

Rafael Ball\*, Dirk Tunger\*\*  
Central Library, Research Centre Jülich  
Germany

## NATIONAL AND INTERNATIONAL PUBLICATIONS REFLECTED IN BIBLIOMETRIC DATABASES [KRAJOWE I ZGRANICZNE PUBLIKACJE W BIBLIOMETRYCZNYCH BAZACH DANYCH]

**Abstract:** Quantifying scientific output is a phenomenon without which science policy and science strategy can no longer be managed. Despite occasional criticism of bibliometric analyses, this tool – when correctly and transparently applied – provides a helpful parameter for science decisions, whereby the publication rate (quantity of publications) and the citation rate (perception of publications) in particular are explored and presented in different contexts.

An important problem associated with this process is the question of the language of publications, the origin of scientists and how they write their names and institutions.

Using best practise studies and examples from the Polish science community, this article will investigate the coverage of national languages and scientists in the citation databases WEB OF SCIENCE (Thomson Scientific) and SCOPUS (Elsevier).

BIBLIOMETRICS – COMMUNICATION SCIENTIFIC – EVALUATION OF RESEARCH – METHODOLOGY – NATIONAL COVERAGE SCIENTIFIC OUTPUT

**Abstrakt:** Ilościowe badania dorobku naukowego są niezbędne dla prowadzenia polityki naukowej i definiowania strategii rozwoju nauki. Mimo sporadycznie pojawiającej się krytyki analiz bibliometrycznych, narzędzie to – zastosowane we właściwy i przejrzysty sposób – pozwala określać parametry pomocne w procesie podejmowania decyzji dotyczących nauki i z tego powodu zarówno wskaźnik publikacji (odnoszący się do ich liczby), jak i wskaźnik cytowań (dotyczący percepcji tych publikacji), są szczególnie często wykorzystywane i przedstawiane w różnych kontekstach. Do istotnych problemów związanych ze stosowaniem tych wskaźników należy kwestia języka publikacji, pochodzenia naukowców oraz sposobu, w jaki zapisują swoje nazwiska i afiliacje. W oparciu o badania najlepszych praktyk i przykłady zaczerpnięte z polskiego środowiska naukowego przedstawiono obecność publikacji w językach narodowych oraz ich autorów w bazach cytowań WEB OF SCIENCE (Thomson Scientific) i SCOPUS (Elsevier).

BIBLIOMETRIA – EWALUACJA BADAŃ NAUKOWYCH – KOMUNIKOWANIE NAUKOWE – METODOLOGIA – ZAKRES JĘZYKOWY BAZ DANYCH

\*

\*       \*

---

\* Dr RAFAEL BALL, Head of the Central Library of Research Center Jülich, Germany (since 1998) and Teaching Professor at the Department of Information Science of the University of Applied Sciences in Chur, Switzerland (since 2002). Studies of biology, slavic languages and philosophy at the Universities of Mainz, Warsaw and Moscow. He took his doctorate in biology and made a additional 2-year education in Information Science in Frankfurt and Freiburg. He has worked in the library sector since 1994, first in the library of the University of Freiburg, Germany. The most important publications: (2006) *Bibliometrische Aspekte der Wissenschaftskommunikation* (co-operation); (2000) *The scientific information environment in the next millennium*. E-mail: r.ball@fz-juelich.de

\*\* DIRK TUNGER, PhD student in the Central Library of the Research Centre Juelich in Germany since 2004. He prepares his theses in bibliometrics. His main research topic is the connection of bibliometric analyses with trend studies.

## Introduction

An interest in bibliometric data and the emergence of analytical methods first began to any appreciable extent in the eighties of the 20<sup>th</sup> century. Initially, mathematicians, information scientists and sociologists concerned themselves with mathematical models in bibliometrics. After that, interest in bibliometrics faded somewhat until in the late nineties information and library scientists took up bibliometrics once more against the background of a new science scene. Large volumes of digital bibliometric data, now easily processible, as well as the necessity of providing reliable, quantifiable information on scientific output and the frequent introduction of performance-oriented allocation of funds in science and research made the question of the possible application of bibliometrics a hot topic again. "The German Scientific Council regards the performance-oriented allocation of funds as a central instrument for supporting research and teaching, which is to be applied both within the faculties and also on the level of the federal state [...]. The aim is to reward high achievers and to provide incentives for increasing performance in research and teaching" [Wissenschaftsrat 2004]. Bibliometrics is thus experiencing a revival, not primarily with respect to mathematical modelling and theoretical principles, but as an instrument of science management and international benchmarking.

Bibliometrics is gradually escaping from the "pigsty" of science evaluation and its reputation as a "conspiratorial element" and is beginning to establish itself as an accepted instrument in the orchestra of the overall evaluation of persons and institutions.

In the USA, for instance, bibliometric data have already been used since the mid-seventies as a basis for funding decisions, and the use of quantitative indicators is just as widespread in the Scandinavian countries as in Switzerland where "research maps" have been drawn for certain disciplines.

For international benchmarking it seems without doubt, that English language is the only accepted language for science measurement. Indeed one ignores a lot of national experiences when science benchmarking is focused only on international publications in English language.

## Relevant databases

For a long time, the Science Citation Index (SCI) [<http://portal.isiknowledge.com>] was the only multidisciplinary database that could be used to quantitatively determine the response to scientific publications. This method of measuring response (as the number of citations per paper) became increasingly important as a decisive factor in the evaluation of scientific output, which no longer consisted solely of unspecific, personal assessments by experts, but rather – immersed in the pool of performance indicators from the field of economics – included quantitative parameters in the evaluation. This development was recognised early on by the founder of the Science Citation Index at the Institute of Scientific Information, which led to the creation of a database that holds an unattested monopoly in the market decades later as the only benchmark for the quantitative evaluation of scientific output [Garfield 1984, pp. 525–535]. This era came to a close in 2005. The large scientific concern "Reed Elsevier" placed a second multidisciplinary database (SCOPUS) on the market in 2004 [[www.SCOPUS.com](http://www.SCOPUS.com)]. On the one hand, this product was to serve as a database of research literature for scientists, and therefore corresponds to a "normal" bibliographic database like those provided by libraries in every

possible discipline. On the other hand, the fact that it records the citations of scientific articles put it in direct competition with Thomson Scientific's existing Science Citation Index.

The introduction of a new product would not have warranted attention from an economic point of view had the second multidisciplinary research database of scientific reception not thrown the scientific community into a state of emergency when it came to the decision which database should now be used as the international benchmark for scientific rankings. If SCOPUS has not yet managed to establish itself among the scientific community, it is certainly not something that the general public is familiar with. Computer science and centres of excellence for bibliometrics, in particular, have now begun to analyse and compare both systems with regard to usability of results for quantitative scientific evaluations. A comprehensive comparison was recently published in the Austrian "Online-Mitteilungen" brought out by the University of Vienna [Gorraiz 2006, pp. 25–30; Wildner 2006, pp. 18–20], while other isolated papers also have dealt with this topic [Ball 2006a, pp. 177–178; Ball 2006b, pp. 293–301; Bar-Illan 2006; Tunger 2005, pp. 17–19]. We detected a research desideratum here – the question of an international benchmark for the evaluation of scientific output concerning the national languages in both databases.

SCOPUS claims that it is a database built on the thorough analysis of over 15,000 scientific journals. These journals come from the STM sector (science, technology, medicine) and the social sciences.

With over 15,000 journals that are regularly evaluated, the number of periodicals covered in SCOPUS is almost twice as high as that of SCI with around 7,500 journals covering the natural and social sciences [For more details, see Journal Citation Report (JCR) in the Web of Knowledge <http://portal.isiknowledge.com>].

The strategies employed by each database are very different. In SCI, only a selection of approx. 7,500 scientific journals are evaluated. The selection is over proportionately based on the average citation rate of articles in the journals. Limiting themselves to a set volume of journals is considered a quality criterion for SCI. The selection of titles is continuously checked and adjusted. If a journal falls below the assessment threshold, then it is dropped from the index, while journal titles that rise above the threshold are incorporated into the index.

SCOPUS on the other hand favours the "as much as you can" strategy. This means that a much greater number of journal titles are taken into account [<http://info.SCOPUS.com/detail/what/>]. This alone illustrates that the citation analyses in SCI and SCOPUS, each based on different data sets, produce very different results that are NOT comparable to each other.

Our Austrian colleagues have also examined both databases with a fine tooth comb. Juan Gorraiz describes the advantages and disadvantages of both databases from a bibliometric point of view. He comes to the conclusion that we cannot say with certainty which of the two databases is better since it also depends to a certain extent on the subject area in question.

## Key Issues

The low presence of non English articles in international scientific literature is a widely described phenomenon in the scientific communication. Even in the 70<sup>th</sup> of the 20. century the dominating languages in scientific literature were english, russian, german, french and japanese [Nicolas 1978].

The current paper is primarily concerned with if the higher number of journals (journal volume) is accompanied by a better coverage of national languages. We want to answer this question with the aid of selected

analyses for polish and german journals and scientific articles, represented in both databases. The investigations were carried out simultaneously in SCI and SCOPUS:

- Coverage of the various national journals in the SCI database
- Proportion of the various nations at the overall number of publications in both databases
- Proportion of polish language articles of polish papers in both databases
- Proportion of english language articles of polish papers in both databases

## Results

### ***Coverage of the various national journals in the SCI database***

The coverage of scientific journals of smaller nations in the database Journal Citation Report was investigated by Bordons et al [Bordons 2002, pp. 195–206; Presmanes 2003, pp. 547–558]. They showed an only low reflection of journal representation of the scientific output. Our findings are reflected in Figure 1; it shows the dissemination of the various national journals reflected in the SCI. There is no wonder, that the USA are number one in the database with more than 38% of all scientific journals. The second place is covered by Great Britain with more than 20%. That means, more than half of the SCI is dominated by journals from an english spoken country. The rest is scattered over more than 100 nations. Poland is at the relatively high position number 13 with 0.9% of all the journals covered by the SCI. For SCOPUS similar data are not available now, because the search strategies are more limited than these of the SCI.

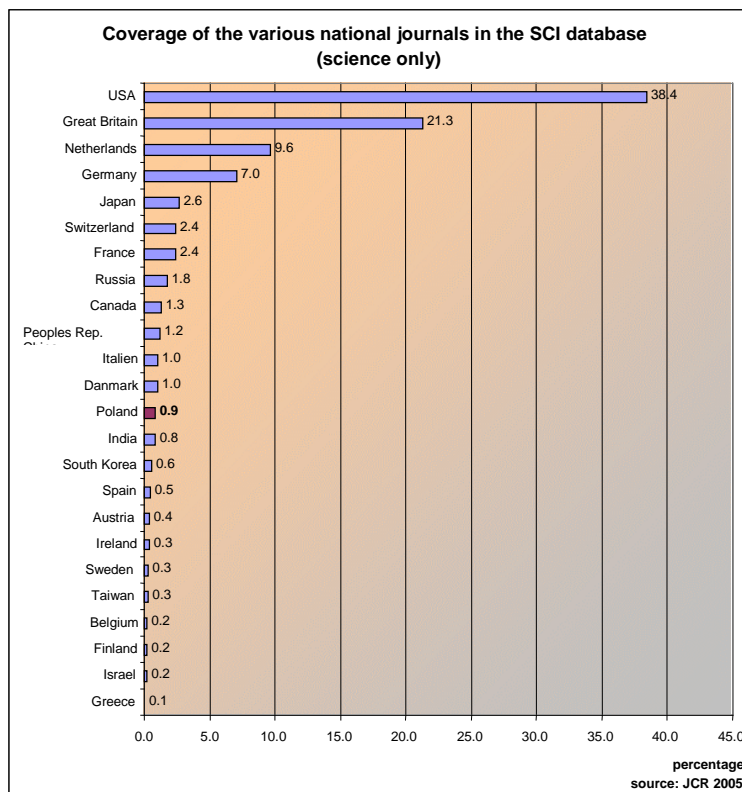


Figure 1: Coverage of the various national journals in the SCI database

### Proportion of the various nations at the overall number of publications in both databases

This investigation reflects to the proportion of the overall publications from various nations of the articles covered by SCI and SCOPUS (Figure 2 & Table 1).

In SCI more than 25% of all articles come from the USA, on position 2 we find very close together GB and Japan with more than 8%. Position 4 and 5 is occupied by Germany (7,7%) and France (5,4%). If we summarize the states of central Europe we come to an proportion of nearly 30% of all articles in SCI. This is the classical triade of leading science regions worldwide (US, Europe, Japan). But in future we will recongnize a new up-coming nation, which publishes more and more: China. China will be soon at position 4 and than we no longer talk about a science triad, but a science tetrad [Glänzel 2006, pp. 59–61]. Articles from Poland we find at the position number 18.

The data of SCOPUS are showing a very similar distribution of the various contries. Only China is more reflected in SCOPUS (5.0% of all articles) than in SCI (4.2% of all articles)

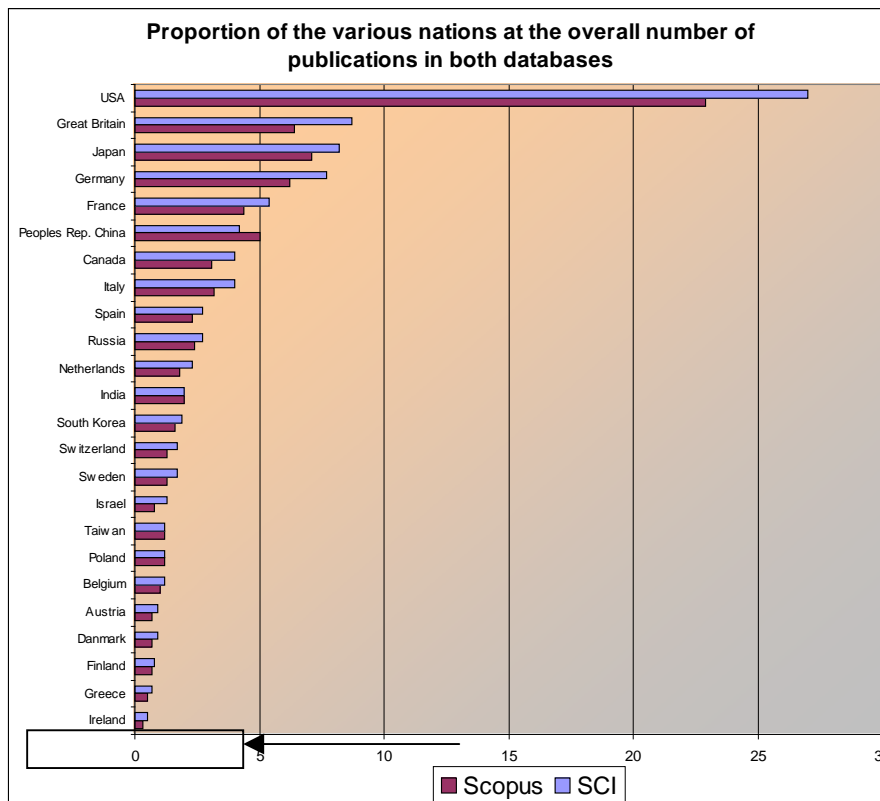


Figure 2: Proportion of the various nations at the overall number of publications in both databases

Table 1: Proportion of the various nations at the overall number of publications in both databases (percentage of the publications from a specific country in each database)

Country	Proportion at the overall number of publications	
	Scopus	SCI
USA	22.9	27.0
Great Britain	6.4	8.7
Japan	7.1	8.2
Germany	6.2	7.7
France	4.4	5.4
Peoples Rep. China	5.0	4.2
Italy	3.2	4.0
Canada	3.1	4.0
Russia	2.4	2.7
Spain	2.3	2.7
Netherlands	1.8	2.3
India	2.0	2.0
South Korea	1.6	1.9
Sweden	1.3	1.7
Switzerland	1.3	1.7
Israel	0.8	1.3
Belgium	1.0	1.2
<b>Poland</b>	<b>1.2</b>	<b>1.2</b>
Taiwan	1.2	1.2
Danmark	0.7	0.9
Austria	0.7	0.9
Finland	0.7	0.8
Greece	0.5	0.7
Ireland	0.3	0.5

### Usage of Polish and English language in Polish papers in both databases

This investigation shows the proportion of Polish language articles in relation to all Polish articles (affiliation) in both of the SCI and SCOPUS during a 30 year period (Figure 3 & Table 2).

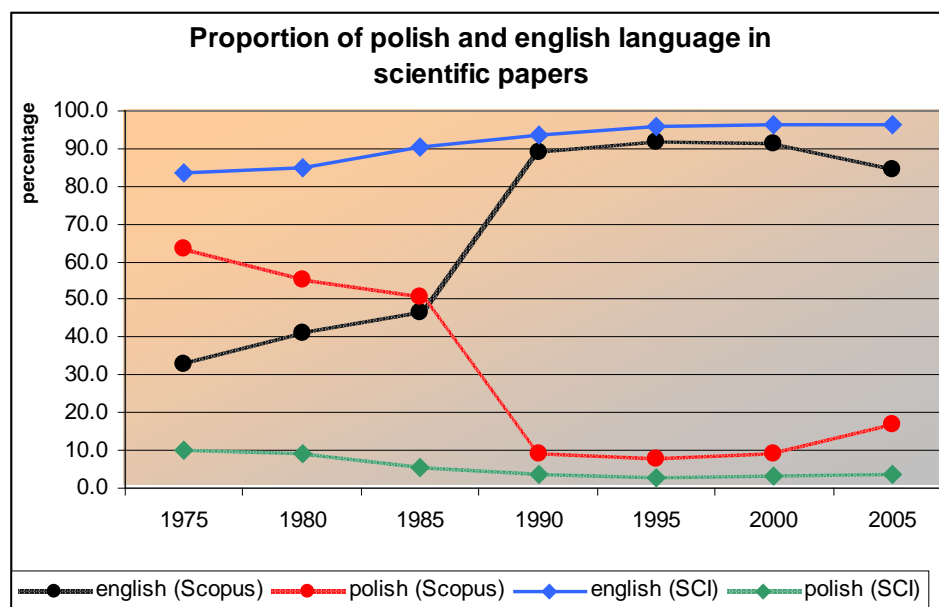


Figure 3: Proportion of Polish and English language in scientific papers from Poland

Table 2: Proportion of Polish and English language in scientific papers and total number of articles in each database

	SCI			SCOPUS		
	percentage english	percentage polish	# articles in SCI	percentage english	percentage polish	# articles in SCOPUS
1975	83.5	10.2	3894	32.9	63.5	4401
1980	84.8	9.2	5485	41.1	55.3	3228
1985	90.2	5.5	5659	46.5	50.7	2448
1990	93.4	3.7	6106	88.9	9.1	3013
1995	95.7	2.5	8231	92.0	7.5	3869
2000	96.1	3.2	10926	91.1	9.1	12666
2005	96.2	3.5	15907	84.4	16.8	20031

SCI: The part of English language articles from 1975 to 2005 is constantly rising from about 80% in 1975 to more than 95% in 2005. In accordance the proportion of Polish language articles decreases from 10% to 3.5% till 2005. These findings indicate, that nearly all articles from Polish scientists covered in SCI are written in English.

SCOPUS: A completely other situation we find in SCPOUS: In 1975 more than 60% of the Polish articles are written in Polish, only 30% in English. In the year 1985 there is a “breakeven point” at which the number of English language publications topped the Polish spoken articles. In 2005 we have a situation very similar to the distribution in SCI: 84% of all polish articles covered by SCPOUS are English language articles (SCI: 96,2%) and only 16,8% are Polish language papers (SCI: 3,5%).

If we look at the total numbers of covered Polish articles in 2005 in SCOPUS (English and Polish: 20031) the distribution reflects the strategy: Elsevier covers as much as possible and so the number of total publications also from smaller nations is actually higher than the appropriate number in SCI (15907), Figure 4.

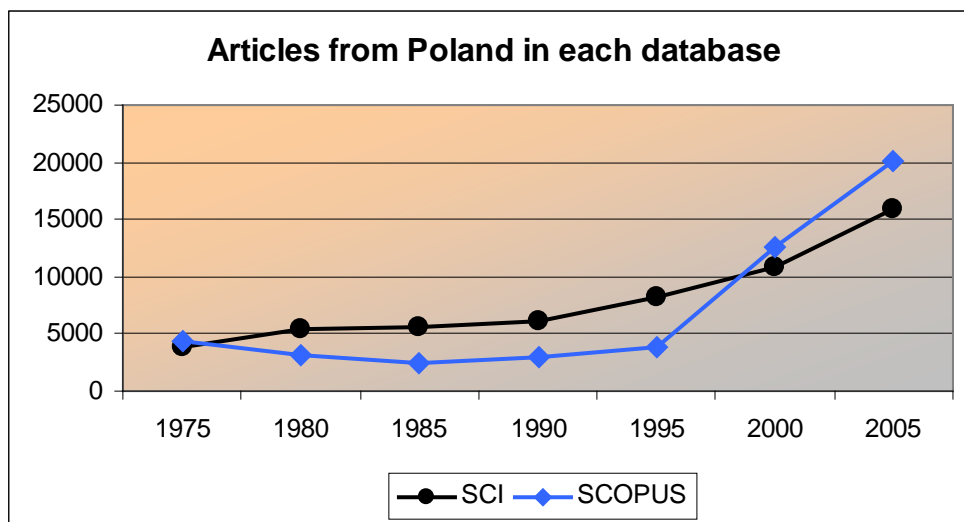


Figure 4: Articles from Poland in each database

## Discussion and Conclusions

We were able to show that depending on the data base chosen, bibliometric analyses of national languages provide very different information. The database should therefore be chosen with great care and on the basis of content. Furthermore, it still remains unclear what effects two relevant citation databases with similar contents will have on the process of scholarly communication. The current strategies employed by the two database pro-

viders are conflicting: the number of journals covered are expanded in terms of breadth in one (SCOPUS) and in terms of depth in the other (SCI).

But obviously the SCOPUS database contains more articles of smaller European nations and more publications in the national languages. Therefore it seems to be better for regional coverage of scientific literature. It can break the English language dominated databases of SCI and could be more of applicable for science evaluation of non englishspoken nations. Another question is the quality standard of the SCI. If one enlarges the basis for science evaluation too much, it could be difficult to compare the results worldwide.

We will do further investigations concerning language and smaller nations coverage of both databases

Despite everything, bibliometrics do reveal the nuts and bolts of the two databases and will conduct other comparative analyses with these data archives. The SCI benchmark is still number one worldwide. It remains to be seen whether SCOPUS will ever be in a position to take over.

## References

- Ball, R.; D. Tunger (2006a). Science Indicators Revisited. in: Book of Abstracts – 9th International Conference on Science & Technology Indicators. 07–09 September 2006, Leuven, Belgium, pp. 177–178; this was also presented as a poster on the conference.
- Ball, R.; D. Tunger (2006b). Science indicators revisited – Science Citation Index versus SCOPUS: A bibliometric comparison of both citation databases in: *Information Services and Use*, 26, 2006, pp. 293–301.
- Ball, R.; M. Stein-Arsic, B. Buscher, E. Bickar (2003). Bibliometrische Analysen als Instrument des Bestandsmanagements in Bibliotheken. *B.I.T.-Online*, 6(4), 2003, pp. 347–351.
- Bar-Illan, J. (2006). Which h-Index? Comparison of WoS, SCOPUS & Google Scholar. in: Book of Abstracts – 9th International Conference on Science & Technology Indicators. 07–09 September 2006, Leuven, Belgium.
- Bordons, M.; M. T. Fernandez, I. Gomez (2002). Advantages and limitations in the use of impact factor measures for the assessment of research performance. *Scientometrics*, 53(2), 2002, pp. 195–206.
- Garfield, E. (1984). Science Citation Index – A New Dimension in Indexing in: *Essays of an Information Scientist*, Vol.7, pp. 525–535, 1984; Reprinted from: *Science*, 08.05.1964, Vol.144, No. 3619, pp. 649–654.
- Glänzel, W.; K. Debackere; M. Meyer (2006). Triad or Tetrad? On global changes in a dynamic world. In: 9th International Conference on Science and Technology Indicators, 7.–9. September 2006, Leuven, Belgium, Book of abstracts, pp. 59–61.
- Gorraiz, J. (2006). Web of Science versus SCOPUS oder das aktuelle Dilemma der Bibliotheken in: *Online-Mitteilungen* No. 85, March 2006, pp. 25–30; electronic version available at: <http://www.univie.ac.at/voeb/php/publikationen/om/> [visited 14.03.2007].
- Nicholas, D.; M. Ritchie (1978). *Literature and Bibliometrics*. London: Clive Bingley.
- Presmanes, B.; E. Zumelzu (2003). Scientific cooperation between Chile and Spain: Joint mainstream publications (1991–2000). *Scientometrics*, 58(3); 2003, pp. 547–558.
- Research Centre Jülich – Central Library (2003). *Bibliometric Analysis in Science and Research – Applications, Benefits and Limitations*. Juelich. <http://www.fz-juelich.de/zb/Bibliometrics> [visited 14.03.2007].
- Trkulja, V. (2005). Elseviers SCOPUS – Weltgrößte Abstracts- und Zitationsdatenbank aus dem wissenschaftlichen Web in: *password* 5/2005, pp. 37–40.
- Tunger, D. (2005). Ist mehr oder weniger Inhalt besser? Vergleich der Inhalte SCOPUS und Science Citation Index in: *Password : Nachrichten & Wissen für Informationsprofis im deutschsprachigen Raum*, Volume 10, pp. 17–19.



- Wagner-Döbler, R. (2003). The system of research and development indicators: entry-points for information agents. Paper presented at the conference: Bibliometric Analysis in Science and Research – Applications, Benefits and Limitations, Research Centre Juelich – Central Library.
- Wildner, B. (2006). Web of Science – SCOPUS: Auf der Suche nach Zitierungen in: Online-Mitteilungen No. 85, March 2006, pp. 18–20; electronic version available at: <http://www.univie.ac.at/voeb/php/publikationen/om/> [visited 14.03.2007].
- German research Council (Wissenschaftsrat) (2004). Leistungsorientierte Mittelvergabe der Hochschulmedizin, online accessible under: [http://www.wissenschaftsrat.de/wr\\_fokus/fokus-mittel.html](http://www.wissenschaftsrat.de/wr_fokus/fokus-mittel.html) [visited 14.03.2007].