

# What to Include in a Social Sciences and Humanities Citation Index - an Empirical Analysis

Gunnar Sivertsen<sup>1</sup> and Birger Larsen<sup>2</sup>

<sup>1</sup> [gunnar.sivertsen@nifu.no](mailto:gunnar.sivertsen@nifu.no)

Nordic Institute for Studies in Innovation, Research and Education (NIFU)  
Wergelandsveien 7, N-0167 Oslo, Norway

<sup>2</sup> [blar@iva.dk](mailto:blar@iva.dk)

Royal School of Library and Information Science, Birketinget 6, DK-2300 Copenhagen S, Denmark

## Abstract

A well-designed and comprehensive citation index for the social sciences and humanities has many potential uses, but has yet to be realised. Significant parts of the scholarly production in these areas are not published in international journals, but in national scholarly journals, in book chapters or in monographs. In this paper we investigate the potential for covering these literatures empirically by using a complete publication output data set from the higher education sector of an entire country (Norway). We find that the so far uncovered literatures are concentrated in relatively few publication channels, which should be promising for a more comprehensive coverage of the social sciences and humanities.

## Introduction and background

A well-designed and comprehensive citation index for the Social Sciences and Humanities (SSH) has many potential uses, but has yet to be realised. A recent initiative in this direction is the so-called “European Scoping Project” that was commissioned in collaboration between the research councils of France, Germany, Netherlands and United Kingdom and coordinated in 2009-2010 by the Science Policy Research Unit at the University of Sussex. The aim of the project has been to give advice on the feasibility and development of a robust bibliometric database for assessing the impact of all types of research output in the social sciences and humanities. The two most general and immediate recommendations in the report (Martin *et al.*, 2010) are (1) to define the criteria for inclusion of SSH research outputs and establish a standardised database structure for national bibliometric databases, and (2) to explore the option of involving a commercial supplier in the construction of a single international database. The report points at the producers of *Scopus* and *ISI Web of Science* as potential suppliers, but advises that work on the second step should begin only after significant progress has been made with the first recommendation. In this paper, we contribute with an empirical investigation of what is covered by *ISI Web of Science* and what is not, using a specific definition and delimitation of the scholarly literatures of the SSH.

A main challenge in the SSH is that the publication and citation patterns are not as concentrated in core international journals as can be seen in the natural sciences. This notion of “core journals” in the natural sciences aided Eugene Garfield in constructing the Science Citation Index. He found that the significant literature appears in a small number of large important journals, and that the returns from adding extra journals outside of this core are marginal from the point of view of research interest, measured as citedness (Garfield, 1979). However, significant parts of the scholarly production in the SSH are published in national journals, book chapters and monographs (Hicks, 2004). As a result of this diversity, the challenge of setting criteria for the selection of source items for the SSH is seen as much greater than for the sciences (Martin *et al.*, 2010). The question still remains whether there are concentration effects even in the more heterogeneous publication patterns of the SSH. We do indeed find such effects in our study. These findings may have direct relevance for the feasibility of a citation index for the SSH.

An important background for our study is that it has become generally accepted in the SSH during the last decades that publications presenting new results from research should be peer reviewed. Peer review is an activity with limited resources that as a rule takes place in publication channels (a journal, a series, or an academic book publisher) that are shared by authors and reviewers from more than one institution. Thus, the peer reviewed scholarly literature cannot be unlimited. It must be concentrated in a limited, but dynamically changing number of publication channels that are acknowledged for competent publishing procedures involving peer review. Researchers in the SSH also often publish without peer review for students and the general public. According to its purpose, this literature appears in a much wider number of publication channels and will always seem to be unlimited from the point of view of publication databases. Since the scope of our study is the feasibility of covering the SSH in a comprehensive citation index, we concentrate on the peer reviewed literature only. The existing citation databases in the sciences were designed to mainly cover peer reviewed literature.

## Methods

Our hypothesis is that SSH does display some level of concentration even when including document types beyond the international journal literature, and that current citation index coverage can be improved significantly by adding a modest number of publication channels from each country.

The analysis is based on the complete publication output during 2005-2009 in the higher education sector of Norway. This sector includes four large universities and 54 other smaller or specialized institutions in higher education with a total of more than 20 000 researchers.

Our data set has been extracted from those collected as part of a nationwide performance based funding model for the higher education sector implemented in Norway in 2005 (Sivertsen, 2006; Schneider, 2009). This 'Norwegian model' has been pointed at as one possible element in the solutions for better SSH coverage (Dolan, 2007; Hicks & Wang, 2009). The model has two components, a National Research Documentation system covering all higher education institutions and collecting publication data according to a definition, and a bibliometric indicator that weights publications according to publication type and the level (prestige) of the publication channel. Only the first component is used in this study, meaning that all publications are given the same weight.

The indicator in the model is based on complete data (in journals, series and books) for the scientific publication output at the level of institutions. The data for the indicator are produced by the institutions themselves in a shared quality assurance system that creates an open and transparent national database with references to all scientific publications from all institutions. Only the scientific and scholarly publications are covered completely according to an agreed definition. A scientific or scholarly publication must (Sivertsen, 2010):

1. present new insight,
2. be in a form that allows the research findings to be verified and/or used in new research activity,
3. be in a language and with a distribution that makes the publication accessible for a relevant audience of researchers,
4. be in a publication channel (journal, series, book publisher) with peer review.

In order to control the third and fourth requirement of this definition and to standardize the data, dynamic lists of so far 20,000 ISSN-titles and 1,000 publishers of books were created. These lists can be downloaded at <http://dbh.nsd.uib.no/kanaler/>. Suggestions for additions are received through the same web page. In addition to requirement 3 and 4, a publication channel must publish on behalf of authors from more than one institution to be included.

Publication counts are fractionalized among authors of the same publication, meaning that contributions by authors outside of Norway's higher education sector are not counted.

Three publication types are included:

- Articles (in ISSN-titles: journals, series, yearbooks)
- Book chapters (Articles or chapters in ISBN-titles)
- Books (ISBN-title)

The data set used in the present study covers the years 2005 to 2009 and is based on data from more than 38 thousand publications with at least one author from Norway (see Table 8).

The national database in Norway is based on import of references from ISI Web of Science by agreement with Thomson Reuters. These references are validated at the institutions and supplemented with non-ISI references. These routines allow for a measurement of the coverage in the ISI Web of Science for each subfield. Our analysis could also have included the similar database *Scopus* by Elsevier by matching our data with their journal lists. We did so in a preliminary analysis and detected that although *Scopus* covers a larger number of journals, the pattern of deficiencies in representing the more heterogeneous scholarly literature of the SSH is exactly the same. In our analysis for this paper, however, we have chosen to compare the data with ISI Web of Science only. We have classified all publications in two categories, "ISI and "Non-ISI".

Research in the SSH is sometimes more relevant to publish in language of the society or culture that is studied than in English. Our data show that Norwegian scholars in the SSH are bilingual in their scholarly publishing patterns. About half of the publications are in Norwegian and the other half in English or other international languages (e.g. in German in the subfield of German Studies). Publications in Norwegian are easily read in Denmark and Sweden, but not outside of Scandinavia. In our study, we have chosen the language of a publication channel as a marker of internationality. Publications in channels where the Scandinavian languages are most frequent, are classified as "Domestic", while publications in channels publishing in international languages are classified as "International".

All publications are classified in five major fields: humanities, social sciences, health sciences, natural sciences, and engineering. Note that psychology is classified in health sciences, not in the social sciences in our data.

## Results

We will first show the publication patterns for all major fields, not only the SSH. Then we will have a closer look at subfields in the SSH. Finally, we will provide some simple measures of the concentration of publication channels in SSH, looking at journal publishing and book publishing separately in the two dimensions domestic and international.

A total of 43% of the total Norwegian output 2005-2009 is produced by SSH: 20% by the humanities and 23% by the social sciences (Table 8). The SSH fields thus represent a sizable share of the output. As expected, the SSH fields display a markedly different publication profile compared to the health, engineering and natural sciences: SSH relies much less on journal articles (47%-56% vs. 72%-90%), SSH publishes markedly more in book chapters (39%-46% vs. 9%-27%), and SSH publishes an order of magnitude as whole books (5%-7% vs. 0.4%-0.8%). In addition, SSH relies on domestic languages to a much larger degree (50%-55% vs. 3%-18%), with as much as 97% of publications in international languages in the engineering and natural sciences (Table 9).

As expected from these figures, the coverage in the existing citation indexes is low: only 11%-20% for SSH vs. 61%-80% for engineering, health and natural sciences in ISI. In all the studied dimensions, the humanities represent the extreme case. The empirical results based on a whole country and over a five-year period thus support the claims that as a whole SSH does

require special effort in order to be covered by citation indexes to the same degree as the engineering, health and natural sciences.

**Table 8. Total number of publications and distribution over publication types (fractionalised) for Norwegian publications (fractionalised) across fields 2005-2009.**

Main field	Publications	Publication type		
		Articles	Book chapters	Books
Humanities	7,650 (20%)	3,590 (47%)	3,556 (46%)	503 (7%)
Social Sciences	8,717 (23%)	4,896 (56%)	3,408 (39%)	413 (5%)
Health Sciences	9,129 (24%)	8,209 (90%)	843 (9%)	77 (0.8%)
Engineering Sciences	5,173 (14%)	3,733 (72%)	1,405 (27%)	35 (0.7%)
Natural Sciences	7,556 (20%)	6,708 (89%)	816 (11%)	32 (0.4%)
Sum	38,225 (100%)	27,135 (71%)	10,029 (26%)	1,061 (3%)

**Table 9. Language distribution and coverage in the ISI citation indexes for Norwegian publications (fractionalised) across fields 2005-2009.**

Main field	Language		Coverage	
	International	Domestic	ISI	Non-ISI
Humanities	3,454 (45%)	4,196 (55%)	855 (11%)	6,794 (89%)
Social Sciences	4,371 (50%)	4,346 (50%)	1,753 (20%)	6,964 (80%)
Health Sciences	7,473 (82%)	1,656 (18%)	6,940 (76%)	2,189 (24%)
Engineering Sciences	4,999 (97%)	174 (3%)	3,180 (61%)	1,993 (39%)
Natural Sciences	7,342 (97%)	215 (3%)	6,078 (80%)	1,478 (20%)
Sum	27,639 (72%)	10,586 (28%)	18,807 (49%)	19,418 (51%)

The analysis of the main fields hides differences that exist among the SSH subfields (see Table 10). A few fields such as Economics and Geography (which in our database only includes the social science publications of the discipline) have profiles that are similar to engineering, health and natural sciences with large shares of journal articles and international language publication – and also quite good ISI-coverage (52%-57%). However, 11 of 19 humanities subfields have 10% or less coverage in ISI, and 5 of 10 social science subfields 20% or less. Only Linguistics in the humanities has 30% coverage, and only two additional fields 20% or more. Overall, there is great heterogeneity within the SSH in the three dimensions we study, but as a rule there is very limited coverage in the existing citation indexes because of book publishing and regional or national journals.

**Table 10. Distribution over Humanities and Social Science subfields by publication type, language and ISI coverage for Norwegian publications (fractionalised) 2005-2009.**

Humanities subfield	Publ.	Publication type			Language		Coverage	
		Articles	Bk chap	Books	Internat.	Domestic	ISI	Non-ISI
Archaeology and Conservat.	433	51 %	44 %	4 %	50 %	50 %	12 %	88 %
Architecture and Design	211	56 %	38 %	5 %	39 %	61 %	5 %	95 %
Art History	155	56 %	33 %	10 %	49 %	51 %	9 %	91 %
Asian and African Studies	144	44 %	47 %	9 %	91 %	9 %	20 %	80 %
Classical Studies	157	62 %	34 %	4 %	49 %	51 %	10 %	90 %
English Studies	230	33 %	58 %	9 %	88 %	12 %	19 %	81 %
Ethnology	211	49 %	45 %	6 %	40 %	60 %	7 %	93 %

Germanic Studies	164	35 %	58 %	7 %	93 %	7 %	14 %	86 %
History	994	46 %	46 %	7 %	38 %	62 %	19 %	81 %
Linguistics	526	58 %	39 %	3 %	81 %	19 %	30 %	70 %
Literature	372	57 %	39 %	4 %	28 %	72 %	10 %	90 %
Media and Communication	578	37 %	55 %	8 %	53 %	47 %	5 %	95 %
Musicology	202	54 %	41 %	5 %	35 %	65 %	11 %	89 %
Philos. & History of Ideas	599	58 %	34 %	8 %	41 %	59 %	10 %	90 %
Romance Studies	183	40 %	51 %	9 %	82 %	18 %	21 %	79 %
Scandinavian Studies	1122	33 %	61 %	6 %	11 %	89 %	0 %	100 %
Slavonic Studies	146	55 %	38 %	6 %	86 %	14 %	7 %	93 %
Theatre and Drama	75	64 %	33 %	3 %	53 %	47 %	7 %	93 %
Theology and Religion	1150	48 %	44 %	8 %	41 %	59 %	9 %	91 %
<b>Social science subfield</b>		<b>Articles</b>	<b>Bk chap</b>	<b>Books</b>	<b>Internat.</b>	<b>Domestic</b>	<b>ISI</b>	<b>Non-ISI</b>
Anthropology	338	55 %	37 %	8 %	66 %	34 %	16 %	84 %
Business & Finance	1562	60 %	36 %	4 %	62 %	38 %	23 %	77 %
Economics	607	78 %	20 %	2 %	78 %	22 %	57 %	43 %
Education & Educational Res.	2396	46 %	49 %	5 %	33 %	67 %	9 %	91 %
Gender Studies	234	46 %	51 %	3 %	31 %	69 %	7 %	93 %
Geography	480	79 %	19 %	2 %	77 %	23 %	52 %	48 %
Law	1278	63 %	29 %	8 %	29 %	71 %	4 %	96 %
Library and Info. Science	229	85 %	13 %	2 %	93 %	7 %	39 %	61 %
Political Science	941	46 %	50 %	4 %	63 %	37 %	28 %	72 %
Sociology	652	44 %	50 %	6 %	44 %	56 %	16 %	84 %

With such diversity, can there be sufficient concentration in the publication channels to make a broader SSH coverage manageable and viable? Table 4 presents the results of an analysis of the concentration of publications over publication channels split into journals and books. In the ‘Norwegian model’, more than 1000 scholarly publishers have been assessed and included in the model on the condition that they perform peer review and represent authors from more than one institution.

In order to produce the results in

Table 11 we ranked all publication channels in each category and dimension (journals versus book publishers, domestic versus international) according to the frequencies of publications. The results show that domestic publications are more concentrated in a few channels than international publications, and that book publishing is more concentrated in a few channels than journal publishing. As an example, we see in table 4b that only 15 academic publishers publish 75 per cent of Norway's scholarly books and book chapters in the SSH, and that the rest of the publications are spread among 115 other publishers. We also see that relatively few domestic scholarly journals publish a quite large share of Norway's journal articles. This means that it should be possible to cover large parts of the peer reviewed literature in the SSH if one is willing to cover the national as well as the international level and include book publishers that can respond to the same quality criteria that are now used for the inclusion of journals in *Scopus* and *ISI Web of Science*.

**Table 11. The concentration of publications in publication channels in the Social Sciences and Humanities for Norwegian publications (fractionalised) 2005-2009.**

<b>a. Journals</b>		<i>Domestic languages: 4119 publications</i>		<i>International languages: 4914 publications</i>	
Maximum share of publications		Minimum number of channels	Share of channels	Minimum number of channels	Share of channels
25 per cent		10	5%	95	4%
50 per cent		29	14%	358	16%
75 per cent		58	28%	916	40%
100 per cent		208	100%	2291	100%

  

<b>b. Book publishers</b>		<i>Domestic languages: 4735 publications</i>		<i>International languages: 3139 publications</i>	
Maximum share of publications		Minimum number of channels	Share of channels	Minimum number of channels	Share of channels
25 per cent		3	2%	8	2%
50 per cent		7	5%	29	6%
75 per cent		15	11%	93	20%
100 per cent		130	100%	468	100%

Only 8 book publishers cover more 25 per cent of all Norwegian SSH book publishing in international languages, and only 93 publishers cover 75 per cent. Table 5 shows the names and numbers of the 12 most frequent international book publishers in the Norwegian data.

**Table 12. List of the 12 most frequent publishers, covering 34% of all scholarly book publishing in international languages in Social Sciences and the Humanities.**

<b>Publisher</b>	<b>Publications</b>	<b>Publisher</b>	<b>Publications</b>
Routledge	259	Cambridge University Press	95
Ashgate	172	Edward Elgar Publishing	85
Oxford University Press	124	Peter Lang Publishing Group	83
Peter Lang	120	Berghahn Books	78
Palgrave Macmillan	115	Walter de Gruyter	74
Brill Academic Publishers	97	Cambridge Scholars Publishing	74

There is indeed concentration among SSH publication channels. Domestic journal and book publishing is concentrated in a few channels, and so is international book publishing. Only international journal publishing seems to be dispersed in many channels according to the pattern already known from existing citation indexes. It seems that the so far uncovered types of literature should be manageable, if not commercially viable, to cover in an international indexing service.

### Discussion and conclusion

Because of differences in methods, material and missions, the peer reviewed scholarly literatures of the social sciences and the humanities are not concentrated in a core of important international journals, such as in most of the natural sciences. On the other hand, if we regard book publishing and articles in domestic scholarly journals, which are also important in the social sciences and the humanities, these publications are for a large part concentrated in very few publication channels. It should therefore be manageable to integrate book publishing and



domestic journals in a comprehensive citation index for the social sciences and humanities. It would require a limitation to the peer reviewed literature and to define criteria for the inclusion of specific journals, series and book publishers. Even with these necessary limitations, an integrated citation index with comprehensive coverage of the humanities and social sciences would not only serve bibliometrics and research management in these fields with better data, but also create an improved bibliographic research infrastructure, which is the original and main purpose of all citation indexing.

## References

- Dolan, C. (2007). *Feasibility study: The evaluation and benchmarking of humanities research in Europe*. HERA, Humanities in the European Research Area (<http://www.heranet.info>).
- Garfield, E. (1979). *Citation indexing - its theory and application in science, technology, and humanities*. New York: Wiley.
- Hicks, D. (2004). The Four Literatures of Social Science. In H. Moed, Glänzel, W. and U. Schmoch (eds): *Handbook of Quantitative Science and Technology Research*. Dordrecht: Kluwer Academic Publishers, p. 473-496.
- Hick, D. & Wang, J. (2009): *Towards a Bibliometric Database for the Social Sciences and Humanities – A European Scoping Project*. Available: [http://www.dfg.de/download/pdf/foerderung/grundlagen\\_dfg\\_foerderung/informationen\\_fachwissenschaften/geisteswissenschaften/annex\\_1\\_en.pdf](http://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/annex_1_en.pdf) (Annex 1 of Martin *et al.*, 2010)
- Martin, B., Tang, P., Morgan, M., Glänzel, W., Hornbostel, S., Lauer, G., Lenclud, G., Lima, L., Oppenheim, C., van den Besselaar, P. & Zic-Fuchs, M. (2010): *Towards a Bibliometric Database for the Social Sciences and Humanities – A European Scoping Project*. Research report produced for DFG, ESRC, AHRC, NWO, ANR and ESF. Available: [http://www.dfg.de/download/pdf/foerderung/grundlagen\\_dfg\\_foerderung/informationen\\_fachwissenschaften/geisteswissenschaften/esf\\_report\\_final\\_100309.pdf](http://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/esf_report_final_100309.pdf)
- Schneider, J. (2009). An outline of the bibliometric indicator used for performance-based funding of research institutions in Norway. *European Political Science*, 8 (3), 364-378.
- Sivertsen, G. (2006). A bibliometric model for performance based budgeting of research institutions. In *Proceedings of the 9th International Conference on Science & Technology Indicators, 7-9 September 2006, Leuven, Belgium*. Leuven: Katholieke Universiteit Leuven – Steunpunt O&O Statistieken, p. 133-135.
- Sivertsen, G. (2008). Experiences with a bibliometric model for performance based funding of research institutions. In Gorraiz, J. and Schiebel, E. (eds). *Excellence and Emergence. A New Challenge for the Combination of Quantitative and Qualitative Approaches. Book of Abstracts. 10th International Conference on Science & Technology Indicators. Vienna, Austria. 17-20 September 2008*. Vienna (AUT): Austrian Research Centers GmbH, p. 126-128.
- Sivertsen, G. (2010). A performance indicator based on complete data for the scientific publication output at research institutions. *ISSI Newsletter*, 6(1), 22-28.