Welcome to the linguistic warp zone: Benchmarking scientific output in the social sciences and humanities[1]

Éric Archambault[*], Étienne Vignola-Gagné[**], Grégoire Côté**,
Vincent Larivière*** and Yves Gingras[***]

[*]*eric.archambault@science-metrix.com*
Science-Metrix, 4572 avenue de Lorimier, Montréal, Québec H2H 2B5 (Canada) and
Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la
science et la technologie (CIRST), Université du Québec à Montréal, Montréal (Québec), Canada

[**]*etienne.vignola-gagne@science-metrix.co, gregoire.cote@science-metrix.com*
Science-Metrix, 4572 avenue de Lorimier, Montréal, Québec H2H 2B5 (Canada)

[***]*lariviere.vincent@uqam.ca, gingras.yves@uqam.ca*
Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la
science et la technologie (CIRST), Université du Québec à Montréal, Montréal (Québec), Canada

Abstract
The goal of this paper is to examine the impact of linguistic coverage of databases used by
bibliometricians on the capacity to effectively benchmark the work of social scientists and
humanities researchers in various countries. The paper firstly examines the strong link
between bibliometrics and the ISI's database and, subsequently, reviews the differences in the
production and diffusion of knowledge in the SSH and NSE. This leads to an examination of
the old debate on the coverage of ISI's databases, more specifically in the case of SSH. The
methods section explain how we have compared the coverage of ISI's databases in the NSE
and SSH to the Ulrich extensive database of journals. Our results show that there is a 20 to
25% overrepresentation of English-language journals in ISI's databases compared to the list of
journals presented in Ulrich. This paper concludes that because of this bias, ISI's databases
cannot be used to benchmark the output of countries in the SSH.

Introduction
Scientometric methods are increasingly used for science policy purposes, more particularly in
Europe. The use of scientometrics for policy purposes has so far been mostly limited to the
natural sciences and engineering (NSE), but this is changing and the extension of this
evaluation process to the social sciences and humanities (SSH) may be a cause for concern. A
number of scholars have highlighted fundamental differences between the scientific
communication practices of scholars in the NSE and those in the SSH (Glänzel and
Schoepflin, 1999; Hicks, 1999 and 2004; van Raan, 2003). For instance, SSH comprise a
greater proportion of scientific publications which are not journal articles, and especially the
proportion of books; referencing practices that translate in fewer citations towards relatively
older literature; and the relatively more local orientation of research in the SSH.

Indeed, while policy makers can count on scientometrics to produce relatively reliable and
valuable results on the research performed in the NSE, their application to the SSH has been
shown to be problematic for many reasons. A number of bibliographic characteristics proper
to the SSH hampers scientometric analyses: research results are disseminated through a much
broader range of media than in the NSE; referencing practices that translate in fewer citations
towards relatively older literature; and the relatively more local orientation of research in the

---

SSH. The problem posed by this last specificity is particularly important since efforts to solve it involve the changing of one of the central tools of scientometrics: the Thomson ISI Citation Indexes. Most bibliometric studies are based on one or more of the Thomson ISI databases: the Science Citation Index (SCI), the Social Science Citation Index (SSCI), and the Arts and Humanities Citation Index (AHCI). The literature has argued the SSCI and the AHCI do not cover adequately journals and articles in languages other than English. In fact, there has been a healthy debate going on for years on this issue. Considering the potential expansion of scientometric evaluation for policy purposes to the SSH, this debate is highly relevant.

The goal of this paper is to shed light on this debate and to examine the impact of linguistic coverage of the database on the capacity to effectively benchmark the work of social scientists and humanities researchers in various countries. The paper firstly examines the strong link between bibliometrics and the ISI's database and subsequently the differences between SSH and NSE. This reviews leads to examine the old debate on the coverage of ISI's databases, more specifically in the case of SSH. The methods section explain how we have compared the coverage of ISI's databases in the NSE and SSH to the Ulrich extensive database of journals. Our results show that there is a 20 to 25% overrepresentation of English-language journals in ISI's databases compared to the list of journals presented in Ulrich. This paper concludes that because of this bias, ISI's databases cannot be used to benchmark the output of countries in the SSH.

The marriage of bibliometrics and ISI's databases
Bibliometrics and scientometrics are a set of methods for measuring the production and dissemination of scientific knowledge. Derek de Solla Price and Vasilij Vasilevich Nalimov were the originators of the discipline, which they developed for the purpose of providing research tools to historians and sociologists of science. However, it was only with the advent of the tools developed by the Institute for Scientific Information (now Thomson ISI) and the research conducted by its founder, Eugene Garfield, that the use of bibliometrics became widespread. With their systematic archiving of articles from a selection of some of the most prestigious and most often cited scholarly journals, the Thomson ISI databases considerably reduce the effort required to carry out bibliometric analysis. This partly explains why there has been a marriage of interest between ISI and the bibliometric community and why the most commonly used databases in bibliometric analyses are by far the ones produced by Thomson ISI.

It must also be stated that ISI production process, although not always as transparent as the bibliometrician would like it to be, is fairly well documented. Established scholars select the journals to be covered in cooperation with users, publishers and members of editorial boards. Selection criteria include publication frequency, compliance with international presentation conventions, and a peer review committee. ISI's databases are particularly useful for a number of reasons (see Katz and Hicks, 1998, among others):

- Their primary advantage is coverage. They cover all research fields, giving ready access to aggregated data. Unlike a number of other databases, which may cover some journals only in part (depending on the relevance of an article to single-discipline databases, for example), these databases systematically index all articles and other items in the selected journals. Nearly 10,000 existing learned journals are covered (Katz and Hicks, 1998).

- The criterion for including a journal in SCI and SSCI is the number of citations it receives. Because citations are perceived to be an indicator of interest and recognition on the part of scholars, the citation count is considered as evidence of the usefulness, quality and/or impact of a journal. According to Eugene Garfield, 90–95% of the most frequently cited articles in the natural sciences are published in a core group of about 2000 journals (CNER, 2002, citing Garfield, 1996). However, it is noteworthy that AHCI inclusion criteria are more subjective. According to Thomson ISI:

  > Publishing standards, including timeliness, are also important in the evaluation of Arts and Humanities journals. Citations in the Arts and Humanities, however, do not necessarily follow this same predictable pattern as citations to Social Sciences and Natural Sciences articles. Citations to an article on the 19th Century Romantic novel, for example, may accrue slowly at first, and then slacken, fluctuating over time in cycles consistent with scholars' varying interest in the topic. In addition, Arts and Humanities journal articles reference non-journal sources heavily (e.g., books, musical compositions, works of art and literature). Consequently, citation data, while sometimes useful, are frequently much less so in journal evaluations in the Arts and Humanities. Arts & Humanities journals are selected by the primary editor with the support of the subject experts from the ISI Arts & Humanities indexing staff. The goal is the collection of superb Arts and Humanities content that reflects the complex nature of cultural phenomena across a broad range of fields. http://www.isinet.com/essays/selectionofmaterialforcoverage/199701.html/

- The Thomson ISI databases contain the institutional addresses for all authors of a given article. Other databases usually contain only the address of the first-named author. Without a complete address list, the analysis of collaboration research analysis is impossible.

- Only the Thomson ISI databases contain citation information, which makes it possible to measure research impact. In the opinion of Katz and Hicks, this attribute alone justifies using these databases as science policy and research management tools.

The Thomson ISI databases, like all databases used for scientometric evaluation, also have limits. These will affect the validity of bibliometric evaluation in both the NSE and the SSH. The main problems with research publication databases comprise (CNER, 2002):

- limited coverage;
- exclusion of certain types of documents;
- classification of journals by discipline;
- changes in journal titles;
- different individual authors with identical names (homographs);
- unequal distribution of work among authors which might not be reflected;
- negative and erroneous citations, self-citation and personal strategies.

Another disadvantage is that they are relatively costly to use (Katz and Hicks, 1998). Since bibliometrics is generally based on the use of bibliographic databases, all the above factors reduce the reliability of bibliometric research evaluation. These problems are common to most the databases and reflect the limitations of compiling and indexing academic journals and articles. But scientometric analyses may also be affected negatively by some practices and situations specific to the field concerned. This is exactly the case with the SSH, as can be seen in the following section.

SSH are not NSE

According to Line (1999), most social sciences can be described as "relatively young, and scarcely organized as coherent disciplines." He suggests that the SSH are fragmented because they do not have international standards for rigorously defining concepts. Terms used vary markedly between regions and over time. Another factor leading to fragmentation is a greater reliance on various means of delivering research results. Hicks (1999) expands on this view by saying that a number of SSH disciplines have more paradigms competing with one another than do those in the NSE, and as a result SSH literature is more fragmented — a situation that hinders the formation of a solid "core" of scientific journals (Hicks 1999), thereby making article-based bibliometric analysis more difficult to conduct successfully.

The SSH differ from the NSE because of fundamental differences in scientific communication practices and their methods (Glänzel and Schoepflin, 1999; Hicks, 1999 and 2004; Moed, Nederhof and Luwel, 2002; van Raan, 2003). Knowledge dissemination media and, by extension, communication media in general are more varied in the SSH than in the NSE. In contrast to the NSE where journals are the dominant form of knowledge dissemination, Hicks (1999) estimates that books make up 40–60% of the literature in the social sciences.

Another problem facing the scientometric analysis of the SSH is what authors have called their local orientation. Whereas the problems identified in the NSE tend to be universal in nature, SSH research subjects are sometimes more local in orientation and, as a result, the target readership is more often limited to a country or region (Glänzel, 1996; Hicks, 1999; Hicks, 2004; Ingwersen, 1997; Nederhof et al., 1989; Nederhof and Zwaan, 1991; Webster, 1998; Winclawska, 1996). Indeed, the literature suggests that the readership in the natural sciences is the international scientific community. Research in physics or biology can be carried out and reused by interested experts anywhere in the world, with the same results; for instance, the properties of an electron are the same everywhere on Earth. In the natural sciences, the outcome of this universality is that dissemination is international. The preferred medium is the scientific article, very often written in English and published in a journal with international distribution.

By contrast, SSH research deals with local and regional problems more often than is the case in the NSE. It follows that SSH research subjects would be specific to a particular culture more often than would the problems tackled in the NSE. Hicks (1999 and 2004) suggests that theoretical concepts in the social sciences are more subtle and cannot be expressed in the universal language of mathematics as much as theoretical concepts can be in the natural sciences. In many cases, the concepts and subjects covered in the SSH can be expressed and understood only in the language of the culture that is shaping them and, consequently, SSH scholars reportedly publish more often in their mother tongue and in journals with a more limited distribution (Gingras, 1984; Line, 1999). Accordingly, SSH scholars publish somewhat more in their own language and in journals with national distribution.

Table 1 presents data on Finnish research output. It includes data on the number of publications written in Finnish, irrespective of type of publication (article, book, conference proceedings), as well as data on Finnish articles in peer-reviewed journals with an international distribution (this data necessarily covers only research articles). The data supports the thesis that the SSH are more locally oriented, since the number of publications in the first category is much lower than in the second. It also supports the thesis that the SSH is becoming increasingly internationalized, because the number of articles published in internationally distributed journals has grown significantly since 1994.

Table 1. Annual output of Finnish scholars

Field	Type of publication	1994	1998	2002
Natural Sciences and Engineering	Local orientation, written in Finnish	3,787	3,032	2,828
	International journal with peer review	6,419	6,702	7,857
	Ratio	0.6	0.5	0.4
Social Sciences and Humanities	Local orientation, written in Finnish	2,871	4,001	3,570
	International journal with peer review	685	984	1,265
	Ratio	4.2	4.1	2.8

Source: Academy of Finland 2003

Enter the warp zone: the good old debate on ISI coverage
If the assumptions mentioned above on the local orientation of SSH research are correct, it is important to determine whether the tools used by bibliometricians to perform international benchmarks reflect this state-of-affairs adequately. In particular, since they are the toll used de facto by the bibliometric community, it is important to determine if Thomson ISI databases provide an un-biased coverage of literature that is more national in scope and with a more local distribution. Importantly, since English is only used by a small proportion of the locals who live in different countries and that SSH publications are seemingly often written by locals for locals, it is important to determine if these databases adequately reflect production in languages other than English. A number of studies provide some empirical data on and interesting insight into a potential distortion in the representation of the SSH offered by the SSCI and AHCI.

A comparison of the UNESCO list of social sciences periodicals with that of ISI reveals some quite significant differences (Schoepflin, 1992). The UNESCO list contains about 2.5 times more academic journals than the SSCI list. US journals account for about 60% of SSCI coverage, yet they represent only 17% of all journals according to the UNESCO list. The comparison shows that SSCI includes more US journals than the number included in the UNESCO list.

Nederhof and Zwaan (1991) noted that the coverage provided by the two databases varied considerably by field, journal importance and language. Only 3% of Dutch articles in public administration were covered by the SSCI, compared with 58% of articles in experimental psychology. With respect to the humanities, coverage varied from 10% of articles in Dutch language studies to 39% of articles in general literature. According to Kyvik (1988, cited by Nederhof and Zwaan, 1991), only one third of Norwegian publications in the SSH during the period 1979–81 were written in a language other than Norwegian, compared with 74% of the publications in the natural sciences.

Authors have made a strong case to the effect that France, Spain and Germany are under-represented in SSCI (Ingwersen, 2000). Even though all three countries enjoy a strong SSH tradition, this fact is not reflected in SSCI. For example, a study covering the periods 1989-93 and 1994-98 reports that Germany is last out of 17 countries ranked according research impact in the social sciences (Ingwersen, 2000). More specifically, Schoepflin (1992) gives the results of surveys of German scholars, who were asked to identify the academic journals with the highest profile and the greatest value for their respective disciplines. Based on their

responses, SSCI covers 94% of German journals in developmental psychology but only 26% in sociology and 8% in education. Mela, Cimmino and Ugolini (1999) argue that:

> bibliometric analyses are biased towards English language journals, and authors of some nations (e.g. France and Germany) with a strong tradition of publishing in their native languages and less prone to submitting papers to internationally peer-reviewed English language journals may be penalized in comparative studies drawing on databases that include few non-English-language publications.

Andersen 2000 (citing Andersen, 1996) suggests that the authors of 60% of the articles indexed in SSCI have U.S. addresses and the authors of 20% of them have U.K. addresses. Since Thomson ISI selects journals according to the number of citations they receive, the citation habits of the various language communities play an important role in the actual compilation of the Thomson ISI databases. For example, U.S. and U.K. sociologists cite articles written in English in 99% of cases. At the same time, those articles account for approximately 70% of international literature in sociology (Yitzhaki, 1998). This factor in itself might explain in part the suspected Anglo-Saxon over-representation in the Thomson ISI databases.

In light of statistics like these, a number of bibliometricians claim that the SSCI and AHCI databases have a bias in favour of English-language journals from Anglo-Saxon countries — specifically, the U.S., the U.K. and, to a lesser extent, Canada and Australia (Andersen, 2000; Glänzel, 1996; Nederhof and Noyons, 1992; Schoepflin, 1992; Webster, 1998). This seems to be confirmed by Godin (2002) who, using the SSCI, placed the United States and the United Kingdom as the most productive countries in the social sciences for the year 2000, followed by Canada and Australia, and then by Germany, the Netherlands and France. Intuitively, it seems very improbable that Canada and Australia would produce more papers in the SSH than a country like Germany with its much larger population and its long traditions of prolific authors such as Kant, Weber, Habermas and so forth. Because of a lack of robust evidence to back up this claim, we decided to verify whether Thomson ISI databases were sufficiently representative of various countries and languages in which knowledge is produced.

Methods

To assess the coverage of national literatures by Thomson ISI, we compared the list of journals covered in ISI's citation indexes with an exhaustive and comprehensive source of scientific journals from all over the world - the Ulrich directory. Analyses of the coverage provided by the Thomson ISI databases are by-and-large based on a comparison of the journals included in them with those in the Ulrich directory. While this method has been used by Braun, Glänzel and Schubert (2000) to assess the field by field coverage of the SCI, none has compared the SCI to the SSI and the AHCI in that respect.

The lists of journals in the Thomson ISI databases — the Science Citation Index (SCI), the Science Citation Index Expanded (SCI_E), the Social Sciences Citation Index (SSCI) and the Arts and Humanities Citation Index (AHCI) — were taken from the Thomson ISI website (http://www.isinet.com/journals/) as of June 3, 2004. Data on the journals in the Ulrich directory was taken from the 2004 second edition of the CD-ROM of Ulrich's International Periodicals Directory (ISSN 0000-0507). Journals in the two sources were matched by means of their ISSN number. Over 95% of the Thomson ISI journals were matched with the Ulrich database. Since 93.5% of the matched ISI journals are peer reviewed (SCI = 98.6%; SSCI = 97.4%; AHCI = 73.4%), the analysis in this report is based only on peer reviewed journals from both Thomson ISI's databases and Ulrich directory.

The language of journal editors and the language of journals are directly coded in Ulrich. More than one language is provided for some journals. In such cases, equal fractions have been assigned to all the main languages of the journal concerned. Thus for a journal that mainly contains texts in English and French and only occasionally includes Spanish texts, 0.5 goes into the "English" column and 0.5 into the "French" column. Although this technique is not perfect, the effect of this distribution is negligible due to the large numbers involved in this analysis. For the analysis of the languages of editors' countries, the first language attributed to countries in the World Factbook[2] was used.

Findings

Table 2 presents statistics on the proportion of journals by country of editor (not publishing company). It gives, in sequence, data on NSE and SSH journals covered in Thomson ISI's databases and in Ulrich's journal database. The Table shows that journals with U.K. editors are very heavily over-represented in the Thomson ISI databases, especially in the SSH. According to Ulrich, 18% of journals have a U.K. editor, whereas Thomson ISI indexes 27% of journals with an editor in that country — an over-representation factor of 55%. SSH journals with editors located in the Russian Federation, the US, Switzerland and the Netherlands are over-represented too, while virtually all other countries are under-represented.

Table 2. Journal coverage rates by country of editor

Country	NSE			SSH		
	Thomson ISI	Ulrich	Difference	Thomson ISI	Ulrich	Difference
United Kingdom	23%	17%	36%	27%	18%	55%
Russian Federation	1.6%	1.4%	12%	0.3%	0.3%	36%
United States	36%	31%	19%	50%	37%	35%
Switzerland	2.7%	2.1%	26%	0.6%	0.5%	8%
Netherlands	9.4%	8.3%	14%	7.7%	7.4%	5%
Canada	1.3%	1.3%	1%	2.5%	3.2%	-21%
France	2.4%	2.6%	-6%	1.0%	1.4%	-24%
Germany	7.7%	6.2%	25%	3.9%	5.9%	-34%
Japan	2.3%	3.7%	-39%	0.5%	1.0%	-55%
Australia	1.2%	2.1%	-42%	1.1%	3.6%	-71%
Spain	0.4%	1.3%	-72%	0.3%	1.0%	-75%
Belgium	0.2%	0.4%	-52%	0.5%	2.1%	-75%
India	0.9%	2.2%	-61%	0.2%	1.6%	-86%
Poland	0.7%	1.6%	-58%	0.2%	1.3%	-87%
Italy	1.1%	1.7%	-38%	0.1%	1.2%	-89%
China	0.9%	2.9%	-69%	0.1%	0.9%	-91%
Brazil	0.3%	1.1%	-72%	0.04%	1.0%	-96%
Other	7.5%	14%	-45%	3.5%	13%	-73%

Source: Compiled from Thomson ISI's and Ulrich's data.

Table 3 gives the distribution of Thomson ISI coverage and Ulrich's listing by principal language of country of journal editor. It shows that only journals with editors in countries

---

[2] http://www.cia.gov/cia/publications/factbook/index.html

where the language spoken is English or Russian are over-represented. For example, SSH journals with editors in French-speaking countries are under-represented by 27% in the Thomson ISI databases.

Table 3. Journal coverage rates by language spoken in editor's country

Language	NSE			SSH		
	Thomson ISI	Ulrich	Difference	Thomson ISI	Ulrich	Difference
Russian	1.6%	1.4%	12%	0.3%	0.3%	36%
English	64%	55%	17%	82%	65%	26%
Dutch	10%	8.6%	11%	8.2%	9.4%	-13%
Czech	0.3%	0.7%	-55%	0.3%	0.4%	-15%
French	2.4%	2.6%	-6%	1.0%	1.4%	-27%
German	11%	9.0%	19%	4.6%	7.0%	-34%
Japanese	2.3%	3.7%	-39%	0.5%	1.0%	-55%
Danish	1.2%	0.9%	40%	0.3%	0.7%	-61%
Afrikaans	0.3%	0.6%	-57%	0.4%	1.1%	-66%
Swedish	0.3%	0.4%	-10%	0.2%	0.7%	-67%
Chinese	1.7%	3.9%	-56%	0.4%	1.7%	-76%
Spanish	1.0%	3.2%	-69%	0.6%	3.0%	-82%
Polish	0.7%	1.6%	-58%	0.2%	1.3%	-87%
Italian	1.1%	1.7%	-38%	0.1%	1.2%	-89%
Portuguese	0.3%	1.1%	-74%	0.1%	1.1%	-92%
Arabic	0.1%	0.6%	-90%	0.0%	0.5%	-100%
Other	2.3%	5%	-57%	1.1%	4%	-76%

Source: Compiled from Thomson ISI's and Ulrich's data.

To determine the role of language factors in journal coverage rates in the Thomson ISI databases, we also considered the actual language of journals. Table 4 shows a clear selection bias in favour of journals in which the articles are written in English. Whereas 75% of peer-reviewed journals indexed in Ulrich are in English, almost 90% of those selected by Thomson ISI are in English, yielding an over-selection rate of about 20%. The only other over-represented language in the Thomson ISI databases is Czech, a language that plays a marginal role in the global science system. French is under-represented by 26%.

Discussion and conclusion
Bibliometric analyses reflect the biases of the databases used and because of the marriage between bibliometrics and ISI's databases, bibliometric analysis can be said to by-and-large reflect the limits of ISI's databases. The findings presented in this paper show when SSCI and AHCI coverage is aggregated, there is a 20-25% bias in favour of English-language scientific output in the SSH.

Thus, these data clearly show that Thomson ISI's SSH journal selection favours English. This may be due to differences in the quality of research output according to the language of articles. According to Hodgson and Rothman (1999), 388, or 84%, of the 463 editors of the 30 most prestigious economics journals are affiliated with U.S. institutions. This might be a sign that English is dominant in Thomson ISI's simply because it is the favoured language for high quality research. Thus, the over-representation of English in the databases would after all be a fair or justifiable reflection of scientific dynamics. This view is supported by some results by Zitt (2003), which show that the exclusion of national oriented journals has a positive effect

on the calculation of their "national impact factor". Evidence such as this would support the view that national oriented journals are of little interest to academics outside of their country. In the opinion of Moed, Nederhof and Luwel (2002), truly academic research should be relevant internationally and local orientation should not be a factor in bibliometric evaluations. From this perspective, research not covered by the Thomson ISI databases simply fails to reach the relevance threshold that would warrant closer evaluation.

Table 4. Journal coverage rates by language of article

Language	NSE			SSH		
	Thomson ISI	Ulrich	Difference	Thomson ISI	Ulrich	Difference
English	89%	78%	13%	90%	75%	20%
Czech	0.04%	0.3%	-85%	0.2%	0.2%	8%
Russian	0.5%	0.9%	-48%	0.3%	0.4%	-24%
French	3.3%	3.4%	-3%	3.2%	4.4%	-26%
Multiple languages	0.2%	0.2%	-14%	0.3%	0.5%	-45%
Dutch	2.2%	2.2%	0%	1.3%	2.6%	-48%
German	3.2%	3.9%	-18%	3.0%	5.8%	-50%
Japanese	0.4%	1.7%	-74%	0.2%	0.6%	-64%
Swedish	-	0.1%	-100%	0.1%	0.4%	-69%
Spanish	0.6%	2.6%	-75%	0.9%	3.0%	-69%
Italian	0.1%	0.8%	-83%	0.2%	1.1%	-80%
Danish	0.04%	0.1%	-50%	0.1%	0.3%	-83%
Portuguese	0.1%	0.7%	-85%	0.1%	1.0%	-86%
Chinese	0.3%	2.4%	-88%	0.04%	1.2%	-96%
Polish	0.05%	0.7%	-92%	-	0.9%	-100%
Arabic	-	0.1%	-100%	-	0.3%	-100%
Turkish	0.01%	0.2%	-95%	-	0.1%	-100%
Other	0.3%	1.6%	-80%	0.3%	2.4%	-87%

Source: Compiled from Thomson ISI's and Ulrich's data.

However, it is questionable whether research articles written in languages other than English are of lower quality in such a high proportion as the bias observed in our data. In fact, it may be difficult to rely solely on Thomson ISI to be the impartial judge of what is and is not quality research output. This position is made the more doubtful considering Thomson ISI self admitted inability to analyse the content of journals in languages other than English — a fact that is clearly stated on Thomson ISI's website:

> English language article titles, abstracts, and keywords are essential [for inclusion in Thomson ISI's databases]. English language cited references are also recommended. Although important scientific information is published in all languages, authors must provide English translations of article titles, author keywords, and abstracts if they hope to reach the widest possible audience. Likewise, as a purely practical matter, it would not be feasible for ISI to take on the task of translating this material[3].

Because its selection criteria require journal bibliographic information to be in English, Thomson ISI may fail to index the content of an excellent journal in philosophy, for example, because its content is only presented in German.

---

[3] http://www.isinet.com/essays/selectionofmaterialforcoverage/199701.html/

The original contribution of this paper is to shed light on how the bias of bibliographic databases has a particularly important effect in the benchmarking of national output in SSH research due to the greater importance of local journals and languages. Having shown that the staple databases used in bibliometric analyses do not espouse well the geographic and linguistic distribution of scientific production, it appears essential to reiterate the warnings already made elsewhere on the danger of relying on these tools to produce rankings of countries in terms of SSH output. It is essential to factor the bias into any international comparative analysis of the SSH since any benchmarking based on SSCI and AHCI will overestimate the production of English speaking countries such as the US, the UK and Canada and underestimate that of Germany, Spain, France and other non-English-speaking countries, and this bias will affect both publication counts and citation analyses.

References

Academy of Finland. (2003). *Scientific Research in Finland: A Review of Its Quality and Impact in the Early 2000s.* Publications of the Academy of Finland.

Andersen H.1996. Acta Sociologica pa den internationale arena. *Dansk Sociologi*, 7(2), 72-79.

Andersen H. 2000. Influence and reputation in the social sciences - how much do researchers agree? *Journal of Documentation*, 56(6), 674-692.

Braun, T., Glänzel W. & A. Schubert (2000). How Balanced is the Science Citation Index's Journal Coverage? A Preliminary Overview of Macro-Level Statistical Data. in Cronin, B. and Barsky Atkins, H. (Eds). *The Web of Knowledge: A Festschrift in honor of Eugene Garfield* (pp. 251-280). ASIS&T Monograph Series.

CNER. 2002. *Évaluation de la recherche publique dans les établissements publics français*. Paris: Comité national d'évaluation de la recherche.

Garfield, E. 1996. A bibliometric analysis of science journals in the ISI database. *Science Citation Index. Journal Citation Reports*. Printed guide to the microfiche edition of the SCI JCR.

Gingras, Y. (1984). La valeur d'une langue dans un champ scientifique. *Recherches Sociographiques*, 25(2), 286-296.

Glänzel, W. (1996). A bibliometric approach to social sciences: National research performance in 6 selected social science areas 1990-1992. *Scientometrics*, 35(3), 291-307.

Glänzel, W. & Schoepflin U. (1999). A Bibliometric Study of Reference Literature in the Sciences and Social Sciences. *Information Processing and Management*, 35, 31-44.

Godin B. 2002. *The Social Sciences in Canada: What Can We Learn From Bibliometrics? Project on the Measurement of the Social Sciences*. Working Paper No. 1.

Hicks D. (1999). The difficulty of achieving full coverage of international social science literature and the bibliometric consequences. *Scientometrics*. 44(2), 193-215.

Hicks D. (2004). The Four Literatures of Social Science. In Moed, H. (Ed). *Handbook of Quantitative Science and Technology Research* (pp. 476-496). Dordrecht: Kluwer Academic.

Hodgson G.M. & Rothman H. 1999. The Editors and Authors of Economics Journals: A Case of Institutional Oligopoly? *The Economic Journal*. 109, F165-F186.

Ingwersen, P. (1997). The Central International Visibility of Danish and Scandinavian Research 1988-1996: A General Overview of Science & Technology, the Humanities and Social Sciences by Online Publication Analysis. *CIS Report 5.3*. Copenhagen: Centre for Informetric Studies, Royal School of Library and Information Science.

Ingwersen, P. 2000. The international visibility and citation impact of Scandinavian research articles in selected Social Science fields: The decay of a myth. *Scientometrics*. 49(1), 39-61.

Katz, J.S. & Hicks, D. 1998. Indicators for Systems of Innovation - a Bibliometrics-based Approach. *IDEA Paper*.

Kyvik, S. 1988. Internationality of the social sciences: The Norwegian case. *International Social Sciences Journal*. 115, 163-172.

Line, M.B. (1999). Social Science information - the poor relation. *65th IFLA Council and General Conference*.

Mela, GS, Cimmino, MA &Ugolini, D. (1999) Impact assessment of oncology research in the European Union. *European Journal of Cancer*. 35(8), 1182-1186.

Moed, H.F., Nederhof A.J. & Luwel M. (2002). Towards performance in the humanities. *Library Trends*. 50, 498-520.

Nederhof, A.J. & Noyons E.C.M. 1992b. International comparison of departments' research performance in the humanities. *Journal of the American Society for Information Science*. 43(3), 249–256.

Nederhof, A.J. & Zwaan, R.A. (1991). Quality Judgments of Journals as Indicators of Research Performance in the Humanities and the Social and Behavioral Sciences. *Journal of the American Society for Information Science*. 42(5), 332-340.

Nederhof, A.J., Zwaan, R.A., Debruin, R.E. & Dekker, P.J. (1989). Assessing the Usefulness of Bibliometric Indicators for the Humanities and the Social and Behavioral Sciences: A Comparative Study. *Scientometrics* 15(5-6), 423-435.

Schoepflin, U. 1992. Problems of Representativity in the Social Sciences Citation Index. *in* Weingart P., Sehringer R. and Winterhager M. (eds.). *Representations of Science and Technology: Proceedings of the International Conference on Science and Technology Indicators, Bielefeld, 10-12 June 1990* (pp. 177-188). Leiden: DSWO-Press.

Van Raan, A.F.J. (2003). The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments. *Technikfolgenabschätzung*. 12(1), 20-29. English translation available: http://www.itas.fzk.de/tatup/031/raan03a.htm

Webster, B.M. (1998). Polish Sociology Citation Index as an Example of Usage of National Citation Indexes in Scientometric Analysis of Social Science. *Journal of Information Science*. 24(1): 19-32.

Winclawska, B.M. (1996). Polish Sociology Citation Index (Principles for Creation and the First Results). *Scientometrics*. 35(3), 387-391.

Yitzhaki, M. 1998. The Language Preference in Sociology. *Scientometrics*. 41(1-2), 243-254.

Zitt, M., Ramanana-Rahary, S. & Bassecoulard, E. (2003). Correcting glasses help fair comparisons in international science landscape: Country indicators as a function of ISI database delineation. *Scientometrics*. 56(2), 259-282.