

Towards comprehensive bibliographic coverage of the scholarly literatures in the humanities and social sciences

Report from a working group

September 2010

At the invitation of Milena Zic-Fuchs, Chair of the Standing Committee for the Humanities (SCH) of the European Science Foundation (ESF), and Marc Heppener, ESF's Director of Science and Strategy Development, representatives of 26 Member Organizations of the ESF met in Brussels on 31 March 2010 to discuss *ERIH*,¹ its results so far, and the expectations of the Member Organizations with regard to its future. The meeting was attended by Sir Roderick Floud, Chair of the Standing Committee for the Social Sciences (SCSS).

At the conclusion of the meeting, it was decided to set up a small working group to provide a short report with recommendations on some of the larger issues that had arisen during the meeting. Istvan Kenesei (Hungary), Pascal Perrin (Belgium)² and Nigel Vincent (UK) agreed to be members while Gunnar Sivertsen (Norway) agreed to chair the group. Milena Zic-Fuchs and Sir Roderick Floud joined the group as chairs of the respective standing committees.

The group met in Brussels on 2 June 2010. The meeting was also attended by Nina Kancewicz-Hoffman, head of the Humanities Unit at ESF, and Balázs Kiss, head of the Social Sciences Unit, as well as by Marc Heppener.

In what follows we set out the group's recommendations with regard to the questions that were discussed by the Member Organizations on 31 March.

The "bibliographic approach"

The working group considers the original aim of ERIH – increasing the visibility of the humanities by documenting, reviewing and developing the infrastructure for scholarly communication and publishing – as still very important, and we will suggest further steps to be taken in this direction along a path which we will refer to as the "bibliographic approach", thereby developing some of the ideas that were discussed at the meeting on 31 March.

The bibliographic approach takes into consideration the need – primarily from the point of view of ongoing research, but also from the point of view of research management and assessment – to cover scholarly publications in both *books and journals* and in both *the humanities and the social*

¹ The **European Reference Index for the Humanities** (ERIH) has been under development under the leadership of the Standing Committee for the Humanities (SCH) of the European Science Foundation since 2002. The aim of ERIH is to enhance the global visibility of high-quality research in the humanities across all of Europe. ERIH consists of lists of carefully selected and categorized scholarly journals. The 'initial lists' are now being revised and finalized by expert panels, and are expected to be published by the end of 2010.

² Pascal Perrin left the working group in midsummer as he left FNRS.

sciences with searchable bibliographic references to the *publications themselves*, not only their publication channels.

The aim of the bibliographical approach is thus to make all the scholarly literatures – in print or electronically distributed – searchable and accessible across countries. The bibliographic references would thereby mirror as fully as possible the output of European research in the humanities and social sciences and show what diversity of European research actually means. This diversity would be reflected not only in national languages but also in national traditions within specific disciplines. Visibility and availability would be gained at the same time.

Scholarly literature is no longer searched in local libraries, but online in electronic library information systems. More and more often, a bibliographic reference will lead on to a free or purchasable full text of the publication. The increasing use of abstracts in international languages and of authors' information with electronic addresses also calls for an internationally integrated bibliographic research infrastructure. The limited coverage of the scholarly literatures of the humanities and social sciences in international journal citation indexes such as the *Web of Science* or *Scopus* can be seen as deficiencies, not in these commercial products themselves, but in the library information systems that they are purchased to be a part of. The "bibliographic" research infrastructure of the humanities and social sciences must not remain underdeveloped just because it is commercially unviable to create and maintain it. The library information systems need to be developed, standardized and integrated in an international infrastructure that gives complete coverage of and access to both the national and the international scholarly literatures. But this aim would be unattainable without definitions and thresholds that serve to delimit the scholarly literatures from other literatures and at the same time stimulate research quality. The role of research councils or other bodies in international collaboration could be to decide the definitions and standards and give advice with regard to coverage.

The working group recommends that, after the completion and publication of the revised lists by the end of 2010, the present journal lists are not expanded and revised before the bibliographic approach and use of lists in this context has been discussed by ESF Member Organisations, ESF Standing Committees for Humanities and Social Sciences and the relevant research communities. Our report is a contribution to the discussion.

Collaboration with the Social Sciences

We strongly recommend that further discussions and developments of bibliographic reference methods on behalf of ESF are made in collaboration with the social sciences (represented in ESF by the Standing Committee for Social Sciences - SCSS). Several journals in the social sciences are already listed in ERIH as some research domains of SCH and SCSS overlap, and the two areas face many of the same problems and challenges with regard to their publishing infrastructure, in particular that publications appear in both books and journals, in both national and international languages, while the coverage of the literature in *Web of Science* or *Scopus* is limited in both areas.

Towards a new bibliometric database?

A report that addresses many of the same problems as ERIH, including those just discussed, was produced for DFG, ESRC, AHRC, NWO, ANR and ESF in 2010: *Towards a Bibliometric Database for the Social Sciences and Humanities – A European Scoping Project* (hereafter: the SSH report). The ERIH meeting in March found much of value in the SSH report and we have continued to do so in the working group. The two most general and immediate recommendations in the SSH reports are to (p. 33):

1. Define the criteria for inclusion of SSH research outputs and establish a standardised database structure for national bibliometric databases
2. Explore the option of involving a commercial supplier in the construction of a single international database

The report advises that work on the second step should only begin after significant progress has been made with the first recommendation. In connection with this advice, some key operational issues are treated on pages 12-17 of the report that will be very important in relation to further discussions of what we suggest here as the bibliographic approach.

The two recommendations mentioned above will in principle be recognizable in our suggestions below. But we are in doubt whether it is either possible or desirable to build a separate and complete bibliometric database for the social sciences and humanities. No other sciences are served with such a database. The databases known as *Web of Science* and *Scopus* are not designed to be complete, and they are not built with bibliometrics as a main purpose. Their purpose is rather to provide researchers all over the world with a tool for bibliographic search and thereby access to the research literature. Only on this basis are these databases commercially viable. The main markets are not research administrations. These products are purchased by consortia of university libraries in all parts of the world on behalf of millions of daily users of the library information systems. Bibliometric databases are only derivatives and by-products of the primary bibliographic information systems. The working group is concerned that the primary bibliographic information systems are insufficient in the humanities and social sciences.

For a better developed research infrastructure

Tools for searching and accessing the research literature are an important part of *the research infrastructure* in the humanities and social sciences as they are in other research domains. Categorized lists of journals provide information facilitating search but do not make journal contents more accessible. Access to the journal articles themselves, or at least their bibliographical metadata, their abstracts and information about the authors, responds to the need of every scholar to search for relevant literature. The library information systems are part of the publicly funded research economies. They consist of purchased commercial products (like *Web of Science* and *Scopus*) supplemented by non-commercial bibliographic databases (library catalogues, national journal indexes, non-commercial indexes such as *Medline*). We give examples below that indicate that the relevant solutions for a better bibliographic coverage of the humanities and social sciences on the international level already exist in several instances on the national level. However, integration and standardization of procedures and data sources at the international level are missing.

Organizations representing research in the humanities and social sciences must now define criteria and standards that the library information systems will have to meet in order to be regarded as a well-functioning part of the research infrastructure in an age of internationalization and electronic information.

The “core” journals and the national literatures

The second main recommendation of the SSH report is to approach the two commercial suppliers of science citation indexes in the hope that they would be able to construct a single bibliometric database. However in present circumstances the working group cannot see this path leading to a solution for books (monographs and edited volumes). We believe it is possible for these commercial suppliers – within strict criteria – to cover all *international* journals and book series, but not the parallel national or regional journals and series, and not separately published books. We are convinced that books and the national research literatures will continue to be an important part of scholarly communication in the humanities and social sciences, and that bibliographic information about them should be *available for all researchers on the international level*. We must therefore have internationally integrated bibliographic data sources at the national level as a supplement to the commercial databases on the international level.

What seems to be a “neglect” of the humanities and social sciences by the two commercial suppliers is more deeply rooted in the commercial idea itself, which is that there is a *core* of large scientific journals that will cover most of the articles that are of interest in the searchable database. This idea rests on generalizations from bibliometric research, among them Bradford's law of scattering (1934), which estimates the exponentially diminishing returns of extending a search for references in science journals. Outside the core of journals, there are more potential new journals to be covered than there are new relevant articles to be found. Therefore, in principle, these commercial databases will not try to be complete in coverage, since they are only cost-effective on a commercial basis by covering the “core”.

Much of the research literature in the humanities and social sciences may be important, but is not currently published in a “core” of outlets. The patterns of specialization in these domains do not favour large common publication channels. Specialization may take place on levels where the research and results are mainly of interest within a certain linguistic, cultural, social or political context. Although we see clear trends towards internationalisation, the use of both the international and the national languages in the humanities and social sciences must be expected, since the missions and purposes in these disciplines point in that direction. Publishing in books as well as journals must also be expected to continue because both formats reflect the necessary research practices. In some disciplines, such as psychology, economics and general linguistics, there is certainly a core of important international journals (and these are well covered by the commercial suppliers), but in other disciplines, more heterogeneous publishing patterns will remain.

The working group is of the opinion that it is of the utmost importance to envisage and develop an international research infrastructure (which can also be seen as the basis for possible bibliometric developments) also for the literatures outside the cores of international scientific journals. It should be possible to create a pan-European research infrastructure (comparable to, e.g., *The European Library*) for bibliographic data covering both the national and the international literatures in both books and journals.

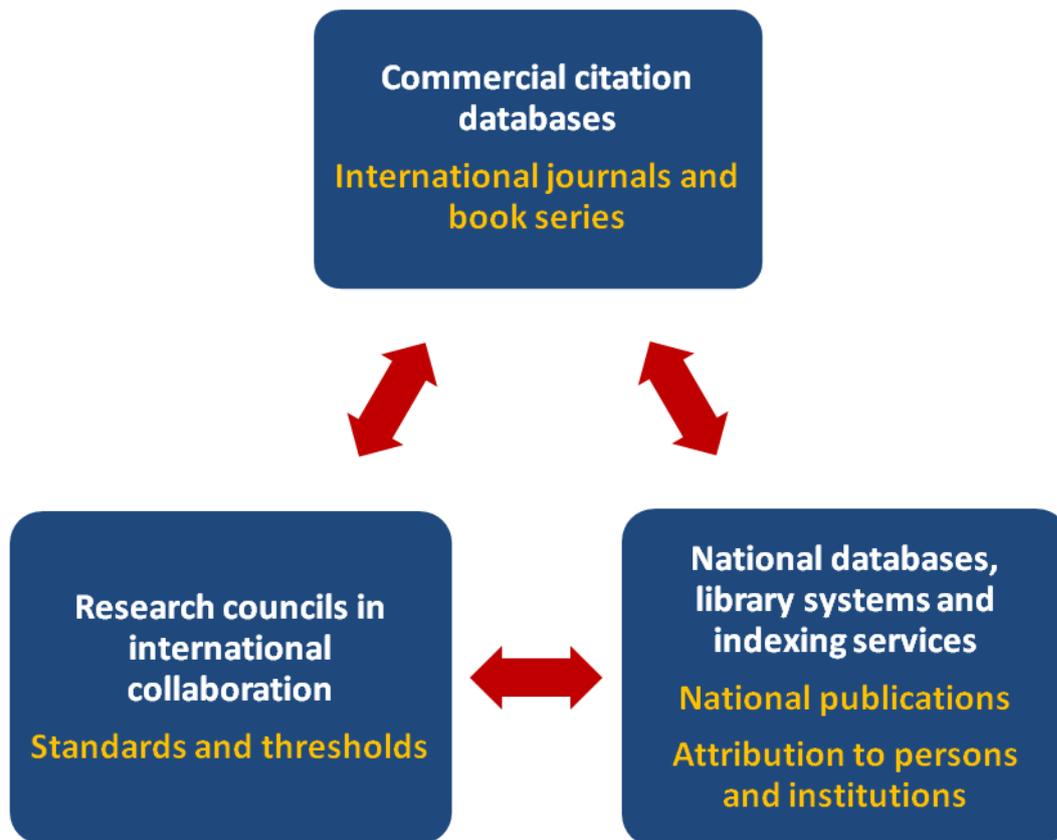
Research councils as collaborating advisers – and lists of publication channels as sources of advice

As seen above, commercially produced bibliographic information can probably only be a part of the complete solution. During the last two years, we have certainly seen an expansion of both *Web of Science* and *Scopus* with many added journals in the humanities and social sciences. In the case of *Scopus*, the ERIH lists were used selectively as a basis for this expansion. But we still see many inconsistencies in the coverage. The main national journal in one discipline may be covered, but not the corresponding journal in the neighbouring discipline or country. That said, since the commercial suppliers cover the costs, they are entitled to decide what journals to cover. Other parties can only give advice.

But the recent use of ERIH for *Scopus* provides a model for progress. The ERIH lists now have two main categories, national and international. Similar lists (covering the social sciences as well) could function as the basis for advising what journals should be covered in different parts of the library information systems. The lists could be updated annually for this purpose. The advice could then be that all *international* journals corresponding to certain criteria are covered by the commercial citation databases while all other journals corresponding to certain criteria should be covered by *internationally integrated* national journal indexes. One of several examples of the latter is the Portal of Scientific Journals of Croatia: <http://hrcak.srce.hr/?lang=en>.

The national indexes would be advised to cover all journals above a certain threshold that are not currently covered by the international indexes. The criteria for inclusion in both parts of the lists should be operational and refer to verifiable information: for international journals, one could demand peer review, an international editorial board, and articles and citations coming from several countries. For other journals, the requirements could be peer review and articles representing several institutions. Both categories of journals should also meet technical requirements such as timeliness, abstracts in international languages and other metadata and standards needed to integrate national indexing services into a common data source. (The Croatian portal mentioned above shows that such requirements can be implemented.) With such a threshold, the national library systems and indexing services cannot argue that the number of journals and articles that need to be indexed is unmanageable. At the same time, the internationally collaborating research councils will be stimulating good publishing practices by providing visibility and respectability to journals on both the international and the national level.

The figure below illustrates how the internationally collaborating research councils may stimulate a dynamic towards a better bibliographic coverage of the humanities and social sciences on the international level.



Books and book series

Book series contain books that are selected and edited in a way that resembles the practice in journals. They have editors who are themselves leading figures in the field, and peer review is performed by independent experts. They may publish both monographs and anthologies. The series often have a general ISSN in addition to the ISBN of the individual titles. Publishing in book series is expanding in the humanities and social sciences because it allows for less costly and more effective distribution of the more specialised research. While the libraries purchase printed copies on demand, electronic copies can be distributed directly to the user, either after a small payment or through open access. Books series can be handled in the same way as the journals in the model explained above. This means that the international book series could be covered selectively by international and national journal indexing services. Anthologies must be indexed down to the level of articles.

For books that are published outside book series, one solution may be to integrate the national, general library book catalogues and make them record anthologies with bibliographic references down to article level. An example of a comprehensive solution in this direction is the international database for economics, ECONIS, which is being produced by the German National Library of Economics: http://www.zbw-kiel.de/e_catalogues/e_econis.htm. Here, the standards and solutions have already been found for a decentralized but integrated coverage of the national and international literatures of the humanities and social sciences.

National and institutional databases

National databases are mentioned in the figure above. They cover the scholarly literatures comprehensively, but only at a national level, and are therefore not primarily used for literature search. An example of a national database is the Norwegian FRIDA, which records all scholarly publications from all Norwegian universities and thereby gives complete coverage of publication data from the humanities and social sciences, but only at national level:

<https://wo.uio.no/as/WebObjects/frida.woa/1/wa/default?la=en&inst=184>

The experience from building national or institutional databases at universities is that scholarly publications on the national level tend to be concentrated in relatively few publication channels (journals, series and book publishers). Such concentration, given that relatively strict criteria are adopted for defining “scholarly publication channels”, is promising for the idea of integrating bibliographic data on both the international and national levels.

Citation indexing

A search for literature in a bibliographic information system can be enhanced by so-called citation indexing, which is the procedure of recording the total reference list in a publication in addition to the publication’s bibliographic data, and then linking these references to the corresponding bibliographic data of the cited publications in the same database. In a bibliographic database, this procedure enhances the efficiency of literature search. In bibliometrics, the procedure provides for the option of measuring the impact of a publication in subsequent publications.

Even with limited coverage of the literature, citation indexing is already useful and promising for both literature search and bibliometric measurement in some disciplines in the humanities and social sciences, but in other disciplines, the usefulness and the validity of indicators is still very limited. Citation indexing of books is needed, and the recorded time span between citing and cited documents has to be expanded to reflect the specificity of humanities and social sciences research developments. If this is accomplished, citation indexing of the scholarly literature of the humanities and social sciences may become very useful for both purposes.

All journals and book series included in *Web of Science* or *Scopus* will have citation indexing, but this is usually not the case in the supplementary data sources on the national level. A pilot study has been launched this year by the Research Institute for Linguistics at the Hungarian Academy of Sciences where the aim is to gain experience with citation indexing of all Hungarian journals in the humanities and social sciences. Note that harvesting the national publications for citations will not only give visibility to this level, but will also enhance the significance of the international journals, references to which in national publications are not recorded in the commercial citation databases. We suggest that pilot studies with citation indexing of literatures outside of the commercial databases are conducted in parallel to the general development of the integrated bibliographic infrastructure, which must at this stage also include integration of data sources that are without citation indexing so far.

The role of journal lists

Since our mandate concerns the “post-ERIH” situation, we conclude here more specifically with regard to the role of categorized lists of selected journals. We have pointed out that lists of *scholarly publication channels* (journals, book series, and scholarly publishers of books) may have a function in advising commercial and non-commercial, international and national, suppliers to the library information systems that these publication channels need to be covered at the level of the individual publication with standardized bibliographic data in data sources that are integrated on the international level. The categorization of the publication channels into national versus international is then not meant as a quality assessment, since both the national and the international literatures are necessary and fulfil different roles in the humanities and social sciences. The categorization instead points at a possible division of labour between the suppliers of bibliographic information systems.

The role of such lists of scholarly publication channels would then be to serve as advice that they should be covered in an internationally integrated library information system. As general effects, this would stimulate:

- high-quality research communication
- cost-effective bibliographic coverage in the library information systems
- complete and reliable publication and citation data

We suggest that this advice is given in the form of dynamic lists of scholarly publication channels that meet certain criteria. These criteria should be developed by ESF and be based on scholarly standards and the aim of efficient international bibliographic access to the national and international scholarly literatures.

The lists must be dynamic since the world of publication channels is changing all the time. The contents and the maintenance of the lists should also be the responsibility of ESF. As mentioned above, the criteria for inclusion in the lists should be operational and refer to verifiable information. This means that there is less need for the qualitative judgement of each publication channel that has been part of the process when the ERIH lists were created. Expert panels will rather be needed to review the lists as such and discuss questions of principle, while the daily maintenance can be an administrative task. We suggest that the information needed to decide whether the criteria are met should be supplied by the publication channels themselves and not be produced by ESF or the panels.

To produce bibliographic information or to integrate bibliographic information systems cannot be the responsibility of the European Science Foundation. ESF can only give advice from the point of view of research and research communication, that is, from the point of view of the users of bibliographic information systems, by collaborating with the producers.

First steps forward

The idea of the bibliographic approach will first need a discussion on the level of research councils. This discussion must include the possible organization and costs of keeping internationally collaborating research councils in the role of giving advice with regard to the coverage of the scholarly literatures in internationally integrated bibliographic information systems. The costs on the side of research councils should only be related to the role of giving advice. Research councils can neither fund nor produce bibliographic information systems. These systems already exist, they already have funding, and they can be improved by advice from the point of view of the users.

The next step for the bibliographic approach would then be to identify the important partners for collaboration. Major representatives of the academic libraries and their information systems, as well as of the main commercial suppliers to such systems, should be approached in order to discuss the options and their costs and gains from building a network of national bibliographic and citation index databases using similar methodologies and supported by similar software.

We would further recommend that the databases included in any such network should ensure coverage not only of printed books and journals but also of electronic outlets, which are of increasing importance in the humanities and social science research communities.

Since the “bibliographic approach” concerns both the national and the international scholarly literatures, and since the international level is not only European, we recommend that collaboration is also sought with non-European bodies akin to ESF that have an interest in developing the infrastructure for scholarly communication and publishing.