

# Thomson Reuters Book Citation Index

Jonathan Adams<sup>1</sup> and James Testa<sup>2</sup>

<sup>1</sup>*jonathan.adams@thomsonreuters.com*

*Evidence* Thomson Reuters, 103 Clarendon Road, Leeds LS2 9DF, UK

<sup>2</sup>*james.testa@thomsonreuters.com*

Thomson Reuters, 1500 Spring Gardens, Philadelphia 19130, USA

## Abstract

Thomson Reuters has developed a novel Book Citation Index. This will become available from late 2011 and will contain a structured selection of scholarly books, categorised by discipline and book type. The Citation Index will be managed using similar inclusion criteria to those applied elsewhere in the Web of Science, so as to ensure that there is comparability across output types. The accurate and comprehensive citation links in the Index are primarily to enable cited items to be readily discovered. It is likely, however, that they will prove an informative source of analytical material for the scientometrics and research evaluation community.

## Introduction

While well-structured and searchable document databases for journal articles are widely available, similar facilities for other types of research output have taken longer to appear. Thomson Reuters has now created a Book Citation Index, available for research use from late 2011. This paper outlines the new Index, its rationale and content.

## Description

The Web of Science is known for its core coverage of top tier international and regional journal literature in the Science Citation Index Expanded, the Social Sciences Citation Index, and the Arts & Humanities Citation Index. In addition, Web of Science covers the Conference Proceedings Citation Index. These resources provide access to sciences and social sciences literature back to 1900 (conference proceedings to 1990) and arts and humanities journal literature to 1975. The indexes are characterized by quality and completeness. Their content has been monitored over decades to arrive at well-defined selection criteria applied in a systematic and consistent manner.

Most fields of research endeavour make extensive use of a diversity of output modes for publishing the results of research. This balance varies across fields. Data analysing the results of the UK's Research Assessment Exercise (RAE) over several cycles show that there has been a progressive concentration on journal articles as the output mode that most effectively represents a researcher's 'best 4' outputs (Adams, 2007). This has seen a shift from conference proceedings as a frequently submitted form of output in engineering and from scholarly monographs in the social sciences. Elsewhere, Lisée et al (2008) have suggested that conference proceedings may generally be diminishing in significance, but the evidence is mixed. Proceedings have grown from 8% of the references in engineering papers in the early 1980s to around 10%. UK RAE data and the Canadian analysis show that they continue to play a particularly important role in specific fields, and critically so in information and communication science and technology, where they account for more than half of the material submitted for UK assessment and, in Lisée et al's data, close to 20% of the references.

In the arts and humanities, however, books and book chapters have remained as a preferred form of representative research output (Weingart et al, 1991; Moed, Luwei & Nederhof, 2001, Hornbostel, 2008). This is very clearly seen in the data below analysed through the RAE cycles of 1996, 2001 and 2008.

**Table 1. Output types submitted for research assessment by UK academics in successive assessment cycles 1996, 2001 and 2008 demonstrating the diversity of output modes identified as significant for representing high quality research outcomes.**

<i>RAE1996</i>	<i>Science</i>		<i>Engineering</i>		<i>Social sciences</i>		<i>Humanities and arts</i>	
	<i>Outputs</i>	<i>%</i>	<i>Outputs</i>	<i>%</i>	<i>Outputs</i>	<i>%</i>	<i>Outputs</i>	<i>%</i>
Books and chapters	5,013	5.8	2,405	8.1	16,185	35.1	22,635	44.4
Conference proceedings	2,657	3.1	9,117	30.8	3,202	6.9	2,133	4.2
Journal articles	77,037	89.8	16,951	57.3	22,575	49.0	15,135	29.7
Other	1,104	1.3	1,122	3.8	4,154	9.0	11,128	21.8

  

<i>RAE2001</i>	<i>Science</i>		<i>Engineering</i>		<i>Social sciences</i>		<i>Humanities and arts</i>	
	<i>Outputs</i>	<i>%</i>	<i>Outputs</i>	<i>%</i>	<i>Outputs</i>	<i>%</i>	<i>Outputs</i>	<i>%</i>
Books and chapters	1,953	2.5	1,438	5.4	12,972	28.6	25,217	46.5
Conference proceedings	751	0.9	3,944	14.9	857	1.9	1,619	3.0
Journal articles	76,182	95.8	20,657	78.1	29,449	65.0	17,074	31.5
Other	618	0.8	408	1.5	2,008	4.4	10,345	19.1

  

<i>RAE2008</i>	<i>Science</i>		<i>Engineering</i>		<i>Social sciences</i>		<i>Humanities and arts</i>	
	<i>Outputs</i>	<i>%</i>	<i>Outputs</i>	<i>%</i>	<i>Outputs</i>	<i>%</i>	<i>Outputs</i>	<i>%</i>
Books and chapters	1,048	1.2	216	1.2	12,632	19.0	21,579	47.6
Conference proceedings	2,164	2.5	326	1.8	614	0.9	897	2.0
Journal articles	80,203	93.8	17,451	95.4	50,163	75.5	14,543	32.1
Other	2,125	2.5	301	1.6	3,018	4.5	8,287	18.3

It is evident from Table 1 that the core literature of research remains not just diverse but strongly disposed to books and to book chapters in the arts and humanities. Books also continue to represent a very substantial component of core literature in the social sciences despite suggestions that European social sciences are shifting towards an American research paradigm.

Butler and Visser (2006) noted that many types of publications are visible in Thomson Reuters Web of Science data including books, book chapters and conferences as well as

journals not indexed. More recently, Nederhof et al (2010) noted that for European social science publications that were themselves highly cited in Web of Science, books may represent between 60 and 80% of references to items not indexed by Thomson Reuters, whereas non-indexed journals make up fewer than 20%. Books (economics, political science) and manuals (psychology) account for the most frequently cited publications.

When applied to the assessment of university departments, this additional literature could have a significant effect on rankings, though it seems likely that much would prove correlative rather than contradictory and an extended analysis might not be equally beneficial in all fields. Nonetheless, many analysts would agree that the possibility of a more comprehensive view would increase both information content and – for evaluation – academic confidence.

The development in 2008 of the Thomson Reuters Conference Proceedings Citation Index (CPCI) established an important separate resource that allowed scholarly researchers to explore the proceedings literature in a searchable format. There had previously been some access through the Science & Technology Proceedings but the CPCI established a more authoritative single database with the same functionality but also a similar set of selection criteria to that understood in the Science Citation Index.

With the arrival of the Book Citation Index (BKCI) in Web of Science, Thomson Reuters will further expand its presentation of citation coverage of the core research literature. As with journals and proceedings, books will be selected according to well- defined criteria. The goal is to provide a comprehensive view of the scholarly literature across all major print and digital formats. By evaluation of the form and content of each publication, Thomson Reuters assures users not only that all Citation Indexes in Web of Science will contain the most relevant and timely research, but also that rigorous bibliographic control will ensure this content is readily discoverable.

Comprehensive does not mean all-inclusive. With books, as with journals and conference proceedings, there is a need to select those publications that will most likely contain significant scholarship. Thomson Reuters has demonstrated that, by applying rigorous selection criteria to the literature it covers, researchers are able quickly and efficiently to discover the scholarly content that best serves their needs.

### **Books Citation Index content**

The BKCI will include scholarly books in all subject areas. Initial content will include more than 1,000 book titles in each of economics, education and history while some eighty subjects identified by current Web of Science social science and humanities journal categories will have some immediate coverage. That said, the ideal subject categorization and granularity for books may differ significantly from that for journals (see below; and Adams, Gurney & Jackson, 2007) and this will require further evaluation with the research community.

The initial subject balance is projected to be: Social Science and Humanities – 56%; and Natural Sciences, Engineering – 44%.

**Table 2. The content of the Book Citation Index, by main subject area**

<i>Subject</i>	<i>Coverage</i>
Clinical Medicine	5%
Life Sciences	6%
Agriculture/Biology	4%

Physics/Chemistry	14%
Engineering/Computing/Tech.	15%
Social & Behavioral Sciences	38%
Arts & Humanities	18%

---

The content of the BKCI is made up of some two-thirds Books in Series with the balance in Monographs, but that may change as more Monographs are indexed. It is evident from prior work (e.g. Larivière et al, 2006) that chapters in book series tend to have relatively high citation counts that may be comparable to reviews in typical serials.

Content will include: original research or reviews; dissertations (as part of an established series); graduate level texts; and translations if scholarly commentary is present.

The basis of the BKCI is an analysis of 250,000 relevant publication records already indexed by Thomson Reuters from prior work. These are then used to identify items frequently cited by journal articles, reviews and conference proceedings and bibliographically describe them in a consistent and unified form. In 2010, some 22,000 additional items not already fully indexed by Thomson Reuters were acquired for evaluation for the BKCI. Of these, about 19,000 were selected for inclusion. In 2011, an additional 15,000 items have been acquired thus far and about 10,000 of these are likely to have been selected and indexed by the time that the BKCI becomes available. The growth profile is likely to be around 10,000 new items per year.

### **How the Book Citation Index will be used**

A Citation Index is primarily established to enable cited items to be discovered in a scholarly search. This will therefore be of significant value to library resources.

It will immediately become feasible for scholars using Web of Science to discover book literature alongside journal literature. This will be valuable in areas less well covered by current journal indexing as users will now receive more complete coverage of key outputs in the social sciences and arts and humanities literature, which is often in deficit in journal literature. More extensive literature reviews are likely to help novice researchers to understand the relevance and significance of the material in books to the results of other searches, rather than being swamped by the established ability to search journal material. They will, in particular, be able to understand the impact of research published in books, track prior research and monitor current developments, track citations to key monographs, and keep track of the influence of leading authorities on their field with cited reference searching. Not least, the BKCI is likely to prove of particular value – at a time of resource constraint – to librarians by helping with collection development and as an aid in acquisition decisions.

The BKCI is being developed in collaboration with major publishing houses and with major holders of book collections, a network that will undoubtedly grow. Publishers have been responding to the needs of researchers by developing digital publishing programs for books. Libraries have been responding by having their book collections digitized and by making sure their catalogues are universally accessible to registered users. Because of this, researchers expect added searchability and anytime/anywhere access to research literature, and this now includes books rather than only serials. Thomson Reuters will support this trend by providing discovery and analysis tools for book content. This will make it possible not only to analyze the citation network between books and wider scholarly research but also to connect to institutional resources through direct links to library catalogs and eBook collections. Bringing together book and journal content where researchers can find and use it supports the academic institutional mission to raise awareness of scholarly resources from the library.

Although ‘search and discovery’ is a prime purpose, it is obvious that a count of citation links will potentially enable additional statistics to be created. This will therefore be an exciting area for academic exploration and for authoritative investigation by the scientometric community. To this end the BKCI will include ‘times cited’ from journals, proceedings, and other books at the level of the Chapter and Book. In due course, citing items may also be collated at the level of Series.

### **The Book Citation Index and research evaluation**

Compared to the half-century of analysis of journal citation patterns, much less is known about citation patterns to books. Larivière et al (2006) have commented that “... books not only form a sizeable part of publications in some disciplines of the social sciences and humanities but are also often cited, and this impact cannot be extrapolated from that of journal articles. In other words, evaluations based only on measures obtained from journal databases are more likely to be less than adequate for disciplines in which less than 50% of references are made to journal articles than for those in which these references account for more than 50%”.

In professional areas, where academic activity is close to practice as it is in law and architecture (and indeed to a significant extent in clinical medicine and dentistry), access to a wider range of output Citation Indexes will be of particular interest. (See for example de Jong et al, 2011). Significant problems have been encountered in adequately describing the range and significance of non-journal material in such areas (Adams, 2008).

It is not immediately evident how a categorical structure for books might be established and whether this could follow the same pattern as that seen for journal articles. It is almost certainly true that the level of granularity at which the data must be analyzed will be critical in establishing sensible and relevant baselines (Adams, Gurney & Jackson, 2008) but whether this might be ‘humanities’, ‘history’ or ‘military history’ is a matter of speculation and waits on authoritative academic research with a good understanding of the structure of the relevant disciplines.

The dynamic connectivity between articles and books has yet to be explored to any degree. The growth rates for book citations are almost unknown though it might be that their citation half-life is significantly longer than that of journal articles. The BKCI will undoubtedly prove a fruitful ground both for those interested in more comprehensive research evaluation and for those wishing to develop a greater understanding of the underlying structure and significance of the total corpus of the research literature.

### **Acknowledgments**

The development of the Book Citation Index has been driven from Thomson Reuters Philadelphia office by the Editorial and Publisher Relations team and particularly by Susan Ciambrano (Product Development Manager).

### **References**

- Adams, J. (2007). *The use of bibliometrics to measure research quality in UK higher education institutions*, pp. 44. Universities UK, London. ISBN 978 1 84036 165 4
- Adams, J. (2008) *Strategic review of the Performance-Based Research Fund*. The assessment process, Wellington: Tertiary Education Commission.  
<http://www.tec.govt.nz/Documents/Reports%20and%20other%20documents/pbrf-strategic-review-of-assessment-process-2008-review.pdf>
- Adams, J., Gurney, K.A. & Jackson, L. (2008). Calibrating the zoom – a test of Zitt’s hypothesis. *Scientometrics*, 75 (1), 81-95.
- Butler, L. & Visser, M.S. (2006). Extending citation analysis to non-source items. *Scientometrics*, 66 (2), 327-342.

- de Jong, S.P.L., van Arensbergen, P. & Daemen, F. (2011). Evaluation of research in context: an approach and two cases. *Research Evaluation*, 20 (1) 61-72.
- Hornbostel, S. (2008). Gesucht: Aussagekräftige Indikatoren und belastbare Datenkollektionen. Desiderate geisteswissenschaftlicher Evaluierung in Deutschland, in: Lack, E & Marksches, C (Eds.): *What the hell is quality?* Frankfurt/New York: Campus Verlag, pages 55-73.
- Larivière, V., Archambault, E., 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.
- Lisée, C., Larivière, V. & Archambault, É. (2008). Conference proceedings as a source of scientific information: A bibliometric analysis. *Journal of the American Society for Information Science and Technology*, 59 (11), 1776-1784.
- Moed, H.F., Luwei, M. & Nederhof, A.J. (2002). Towards research performance in the humanities. *Library Trends*, 50 (3), 498-520.
- Nederhof, A.J., van Leeuwen, T.N. & van Raan, A.F.J. (2010). Highly cited non-journal publications in political science, economics and psychology: a first exploration. *Scientometrics*, 83 (2), 363-374.
- Weingart, P., Prinz, W., Kastner, M., Maasen, S. & Walter W. (1991). *Die Sog. Geisteswissenschaften: Aussenansichten: die Entwicklung der Geisteswissenschaften in der BRD, 1954-1987*. Suhrkamp Taschenbuch Wissenschaft; ISBN 978-3518285657