CONTRIBUTIONS to SCIENCE, **4** (1): 69-73 (2008)
Institut d'Estudis Catalans, Barcelona
DOI: 10.2436/20.7010.01.37 ISSN: 1575-6343
www.cat-science.cat

The Thomson Scientific journal selection process*

James Testa**
Thomson Scientific, Philadelphia, USA

James Testa began his career at the Institute for Scientific Information (ISI) in 1983 as the Manager of Publisher Relations building and maintaining relationships with the world's scholarly publishing community. Since 1996 he has worked in Editorial Development. As Director of Editorial Development he has focused on guiding a team of information professionals in the selection of the world's most important and influential scholarly literature published in journals, books, and proceedings. Since 1998 this work has expanded to include the evaluation and selection of websites of interest to the scholarly research community. Over the past ten years he has worked closely with publishers and editors in developing countries, particularly Asia and Latin America, guiding them in their efforts to bring their research publications into the international mainstream.

The Web of Knowledge is an electronic platform that consists of important searchable bibliographic databases and key analytical tools such as the Journal Citation Reports (JCR). Several of these bibliographic databases provide extensive coverage of the literature of specific subjects such as the Life Sciences, Agriculture and Forestry, Food Science, Physics and Engineering, Medicine, Behavioral Sciences, and Animal Biology (Figure 1). The international patent literature and conference proceedings are also key components of the Web of Knowledge. In all, this environment covers over 22,000 unique journals, more than 60,000 conferences from 1990 to the present, and more than one and a half million patents.

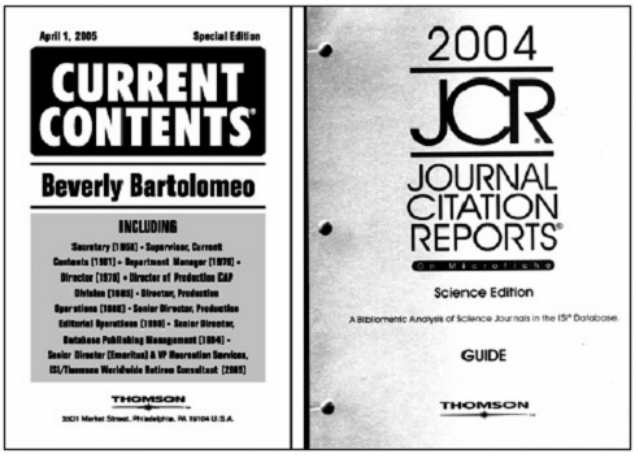

Figure 1. *Current Contents* and *Journal Citation Reports*, two of the publications and bibliographic databases of Thomson Scientific.

The cornerstone of the Web of Knowledge, however, is the highly selective, multidisciplinary Web of Science (WoS) which the Institute for Scientific Information (ISI) and now Thomson Scientific, has developed and maintained for over 40 years. The WoS comprises several multidisciplinary citation indexes. In addition to its three major components, the Science Citation Index Expanded, the Social Sciences Citation Index, and the Arts and Humanities Citation Index, it includes Index Chemicus and Current Chemical Reactions. Since significant resources are dedicated to the selection of the most important and influential journals for their inclusion in the WoS, we will examine the journal selection process used to achieve this goal, particularly with regard to citation analysis and the role of the impact factor.

The Journal Selection Process. The Journal Selection Process has been applied consistently for more than forty-five years. Every journal included in the WoS has been evaluated according to the highest standards, which were originally defined by Eugene Garfield, the founder and president Emeritus of ISI. This steady application of quality has enabled the continuous development of the WoS, whose primary purpose is to provide comprehensive coverage of the world's most important and influential scholarly journals. It would appear that in order to be comprehensive, an index to journal literature should cover all the scientific scholarly journals published. Nevertheless, this approach would not only be economically impractical but unnecessary, since it has been demonstrated that in fact, it is only a relatively small number of journals which publish the bulk of scientific results. An analysis of 7622 journals covered in the 2003 JCR revealed that as few as 300 journals account for more than half of what is cited and nearly one third of what is published. It has also been shown that a core of 3,000 of JCR journals accounts for about 80% of published articles and nearly 95% of cited articles. Journal selection and evaluation

* Based on the Lecture given by the author at the Institute for Catalan Studies, Barcelona, on 25 January 2006.

** Author for correspondence: James Testa, Thomson Scientific, Philadelphia, Pennsylvania (USA). Tel. (001) 215 3860100. Fax (001) 215 3866362. Email: james.testa@thomson.com

are conducted on an ongoing basis at ISI with journals added and deleted from the database as frequently as every two weeks. This core, therefore, is not static; its basic composition changes constantly since new promising journals join it while those which are less useful to the subscribers, disappear. Every year, 2000 journals are evaluated, but only the most important and influential international and regional ones are selected, approximately 10%. Each journal goes through an extensive evaluation process before being selected or rejected. The team of editors who select publications for the WoS have educational backgrounds relevant to their areas of responsibility as well as experience and education in information science. Their knowledge of the literature of their field is extended by consultation with established networks of advisors who participate in the evaluation process when needed, as they review every new journal published in the field and continually re-evaluate the existing coverage. Many factors are taken into account when evaluating journals for coverage, and by combining and interrelating data, the editor is able to determine the journal's overall strengths and weaknesses. There are four main parts to the Thomson Journal Selection process:

Journal Publishing Standards. (a) *Timeliness*: Thomson products are built upon the most recent, up-to-date information. Therefore, to avoid journals that publish old, late, or outdated information, it is essential that each journal publishes according to its stated publication schedule. In order to be considered for coverage in the WoS database, three consecutive issues published on time must be received to confirm this. The same standards hold for electronic journals, although since they normally publish articles one at a time rather than collecting articles for release as an 'issue', editors simply look for a steady flow of articles over several months. (b) *International editorial conventions*: since the point of indexing articles for the Thomson Scientific database is to able to retrieve them in a search, certain editorial conventions are essential to make the articles easier to find. These include: informative journal titles, descriptive article titles, full address information for every author, and complete bibliographic information for all the cited references. (c) *English-language bibliographic information*: English has become the universal language in science at this time in history. It is for this reason that Thomson Scientific tries to focus on journals that publish their full text in English or, at the very least, their bibliographic information in English so that they can reach the widest possible audience. At the moment, there are still many journals covered in the WoS that publish only their bibliographic information in English with full text in another language. However, going forward, it is clear the most important texts for the international research community, and particularly in the natural sciences, will have to have their full text published in English. (d) *Peer review*: We expect peer review to be a part of the publishing process for all science journals. Application of the peer review process is normally done by the journal publisher, and it is an assurance and an indication of a journal's standards, the overall quality of research presented, and the completeness of the cited references.

Editorial Content. Scientific research gives rise to new fields of study, and new journals emerge as published research on a new topic achieves critical mass. The Thomson editor determines if the content of a new journal will enrich the database or if the topic is already covered adequately. First, the editorial content of the journal is examined to determine its possible placement in one or more products. Next, the most appropriate category for the journal in the product is determined. The goal of this process is to ensure that the overall editorial content of Thomson journal products is comprehensive and complete, and that journal coverage in every subject meets the needs of our customers.

International Diversity. The presence of an international base among the contributing authors and the journal's editors and editorial advisory board members is a good indicator of the health of a journal and a predictor of its ability to continue to publish important research in the future. Today's scientific research takes place in a global context and an internationally diverse journal is more likely to have importance in the international community of researchers. Thomson also seeks to cover the best regional journals. All of them have met our selection criteria and represent a sample of the top-quality research from certain developing countries. For example, in the infectious diseases category there are 41 journals published in eight countries—not surprisingly, most are published in the United States and in the European Union. A study at the article level, however, reveals a community of authors representing 115 countries not including the USA or the EU. This is the international face of science that we seek to represent in our Web of Science.

Citation Analysis. Authoritative citation analysis is a key factor in the evaluation of journals. The cited references—to articles in both covered and not covered journals—from all 8,700 journals covered in the Web of Science are captured in order to measure the importance, influence and overall usefulness of a journal. Because the number of journals varies greatly among disciplines, discipline-specific citation rates also vary greatly. Small fields like crystallography or botany do not generate as many articles or citations as do larger fields such as biotechnology or genetics. Likewise, in some areas, particularly in the arts and humanities, it may take a relatively long time for an article to attract a meaningful number of citations, whereas in other areas, such as the life sciences, it is not unusual for citations to occur rapidly and peak after only a few years. These facts must be taken into consideration if citation data are to be used correctly.

Thomson Scientific's citation analysis takes place on at least two levels: we look for citations to the journal itself and we also examine the citation record of the contributing authors and editorial board members. This is particularly useful in evaluating new journals where a citation history does not exist. It helps us understand the extent to which the contributing authors and editorial board members have participated in the world of scholarly communication and is often an indicator of the journal's future success. For established journals, we use journal level citation data, particularly the impact factor—the average number of times recent articles in a journal were cited in a par-

ticular year. For example, in the 2004 the impact factor for the journal *Emerging Infectious Diseases* was 5.643. This means that on average, each article published in 2003 or 2002 received around five citations in 2004 (Table 1). We also note that of the 374 articles published in 2004, 371 were original research and only three were review articles. Clearly this journal publishes mostly original research and would therefore not be compared directly to a journal that published mostly review articles (Table 2).

Self-citation rates are also taken into consideration. The self-cited rate relates a journal's self-citations to the total number of times it is cited by all journals including itself. As with all citation analyses, self-citation rates are considered in the context of other journals of similar editorial scope and format. In the case of *Emerging Infectious Diseases*, 279 out of the 3149 citations to articles published during the impact factor years are self-citations, which results in a quite modest self citation rate of 8.80%. Eighty per cent of all journals listed in the JCR Science Edition have self-citation rates of less than 20%. This shows that self citation is quite normal for most journals and is actually expected. However, significant deviation from this normal rate requires examination to determine if excessive self-citations are being used to artificially inflate the impact factor, thus weakening its integrity. If we determine that self citations are being used improperly then the journal will be considered for de-selection. As governments place more importance on impact factors as indicators of quality, Thomson Scientific is dedicated to insuring that it remains a solid indicator of quality.

Citation indexing is a core competency of Thomson Scientific. Relying on the accumulated experience of 40 years of citation indexing plus the application of technology by its citation experts, the Web of Science can claim a nearly 100% accuracy rate in linking citations to the correct source record. This is possible only because the sheer number of source items and citations in the WoS is so vast that we are able to bring together, in a statistically meaningful way, all the elements necessary to absolutely identify a source publication when it appears as a cited reference. With the release of the Century of Science in 2005, WoS coverage now goes back to 1900. In order to include journal literature from 1900 to 1944 in the WoS file, citation data were absolutely essential in the selection of journals for this project. We identified these journals by isolating the specific articles from the period 1900 to 1944 that had been cited 50 or more times by the journals covered in WoS from 1945 to the present. From that starting point we built a balanced list of journals representing all areas of biomedical sciences, physical sciences and technology.

While it is true that for most journals the peak years of citation impact are within 3–5 years of publication, it is also true that many journals have significantly longer cited half lives—the median age of its articles cited in the current JCR year. Over 1100 journals in the 2004 JCR Science Edition (19%) and 400 journals in the 2004 JCR Social Science Edition (23%) have a cited half-life of 9.5 years, which means that half the citations they receive are to articles no more than 9.5 years old, and that half are older than 9.5. These journals with a long cited half life cover every area of the natural sciences. However, there are many important physics, chemistry, and math journals in this group. The older articles from these journals have proven to be a significant resource for researchers in the physical sciences. Clearly, articles more than ten years old are also valuable for effective research in the Social Sciences as well.

Cited reference searching is perhaps the most significant by-product of the citation index and an indispensable tool in the discovery process. Searching cited references for prior work on a subject is a sure way to avoid redundancy and to effectively build on past research. It is also an indispensable tool in evaluating the work of individual researchers and for measuring the importance or influence of their work in the field.

Table 1. Calculation of the Impact Factor of a journal

Journal: *Emerging Infectious Diseases*	
2004 Impact Factor : 5.643	
Cites in 2004 to articles published in:	2003 = 1577
	2002 = 1572 Total 2 years = 3,49 articles
Number of articles published in:	2003 = 281
	2002 = 277 Total 2 years = 558 articles
Calculation: *Cites to Recent Articles*= Number of Recent Articles	*3149* = **5.643** 558

Table 2. Citable items

	Citable items			*Other items*
	Articles	*Reviews*	*Combined*	
Number in JCR year 2004 (A)	371	3	374	7
Number of references (B)	8041	52	8093	6
Ratio (B/A)	21.7	17.3	21.6	0.9

Citation Indexes

Science Citation Index Expanded and the Science Citation Index. The Science Citation Index Expanded (SCIE) is an international, interdisciplinary index of the literature of science, medicine, agriculture, technology, and the behavioral sciences covering approximately 6200 journals, and available online through the WoS. Every journal in the SCIE has been carefully selected according to the process outlined above. The Science Citation Index (SCI) is a subset of the SCIE that is available on CDROM and print, and has the same editorial scope as the SCIE. However, the SCI comprises approximately 3800 journals, since its coverage is limited to only the most highly cited, high impact journals in each category, at the same time balancing original research and review journals. Geographic origin of the journal is also given consideration. In other words, we try to include the best journal from each country in the SCI, provided it meets all the standard selection criteria.

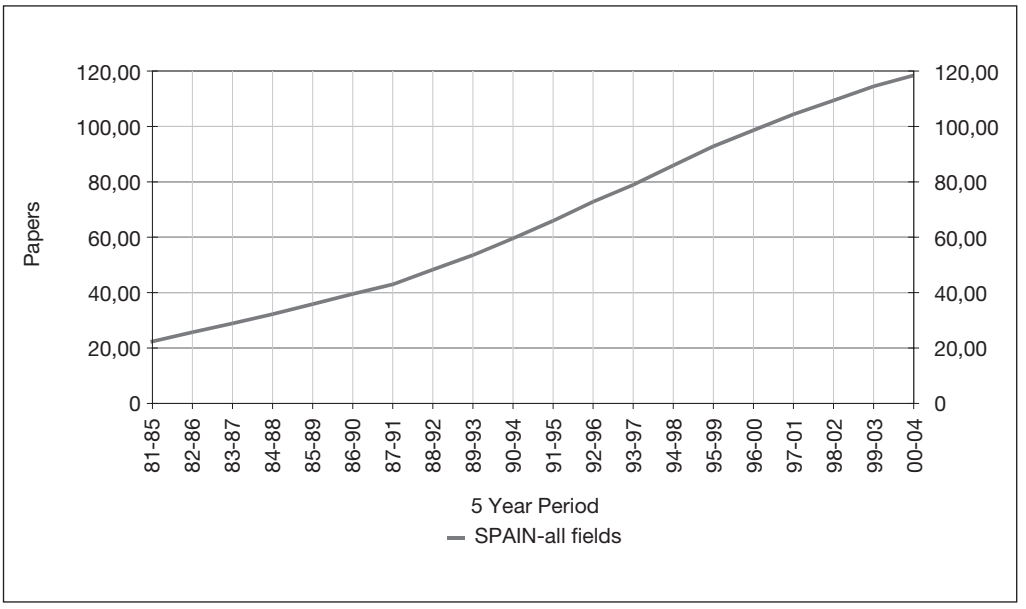

Figure 2. Number of Spanish papers in all the fields. Five-year overlapping periods (1981-2004).

Social Sciences Citation Index. The Social Sciences Citation Index (SSCI) covers 1798 journals. All journals undergo the same thorough evaluation as journals in the natural sciences. Publishing standards, editorial content, international diversity, and citation data are all considered, and standard citation metrics are employed while keeping in mind that, in general, impact factors and overall citation rates in the social sciences are lower than those in the natural sciences and that these tend to rise, peak and decline. Regional studies also have a special importance in the social sciences. In categories such as Education, Economics, International Relations and Area Studies, regional research holds a value not just for the country in which it is produced, but potentially for the wider international community. Coverage of non-English language, regional journals in the Social Sciences Citation Index is becoming more and more important to Thomson Scientific, and we are now in the process of reviewing a number of the best social sciences journals from the European Union including 21 from Spain. Ten of these Spanish Social Sciences journals will be added to the SSCI in the coming weeks.

Art and Humanities Citation Index. The Arts & Humanities Citation Index (AHCI) covers 1123 journals. Although they meet all Thomson Scientific selection criteria, citations in the Arts and Humanities do not necessarily follow the same predictable pattern as citations to social sciences and natural sciences articles: citations tend to experience a cyclical variance depending on scholarly interest in the topic. In addition, arts and humanities journal articles frequently reference non-journal sources (e.g., books, musical compositions, works of art and literature). Consequently, citation data, while sometimes useful, are frequently much less important in journal evaluations in the arts and humanities.

Citation Indexes in Spain. There are currently 53 journals in the WoS that are published in Spain: 32 in the SCIE (5 of which are also in the SCI on CDROM), 19 in AHCI, and 2 in SSCI. Twenty-one more Spanish journals are expected to be added to this figure in the near future, 10 of which will be added during the next few weeks. From 1981 to 2004 there has been a steady stream of articles from Spanish authors in Thomson Scientific products. In 2000-2004, the last five year period measured, Thomson Scientific covered over 118,000 papers by Spanish researchers. The geographic origin of these articles is determined by the author's address regardless of the country of publication of the journal [this information can be found in the National Science Indicators file] (Figure 2).

Open Access Journals

We have recently witnessed a major shift in the landscape of publishing and open access has become a dominant topic. Open access (OA) is simply defined as not charging readers or their institutions for the right to access, download, copy, print, distribute or search an article. (It should be noted that in many cases the cost of producing OA journals is absorbed by the contributing authors who pay a fee to the journal when their papers have been accepted for publication. In these cases the cost of publication has merely shifted from the user to the author). In order to identify OA journals, Thomson Scientific uses a number of sources, including J-Stage <http://www.jstage.jst.go.jo/browse>, the Scientific Electronic Library Online (SciELO) <www.scielo.br>, <wwwscielo.isciii.es>, and the Directory of Open Access Journals (DOAJ) at Lund University <www.doaj.org>. These sources list a more than 2,000 journals, of which there are currently over 260 OA journals that meet the Thomson Scientific selection criteria and are covered in the WoS. These journals are spread across all areas of coverage, though most are in the natural sciences. The number of OA journals is growing rapidly, since new journals are adopting this model and older journals are changing their business model to OA.

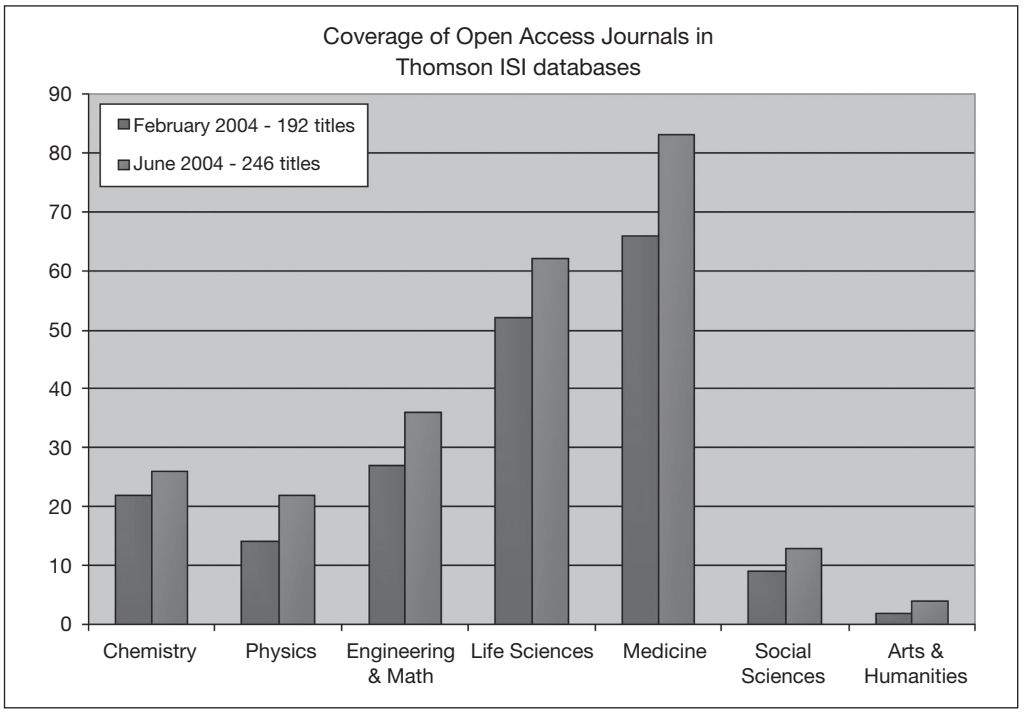

Figure 3. Coverage of Open Access Journals in Thomson ISI database

Surely this trend will continue and that OA will become more widespread (Figure 3).

All of the OA journals covered by Thomson Scientific have been evaluated and selected according to the same standards as journals with a more traditional business model. What matters to Thomson is the potential value of the contents of the journal to users of the WoS.

As a whole Thomson remains committed to bringing the users of its products the most important and influential scholarly information in a timely manner. It focuses on the world's most important journals through WoS, so that investigators will find the best, most relevant, and influential research quickly and efficiently.