrics for Publications*, *GSMP*, *Microsoft Academic Search*, *MAS*, *El profesional de la información*, *EPI*.

Introduction

Evaluating the research performance of faculty members, departments, and universities has been traditionally based on the opinions of subject experts about the quantity and quality of published research, through the prism of the perceived reputation of the publishing venues, which have been regularly rated by expert groups into 4-5 tiers/clusters. There is a strong drift away from this approach toward purportedly more objective, transparent, and cost effective ranking of publishing outlets, predominantly academic journals.

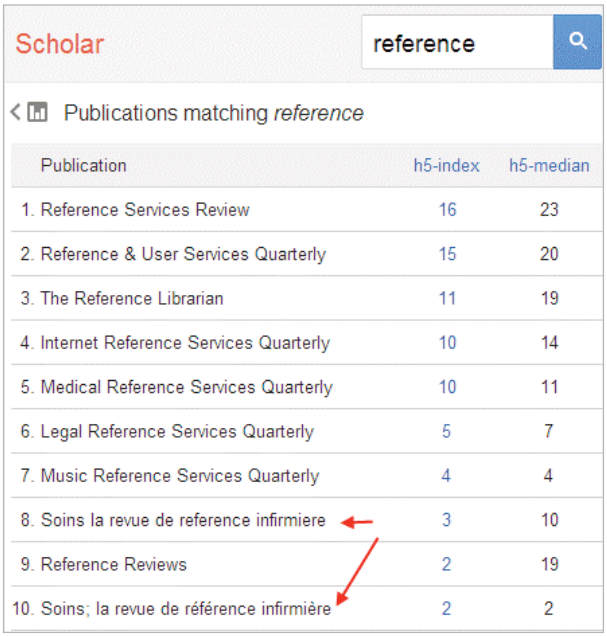
From 2006, there has been a surge in the number of publications about bibliometric databases, and their use in research assessment of individuals, research groups, universities and countries based on the journals where faculty members published and received citations from. In many countries bibliometric services and databases are used to rank journals by their scholarly reputation, and to use their rank positions by various indicators as a proxy in support of decisions in tenure, promotion and grant applications. These databases require knowledge about many factual, bibliographic and bibliometric aspects of

their content and software feature which are not obvious, transparently revealed and sufficiently intuitive for the administrators and subject/disciplinary experts who are involved at universities and research institutions involved in the decision-making process.

In the next few years academic and special librarians will be hopefully involved in educating the administrators and the experts involved in the process about the advantages and the shortcomings of the cited reference enhanced databases which would be used in collecting, corroborating and interpreting the metric indicators about publishing performance. For decades the classic *JCR* (*Journal Citation Reports*) developed by **Eugene Garfield** has been the primary source for collecting and publishing bibliometric information for ranking thousands of journals in nearly 200 subject areas in the Sciences and Social Sciences domains by many indicators of their scholarly familiarity/popularity and influence/impact/prestige.

Now there are several other multi-disciplinary, subscription-based and open access databases enhanced by or created from cited references (such as *Scopus, SCImago, Eigenfactor, Microsoft Academic Search* (*MAS*), *Highwire Press, J-STOR, Google Scholar Metrics for Publications* (*GSMP*) for creating multi-faceted journal league lists from *Web of Science* (*WoS*) *JCR*, *Scopus*, *Google Scholar* (*GS*), *GCrossRef*, and other mega-databases. It is expected by this author that *ProQuest* and *Ebsco* will also release comprehensive journal ranking lists from subsets of from the millions of their records which have been enhanced by cited references and citation counts.

This short paper discusses only some of the key content and software features which must be analyzed by skilled librarians/information professionals to deliver reliable, accurate, and appropriately comprehensive set of bibliographic data and bibliometric indicators about the status of the journals listed in the promotion, tenure and grant applications, as well as in journal desiderata for purposes of collection development in academic and special libraries. The author is committed to publish a full paper in as a case study for discovering, using, calculating, and critically interpreting bibliographic and bibliometric data - using the *SCImago* database. It is a high quality, very well designed, open access digital resource for creating league lists by several indicators focusing on the library and information science (LIS) subject field. Several of these indicators and software features have been analyzed, criticized and advocated by this author for decades in journal articles, conference papers, database reviews, books and book chapters. Many have been created by the best information specialists, or experts



Publication	h5-index	h5-median
1. Reference Services Review	16	23
2. Reference & User Services Quarterly	15	20
3. The Reference Librarian	11	19
4. Internet Reference Services Quarterly	10	14
5. Medical Reference Services Quarterly	10	11
6. Legal Reference Services Quarterly	5	7
7. Music Reference Services Quarterly	4	4
8. Soins la revue de reference infirmiere	3	10
9. Reference Reviews	2	19
10. Soins; la revue de référence infirmière	2	2

Figure 1. Double entry for the same journal

in other disciplinary fields who will be given credit for their contributions in the *EPI* paper.

Eugene Garfield developed his set of bibliometric indicators in order to select the most important journals for his first experimental *Genetics Citation Index* databases in the early 1960s, which demonstrated the power of citation analysis (and indirectly its use as an aid for collection development), the impact factor became a proxy for judging the quality of articles published, irrespective of the fact that the impact factor was meant to express the prestige of the journals based on the average number of citations received by research papers published one or two years before the census year, i.e. in *JCR*-2011 the citations received by papers published in 2009 or 2010. **Garfield** has been well aware of the controversy surrounding the impact factor, and concerned by its misuse, but he passionately stands by his seminal idea of measuring the impact of scholarly journals and rightly so.

The impact factor or the plain citation rate (citations/paper), indeed can be an appropriate indicator for estimating the standing of journals in a given discipline, but not necessarily of the quality of the individual researchers (or in aggregate mode, that of research groups, colleges, universities), whose research papers are cited far less or far more often than the average impact factor of the journal and its rank position would indicate.

In addition, there was another twin-event at the end of 2004, and the publishing of a seminal article in 2005 about a new indicator, which brought the issue of measuring and comparing the quantity and impact of scholarly research, and the ranking of individual researchers, research

groups, departments, colleges and universities, and even nations, in center stage. *Elsevier*, the largest publisher of scholarly journals, announced the launch of its *Scopus* database, a cited reference enhanced multidisciplinary database on a very sophisticated, still user-friendly software platform, and *Google, Inc.* launched the *Google Scholar* database. The license fee of *Scopus* was significantly lower than that of the *WoS* database, and *Google Scholar* was free of charge, at a time when the universal adoration for the *Google* search engine (and brand name) was at its highest level.

The debates about the pros and cons of the impact factors (with a 2-year and 5-year version) versus the h-index, g-index and other important other variants dramatically increased the number of publications about measuring scholarly productivity and impact. The cited references added great value to the traditional bibliographic metadata, and added significant expenses to enhance database records with bibliometric metadata (except for the *WoS* database which was created with such metadata from the beginning). They require even more scrutiny by the librarians and other information professionals than the bibliographic metadata elements, to minimize the consequences of errors of commission and errors of omission or at least alert potential users of cited reference enhanced databases about the scale of problems such as the "metadata mega mess" before the clean-up of *Google Scholar* in 2011.

Not even these steps can guarantee perfectly accurate results and metrics-based ranking lists of researchers, or journals because of the increase in the motivation of the manipulation of bibliometric indicators, fraudulent publications, the "duping" and spamming of open access, autonomously indexed cited reference enhanced databases.

Metrics-based assessment of scholarly publishing productivity and impact and ranking of journals is very likely to stay with us. It may even preempt the traditional perceptions-based ratings of journals which happened in the 2012 round of the *ERA*, the *Excellence in Research for Australia* initiative, which dropped the comprehensive, perception-based rating list of nearly 22,000 scholarly journals created for the *ERA*-2010 round by more than 700 professors and other experts of the disciplines.

Librarians and other information professionals must be aware of the advantages and disadvantages of metrics-based research assessment, and journal rankings, as they will be very much involved in producing metrics-based league lists, as well as interpreting and explaining the results.

Database size, source list and time-span of coverage

Database producers often use the size of a database as a top selling point. In itself it is not a decisive point as there may be millions of skeletal bibliographic records without abstracts and other value added indexing and bibliographic data freely available information, mostly to pad the database. *GS* has been notorious of keeping the size of the database a secret, and *GSMP* follows this tradition, by not providing information about it. From the perspective of journal league lists the size of the database subset enhanced by searchable and adequately tagged cited references and ISSN are the most critical. In some, such as *Scopus* and *SCImago*, the time span is clearly indicated (1996); in *WoS* cited reference enhancement can be taken for granted for all the items –except the ones for articles which did not cite any document. In others the time span of such subset is journal-specific.

The need for searchable, accurate and adequately tagged cited references is obvious as year of cited publication, cited author and journal names, their chronological-numerical designation are needed for correct matching. The cited papers' title is much appreciated by mere mortals, but many scientific journals do not use this data element in the cited references. The ISSN of the citing and cited journals is a key element for uniquely identifying a serial, and for easily searching by it. The ISSN of the cited source would be much appreciated but it is practically never required for the reference list. Most of the databases make available information-rich journal

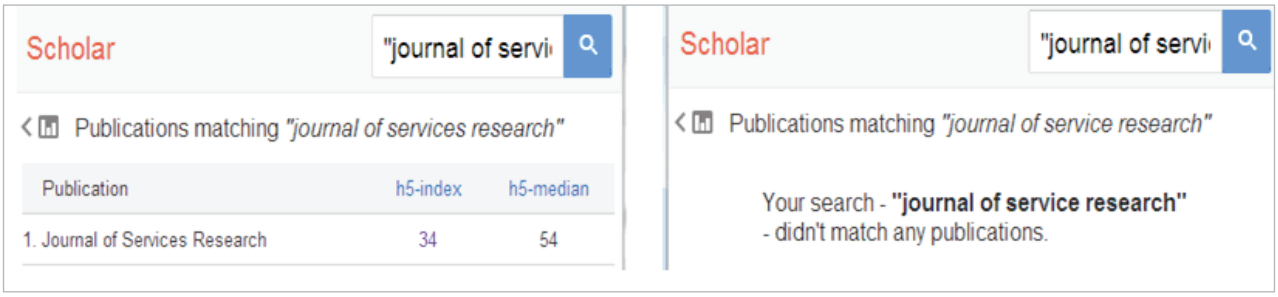


Figure 2. Hi-jacked citations from the *SAGE* publication *Journal of Service Research*

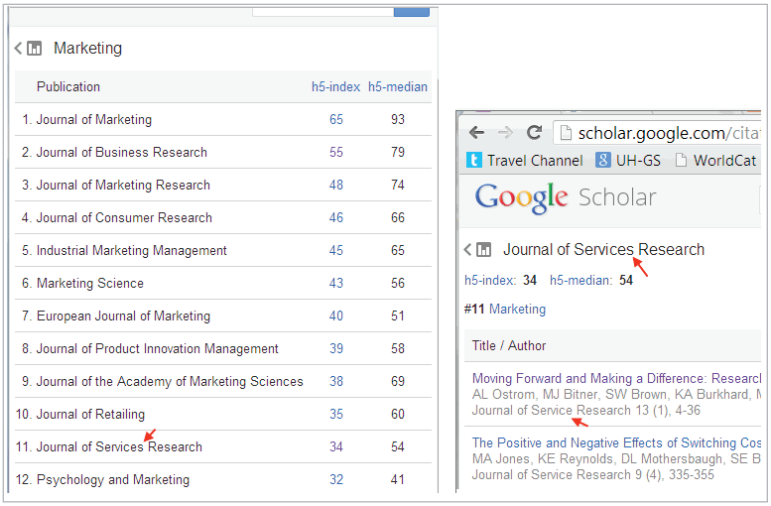


Figure 3. A disservice for *Journal of Service Research* in *GSMP*

lists which include the source journals' title, ISSN, start and end of coverage dates for indexing-only, indexing & abstracting, full-text coverage. Many of the ones with cited reference enhancements also provide the time-span for those types of records by journals.

The lack of ISSN in *GS* as a searchable and displayed data element of source journals is the cause of many of the mis-identified, mis-attributed and duplicate journal entries. Some of these have made it even to *GSMP*, the significantly (but not sufficiently) cleaned up subset of *GS*. The ISSN would make searching for single word titles or titles with common words which are part of many other journal names much more easy without overwhelming the users, such as *Computer* –published by *IEEE*- through its ISSN (0018-9162) -forcing them to wade through a list not sortable by title or any data element in *GSMP*, or exclude those journals which have the target journal title as part of their own title. In a relatively short result list elementary errors and inconsistencies can be easily recognized, such as the French journal appearing twice in a result list. The smart systems such as *SCImago*, *Scopus*, *WoS* and *JCR* designed by developers familiar with the trials and tribulations of searching by short journal names offer the option of exact searching.

The use of ISSN also could have helped in avoiding the misattribution of all the documents and citations of papers published in *Journal of Service Research*, a *SAGE* journal to the much less prominent *Journal of Services Research*. It adds insult to injury that typing the name of the *SAGE* journal returns no hits in *GSMP*.

Ironically, the journal name was correct in *GS*, so the clean-up process for *GSMP* produced its own mess-up, which could have been detected even by minimum-pay hires, or prevented by a little more intelligent programming.

The database time span is usually highly inflated in most databases as the earliest year is often given for the oldest articles even if there are only a handful of them. *GS* does not inform the users about it. Of course the time span of coverage at the journal level varies in every database. The issue in this regard is to learn the time span of coverage of the journals whose bibliographic data and bibliometric indicators must be found. A shorter time span of coverage of some journals than the target window may explain the significantly lower document and citation counts. The adequate time span is not sufficient if the coverage of the journal(s) is gappy. Citations to papers which do not have a master record cannot be "hanged on to" a master record.

Journal sets by topical clusters

The key purpose of any journal league lists is to show the standing of journals among their peers by certain criteria. It is unlikely that any such clustering would please everyone. The compromise is adding many journals to more than one cluster. There is no other option for such journals as *Journal of Educational Psychology*, for example. It is assigned in *GSMP* to Education #1), Educational Psychology (#1), Social Sciences (#11), and Psychology (#16). The number in parentheses indicates the rank position of the journal in the (sub)category. This facilitates to have the right perception about the standing of the journal when judging the publishing activity of a faculty member in the Education versus the Psychology Department.

Some of the categories may have too many journals. It can be unfair, for example, when a marketing journal must compete with journals

LIBR INFORM SCI - Library and Information Science	6	1
J EDUC LIBR INFORM SCI - Journal of Education for Library and Information Science	1	1
PORTAL-LIBR ACAD - Portal-libraries and The Academy	318	0
LEARN PUBL - Learned Publishing	2	0

Figure 4. Absurd entries for quite productive, and widely cited LIS journals in *MAS*

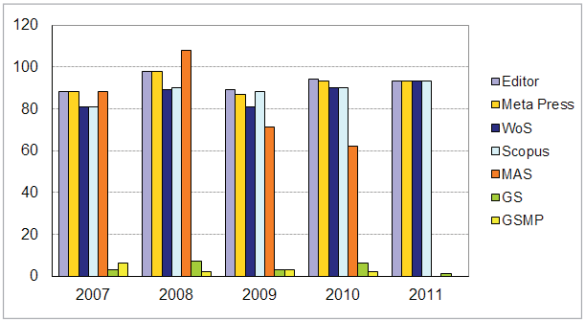


Figure 5a. Absurdly low number of documents in *GS* and *GSMP*

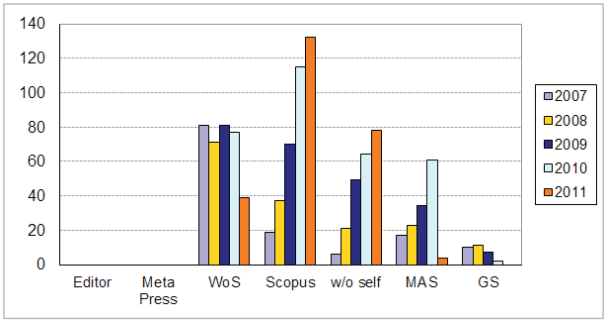


Figure 5b. Absurdly low citation counts for *EPI* in *GS* and consequently *GSMP* throughout 2007-2011

which have far broader subject coverage, and hence have the chance from getting citations from more disciplinary areas. In *Scopus* and *SCImago* the marketing category contains 106 journals, many of them related to administration, governance, general business and management subjects, such as *Administration and Society, Governance*, *Journal of Internet Banking and Commerce*, *Journal of Public Administration Research and Policy*, *Journal of World Business*, *Public Administration Review*.

Some journals don't seem to have any relationship to marketing, such as *Journal of Vinyl and Additive Technology*, *Keramische Zeitschrift*, *Laser Focus World*, *Packaging News*, *PaperBoard Packaging*, *Pulp and Paper International*, *Pulp and Paper*, *Pump Industry Analyst*, *Veterinary Radiology and Ultrasound*, *Veterinary Surgery*, *Weed Research*. On the other hand, the Marketing category does not include, *Journal of Consumer Affairs*, *Journal of Consumer Policy*, *Journal of Personal Selling and Management*, *Services Marketing* which are definitely marketing-related. Some of these are at the very top or in the top-20 positions in the Marketing category.

On the other extreme, *JCR* does not have a category for Marketing, even if it covers dozens of marketing-focused and marketing-related journals. *Microsoft Academic Search* (*MAS*) does not have a category for Marketing either, and the marketing journals are lumped into the huge *Business Administration & Economics* subcategory of 456 journals. This makes it very cumbersome to check a list of 30-35 marketing journals.

The scores and indicators

The primary purpose of any league lists is to prove by scores and derived indicators the standing of journals in their category. Sometimes it backfires directly, and makes the reliability of the whole league list questionable. The number of journals in a category does not necessarily imply good coverage of the journals in a disciplinary area. For example, in the Library Science category, *MAS* includes 72 journals. Some of the entries are very absurd as indicated by the figure below, and do not require any explanation for the readership of the *Anuario ThinkEPI*, just thinking about their implications and consequences. The second column represents the number of documents published in the journals, and the third column shows the number of citations the journal received between 2007 and 2011.

As a final bit of food for thought the Figure 5a and 5b shows the summary raw data for *El profesional de la información* (*EPI*) from various databases used for creating/presenting journal league list information for the 5-year period 2007-2011. I asked the editor to give me the number of papers published in *EPI* in the given time frame as an ultimately reliable base-line. I looked up the journal in various databases, including the host database platform of *EPI: MetaPress*. Figure 5a t shows how many publications were published in the journal according to the editor, and the databases used in the testing. The second one shows the number of citations the journal received year by year in *WoS*, *Scopus* (including and excluding self-citations), *MAS*, *GS* and *GSMP*. The consistently puny numbers for *GS* (which is notorious for including duplicate, and triplicate entries found in the journals' web-site, in subject repositories, institutional repositories, as well as highly inflated citation counts from all types of Internet sources, including lecture notes, reading lists, bibliographies, and student papers.

Considering the world wide adulation of *Google Scholar*, librarians and other information professionals will often have a busy time to explain these serious discrepancies and shortcomings to administrators and faculty members who assume that using *Google, Inc.* data and indicators would be a panacea for learning about the standing of the journals as indirect proof of evidence for quality of research. In a forthcoming paper in *EPI* the details of the most feasible very well designed open access, *SCImago* will be evaluated to demonstrate its power, richness of indicators, appealing design and to make some recommendations for new features.

Creación de Archivos y Bibliotecas Virtuales

Desde la digitalización de materiales bibliográficos hasta la asignación de metadatos y su implementación en la red, conforme a la normativa internacional.

Productos para crear Bibliotecas Digitales y Virtuales

DIGIBIB 8.0
Solución avanzada para la creación de Bibliotecas Digitales y la Gestión Bibliotecaria Multilingüe.

DIGIARCH 2.0
Sistema digital de descripción y gestión archivística. Descripción en ISAD(G) y EAD 2.0.

DIGIDIR 2.1
Directorio para Archivos, Bibliotecas y Museos con generación automática de estadísticas y sistemas de información geográfica (GIS).

OAsIs-PMH 2.0
Sistema integrado de recolección de diversos esquemas de metadatos:
• DCMI sin cualificar • MARC 21 • EAD
• SWAP • mod_OAI • Linked Open Data

ADAPTACIÓN A EUROPEANA
Implementación del esquema ESE 3.4.1 (Europeana Semantic Elements) y EDM 5.2.3 (Europeana Data Model) Adaptado a la Agenda Digital Europea 2020.

DIGITALIZACIÓN AVANZADA
Con asignación dinámica de metadatos.

- Recolección en la Web para Entidades e Instituciones de Memoria en OAI-PMH y Dublin Core cualificado con ESE 3.4.1
- Consultoría y mappings a EDM 5.2.3 (Europeana Data Model)
- Bibliotecas digitales que permiten la creación, recuperación y recolección de metadatos (MARCXML, DCMI y RDF y RDFs)
- Archivos Web que facilitan la creación, recuperación y recolección de metadatos (EAD 2.0 y EAC 2010)
- Implementación de la Europeana OpenSearch API
- Adaptación del repositiro OAI para la transmisión de instancias RDF según ORE
- Repositorios Institucionales DIGIPRESV para Preservación Digital a largo plazo mediante PREMIS 2.2 y OAIS ISO 14721
- Intercambio de metadatos en METS 1.9.1 (diferentes Profiles) integrando todos los esquemas de metadatos
- Creación de METSRights para el control de los derechos de autor
- Reconocimiento Óptico de Caracteres OCR y generación dinámica de ALTO (Analyzed Layout and Text Object)
- Generación e integración de registros SKOS mediante MARC 21(Up.13)/RDA
- Creación de eBooks o libros digitales en formatos: ePub y Mobipocket.
- Adaptación de DIGIBIB a Linked Open Data

Validación en el Data Providers de la Open Archives Intiative.
Genera un Sitemap para Google.

ISO 9001 BUREAU VERITAS Certification
Nº 6002188

AENOR Producto Certificado Proyecto I+D+i

C/ Claudio Coello, 123. Madrid. Tel.: 915 81 20 01. digibis@digibis.com

www.digibis.com