Towards more consistent, transparent, and multi-purpose national bibliographic databases for research output

*National bibliographic databases for research output collect metadata on universities' scholarly publications, such as journal articles, monographs, and conference papers. As this sort of research information is increasingly used in assessments, funding allocation, and other academic reward structures, the value in developing comprehensive and reliable national databases becomes more and more clear. **Linda Sīle**, **Raf Guns** and **Tim Engels** outline the challenges faced by those developing national bibliographic databases for research output, from the need for reliable (persistent) identifiers, through to the new and evolving contexts for data use.*

On 10-11 September, 31 research information professionals from 13 countries gathered in Antwerp, Belgium to exchange experiences and insights on the maintenance of national bibliographic databases for research output. These databases collect metadata on universities' research outputs (e.g. scholarly monographs, journal articles, conference papers); examples include Cristin in Norway or MTMT in Hungary. During the workshop, participants had the opportunity to learn about databases implemented all across Europe (see the full programme for details and presentations). In addition, discussion sessions revealed a series of challenges encountered in database work. Here, we summarise these challenges, their difficulties, and go on to propose possible solutions.

The impetus

Over recent decades it has become common to use research information in research assessments, funding allocation systems, and other reward structures within academia. One problem that keeps surfacing, however, is an absence of comprehensive research information. By now we're all familiar with journal impact factors and other publication and citation-based indicators and their use as proxies for research activity or some aspect of research quality. At the same time, thanks to initiatives such as DORA and the Leiden Manifesto, the belief that limited research information of this kind does not do justice to research is becoming more and more widespread. Some European countries have tried to remedy this need for comprehensive research information by setting up national bibliographic databases for research output.

The need for further discussion of the more practical aspects of national databases emerged from work carried out within the European Network for Research Evaluation in Social Sciences and Humanities (ENRESSH). The idea was to describe the current state of national bibliographic databases for research output in Europe specifically within the social sciences and humanities – the research fields for which the problem of coverage is most pressing. However, the need for reliable and comprehensive research information pertains to all knowledge domains, and all countries.

Persistent identifiers – the key?

Even though the various national databases differ in terms of implementation and organisation, the challenges faced are rather similar. These range from specific technical tasks such as record deduplication, to more theoretical debates on principles that should guide the work with national bibliographic databases. However, perhaps the most pertinent issue is the need for reliable (persistent) identifiers for all research information entities. Author identifiers, research organisation identifiers, digital object identifiers – all of these could, in the long run, establish a more stable research information environment where different datasets can be integrated across institutional, regional, and even national contexts.

Impact of Social Sciences Blog: Towards more consistent, transparent, and multi-purpose national bibliographic databases for research output

Page 2 of 3

But we are not there yet. Most national systems, if they rely on any register at all, rely on national registers of researchers or research organisations. Therefore these systems aren't necessarily well-suited to the increasingly international landscape. For example, a typical challenge pertaining to author identifiers stems from international mobility. Moving across national borders throughout an academic career is the reality in most research communities. Yet database designs are only now beginning to tackle the issue of foreign authors. If there is a national register of researchers used as the reference point for a national bibliographic database, should the records for authors with international background be treated in the same way? How about research outputs produced prior to one's arrival in the specific country? Should data on such research outputs be included to facilitate access to scholarly literature? Or should such outputs be excluded to keep intact the validity of bibliometric indicators used in, say, national research assessments?

Towards transparent, multi-purpose research information systems

These questions bring us to more of a metadebate. The ways in which national bibliographic databases for research output are used are not fixed. Even if databases have been implemented with a specific purpose in mind, such as performance-based funding allocation on a national level, new uses often emerge organically. Data may be used to allocate funding or evaluate research at lower levels (e.g. within faculties and departments at universities). Such bibliographic data are often a valuable source for bibliometric research or information retrieval more broadly. This creates new contexts for data use and, more importantly, highlights yet more aspects of databases that do not (inter)operate as well as they could. Identifying how data is used in ways beyond the original purpose of output databases appears to be a good source of ideas for improvement.

On the one hand, by opening up the databases (and, more broadly, research information systems) to a wide range of purposes, we add to the workload of those who maintain them. On the other hand, it seems worthwhile to treat these national bibliographic databases as multipurpose systems which can be of use not only for managerial purposes (e.g. strategic decision-making, assessment, funding allocation) but also for information retrieval and research purposes.

What next?

What the vivid discussions during the workshop demonstrated is the urgent need for cross-context communication. Experience exists, it just needs to be shared. Similarly, new developments are picked up and implemented unevenly. At the same time, long-established traditions continue to be maintained thus leading to persistent quality of research information in the long term. How, and to what extent, we can learn from each other is a question requiring more such events for research information specialists.

To facilitate exchange of information on national bibliographic databases for research output we have launched a dynamic overview of European databases. Currently, the overview is based on findings from two surveys, discussed recently in *Research Evaluation* (DOI: https://doi.org/10.1093/reseval/rvy016). In future, however, the overview will be broadened in scope and updated.

In light of this, we are happy to announce that, within the framework of ENRESSH, there will be a training school on databases for the social sciences and humanities research output for research information specialists. The training school will take place from 21-25 October 2019 in Poznań, Poland, and will offer good practice examples from across Europe and serve as a platform for continued discussions on current issues in database work.

Note: This article gives the views of the authors, and not the position of the LSE Impact Blog, nor of the London School of Economics. Please review our comments policy if you have any concerns on posting a comment below.

Featured image credit: Samuel Zeller, via Unsplash (licensed under a CC0 1.0 license).

About the authors

Impact of Social Sciences Blog: Towards more consistent, transparent, and multi-purpose national bibliographic databases for research output

Page 3 of 3

Linda Sīle is a PhD student in social sciences at the University of Antwerp. Her thesis explores the relation between context-specific features of bibliographic data and bibliometric knowledge. Her ORCID is: 0000-0003-1435-999X.

Raf Guns is coordinator of the University of Antwerp branch of the Flemish Centre for Research & Development Monitoring (ECOOM). His research interests include bibliometrics, data science, and social network analysis. His ORCID is: 0000-0003-3129-0330.

Tim Engels is head of the Department of Research Affairs and Innovation at the University of Antwerp. He supervises the University of Antwerp branch of ECOOM since 2009. His research focuses on research evaluation and publication patterns in the social sciences and humanities. His ORCID is: 0000-0002-4869-7949.