Comprehensiveness of national bibliographic databases for social sciences and humanities: Findings from a European survey

Linda Sīle[1],*, **Janne Pölönen[2]**, **Gunnar Sivertsen[3]**, **Raf Guns[1]**,
Tim C. E. Engels[1], **Pavel Arefiev[4]**, **Marta Dušková[5]**, **Lotte Faurbæk[6]**,
András Holl[7], **Emanuel Kulczycki[8]**, **Bojan Macan[9]**, **Gustaf Nelhans[10]**,
Michal Petr[11], **Marjeta Pisk[12]**, **Sándor Soós[7]**, **Jadranka Stojanovski[9,13]**,
Ari Stone[14], **Jaroslav Šušol[15]** **and Ruth Teitelbaum[16]**

[1]Centre for R&D Monitoring (ECOOM), Faculty of Social Sciences, University of Antwerp, Antwerp 2020, Belgium, [2]Federation of Finnish Learned Societies, Helsinki 00170, Finland, [3]Nordic Institute for Studies in Innovation, Research and Education, Oslo 0608, Norway, [4]Scientific Electronic Library—eLIBRARY.RU, Moscow 117246, Russia, [5]Slovak Centre of Scientific and Technical Information, Bratislava 811 04, Slovak Republic, [6]Danish Agency for Science and Higher Education, Copenhagen K DK-1260, Denmark, [7]Department of Science Policy and Scientometrics, Library and Information Centre of the Hungarian Academy of Sciences (MTA), Budapest 1051, Hungary, [8]Scholarly Communication Research Group, Faculty of Social Sciences, Adam Mickiewicz University in Poznań, Poznań 61-712, Poland, [9]Centre for Scientific Information, Rudjer Boskovic Institute, Zagreb 10000, Croatia, [10]Swedish School of Library and Information Science, University of Borås, Borås S-501 90, Sweden, [11]Research Office, Masaryk University, Brno 601 77, Czech Republic, [12]Research Centre of the Slovenian Academy of Sciences and Arts, Ljubljana 1000, Slovenia, [13]Department of Information Sciences, University of Zadar, Zadar 23000, Croatia, [14]The Planning and Budgeting Committee, The Council for Higher Education, Jerusalem 91000, Israel, [15]Faculty of Arts, Comenius University in Bratislava, Bratislava 814 99, Slovak Republic and [16]The Henrietta Szold Institute, Jerusalem 91000, Israel

*Corresponding author. Email: Linda.Sile@uantwerpen.be

Abstract

This article provides an overview of national bibliographic databases that include data on research output within social sciences and humanities (SSH) in Europe. We focus on the comprehensiveness of the database content. Compared to the data from commercial databases such as Web of Science and Scopus, data from national bibliographic databases (e.g. Flemish Academic Bibliographic Database for the SSH (VABB-SHW) in Belgium, Current Research Information System in Norway (CRISTIN)) are more comprehensive and may, therefore, be better fit for bibliometric analyses. Acknowledging this, several countries within Europe maintain national bibliographic databases; detailed and comparative information about their content, however, has been limited. In autumn 2016, we launched a survey to acquire an overview of national bibliographic databases for SSH in Europe and Israel. Surveying 41 countries (responses received from 39 countries), we identified 21 national bibliographic databases for SSH. Further, we acquired a more detailed description of 13 databases, with a focus on their comprehensiveness. Findings indicate that even though the content of national bibliographic databases is diverse, it is possible to delineate a subset that is similar across databases. At the same time, it is apparent that differences in national bibliographic databases are often bound to differences in country-specific arrangements. Considering this, we highlight implications to bibliometric analyses based on data from

1

national bibliographic databases and outline several aspects that may be taken into account in the development of existing national bibliographic databases for SSH or the design of new ones.

Key words: social sciences and humanities; SSH; research information; Europe; database

1. Introduction

One of the major challenges in bibliometrics-supported evaluation of research in the social sciences and humanities (SSH) is the absence of comprehensive bibliographic data suitable for bibliometrics. Both the Leiden Manifesto (Hicks et al. 2015) and the San Francisco Declaration on Research Assessment (DORA 2012) have highlighted the need to take into account the diversity of research output types across different knowledge domains. In SSH, scholars often communicate using a broad range of media (e.g. articles in national journals, monographs, and book chapters, in addition to articles in internationally oriented journals; see Hicks 2004). The challenge, however, in attempts to take the SSH specifics into account is the limited coverage of the often used international proprietary databases. Even though the coverage of SSH research output has been increasing in the Web of Science (WoS; Michels and Schmoch 2012), the share of SSH publications included in WoS remains rather low (e.g. Kulczycki et al. 2018). In using the data that do not reflect the richness of SSH research, there is a risk to marginalize socially relevant research or research carried towards ends, which are not captured using indicators based on the data on articles in internationally oriented journals. This is highlighted in, for example, the Leiden Manifesto (Hicks et al. 2015). Alternative data sources that can lead to more accurate insights into SSH research output are national bibliographic databases. A number of countries have set up national bibliographic databases (e.g. Flemish Academic Bibliographic Database for the SSH (VABB-SHW) in Flanders in Belgium, Croatian Scientific Bibliography (CROSBI) in Croatia, Current Research Information System in Norway (CRISTIN), and Russian Index of Science Citation (RINC) in Russia; Verleysen, Ghesquière, and Engels 2014, Stojanovski 1999, Sivertsen 2016, Moskaleva et al. 2017) or implemented other bibliographic data collection initiatives (e.g. Research Core Dataset in Germany; Biesenbender and Hornbostel 2016). Among the main goals of these initiatives is to achieve more comprehensive coverage of national research output, thus overcoming the limited coverage of commonly used citation databases (e.g. WoS and Scopus), especially with respect to SSH.

The need for comprehensive data suitable for bibliometrics of SSH has been acknowledged also on the European policy level (e.g. Martin et al. 2010, Mahieu et al. 2014). Furthermore, acknowledgment of the role that existing national databases may play in the enhancement of the visibility of SSH can be found in the memorandum of understanding of the COST Action 'European Network for Research Evaluation in Social Sciences and Humanities' (ENRESSH), a network launched in 2016 (COST Association 2015).

The use of national bibliographic data in bibliometrics-supported research evaluation, however, is challenged by limited information about the content of national bibliographic databases. For this reason, in autumn 2016, a study was launched to, first, identify currently existing European national bibliographic databases storing data on publications in SSH and, secondly, to determine the extent to which the currently existing national bibliographic databases are suitable for bibliometric explorations of SSH. Here, we present key findings of this study concerning the comprehensiveness of national bibliographic databases for SSH in Europe.

What follows is a description of methods that were used to identify and describe national bibliographic databases. In the findings section, we begin with an overview of the identified databases (n = 21). Then, we continue with a more detailed description of a selection of 13 databases. Finally, we discuss findings highlighting challenges that the various database set-ups pose for bibliometric analyses of the SSH. The supplementary material comprises the following: a list of questions included in the second of the two questionnaires we used (Supplementary Table S1; details on this follow); an overview of the 21 national bibliographic databases with information on the timespan of the bibliographic information included, on the inclusion of the most common research output types, and on the data collection approach (Supplementary Table S2); an overview of the 13 national bibliographic databases with information on several aspects of comprehensiveness, namely, types of research organizations, types of organizational units, seniority and job positions of authors, academic disciplines within SSH, language, and the intended audience of publications (Supplementary Table S3).

2. Methods

The study was conducted within the framework of the European Cooperation in Science and Technology (COST) action ENRESSH. The study was organized in two stages: in the first stage, our main aim was to identify national bibliographic databases in Europe and acquire some basic information on database setups (scope: 41 countries; responses received from 39 countries). In the second stage, we sought more detailed information on the content of a selection of the identified databases (scope: 17 databases; participation approved in relation to 13 databases).

2.1. Key terms

By 'bibliographic database' or 'database for research output', we mean a structured set of bibliographic metadata (e.g. title, publication type, year, and author) in line with requirements for data when calculating the most basic indicator of research output, namely, the number of publications, similar to that suggested by Moed and colleagues (Moed et al. 2009). We used this rather broad definition assuming it to be more appropriate in a context where information on data collection practices across countries is limited.

The term 'research output' denotes publications and other artefacts (both peer reviewed and non-refereed) communicating or representing results of scholarly inquiries to audiences of any kind. For databases that store data exclusively on publications, we use the term '[data on] publications' instead.

Definitions for several key terms we derive from the Frascati Manual (OECD 2015). The term 'social sciences and humanities' refers to those academic disciplines that are recognized as SSH

within the Fields of Research and Development (OECD 2015 pp. 57–9). The term 'research organization' is treated as a synonym to the term 'institutional unit' (EC et al., 2009: 61, para 4.2 cited in OECD 2015 p. 82). In operationalizing types of research organizations, we distinguish between two sectors: (1) higher education sector and (2) the three other sectors (business enterprise, government, and private non-profit). For the higher education sector, we identified universities, which typically have the right to confer doctorates, as a subset of all higher education institutions. We furthermore made a distinction between two general categories in terms of the sources of funding: (1) State and (2) Other.

Finally, 'comprehensiveness' here refers to the extent to which a certain database includes data on the total volume of research output. Within this study, the focus is on comprehensiveness in relation to the total volume of the SSH research output of a particular country. On the latter point, it should be highlighted that in this study we considered both databases that store data specifically on SSH (e.g. Lituanistika in Lithuania) and also more generic databases that include data on research output from any discipline (e.g. CROSBI in Croatia).

2.2. Stage 1: Identification of national bibliographic databases for SSH in Europe

Participants of the study were representatives of 39 of 41 countries within Europe and Israel (See Table 1). The main data collection instrument in Stage 1 was a questionnaire with 31 questions. The questionnaire as well as further information on methodological aspects of this first stage of the study can be found online in a report (Sīle et al. 2017). Here, we summarize findings concerning these questions:

1. Is there a national database on SSH research output?
2. What is the timespan for research output included in the database?
3. Which research output types are included in the database?
4. How are the data collected?

Answers to Questions 1–3 were summarized using the collected data without any adjustments. In cases where the comprehensiveness of a database is lower in earlier years, we use the time span during which the database covered the research output most comprehensively.

An overview of the 21 databases we identified and described can be found in the Supplementary Table S2. For some countries, we identified several national databases, yet in this study, we described only one for each country with an exception for Israel. Similarly, this overview is based on national databases that were reported as such by the study participants. Consequently, this overview does not contain data on, for example, PASCAL and FRANCIS in France, Digital.CSIC in Spain, and the project 'Research Outcomes' in UK. In addition, we identified that in Germany (Social Science Open Access Repository, SSOAR), Ireland (RIAN.ie Open Access in Ireland), Iceland (Opin Vísindi), and Portugal (The Scientific Open Access Repository of Portugal), there are national bibliographic databases (or repositories) that collect data specifically on open-access research output. Due to this focus, we do not consider these databases in this article.

Details on databases SSOAR, RIAN.ie, and Opin Vísindi can be found in the report (Sīle et al. 2017).

2.3. Stage 2: Content and comprehensiveness of 13 national bibliographic databases in Europe

Databases to be studied in the second stage were selected if they met the following criteria: (1) store data on more than one research output type and (2) include data on publications in more than one language (17 meet these criteria). Participation was confirmed concerning 13 databases in Belgium (Flanders), Croatia, the Czech Republic, Denmark, Finland, Hungary, Israel, Norway, Poland, Slovakia, Slovenia, Sweden, and Russia (see Table 1).

At this stage of the study, the main data collection instrument was a questionnaire consisting of 49 questions about the content of the database (description and bibliometric indicators), data processing, and technical specifications. Anticipated time required to complete the questionnaire: 8 h. The findings presented here are based on a small part of the data collected, namely, data on the content and comprehensiveness of databases (Supplementary Table S1).

The design of the second questionnaire was a result of a collaborative work among the first five authors of this text. In addition to the questionnaire, we developed a manual providing complementary information on each question.

We described comprehensiveness of databases in relation to different aspects of comprehensiveness. For example, we distinguished between academic units, referring to units that are tasked primarily with academic duties (e.g. departments, faculties), and administrative units, referring to those types of units that are tasked with administrative or other non-academic duties (e.g. library, finance department). This distinction helps to understand whether databases include publications authored by persons affiliated to, for example, a university department without academic duties. Another aspect is job positions of authors (e.g. are publications authored by doctoral students included?), academic disciplines within SSH (e.g. are publications from all SSH disciplines included?), language (e.g. are publications in any language included?), and the intended audience of publications (e.g. are publications addressed to the general public included?). An overview on these aspects of comprehensiveness is included in the Supplementary Table S3.

Next, we described procedures implemented to ensure comprehensiveness. A straightforward approach to make sure that a database captures data on all publications that fall within the inclusion criteria is to introduce a procedure whereby either authors or institutions reporting data confirm the completeness of data. By completeness we mean that all relevant research output is reported or made available. Another procedure is to link national and/or institutional incentives to data within a database: to introduce a mandate to report data, use data for research evaluation purposes, and/or use data for research funding allocation purposes. The category 'research evaluation' refers to any research evaluation activities that are not linked to funding allocation mechanisms. With incentives as procedures ensuring comprehensiveness we mean that the presence of incentives often leads to more comprehensive data.

Finally, to acquire a more general understanding on what principles guide the data collection, we produced narrative descriptions of 13 databases using the following structure:

Table 1. Overview of identified national bibliographic databases for SSH

Country	National bibliographic database[a]	Any academic discipline[b]	More than one research output type and more than one language[c]	Database analysed in the Stage 2
Albania	Database is being implemented	–	–	–
Austria	–	–	–	–
Belgium	Flemish Academic Bibliographic Database for the SSH (VABB-SHW)	No	Yes	Yes
Bosnia and Herzegovina	–	–	–	–
Bulgaria	–	–	–	–
Croatia	Croatian Scientific Bibliography (CROSBI)	Yes	Yes	Yes
Cyprus	–	–	–	–
Czech Republic	Registry of Information about Results (RIV)	Yes	Yes	Yes
Denmark	The Danish Bibliometric Research Indicator (BFI)	Yes	Yes	Yes
Estonia	Estonian Research Information System (ETIS)	Yes	Yes	–
Finland	VIRTA Publication Information Service (VIRTA)	Yes	Yes	Yes
France	–	–	–	–
Germany	–	–	–	–
Greece	Greek Reference Index for the Social Sciences and the Humanities (GRISSH)	No	No	–
Hungary	The Hungarian Scientific Bibliography (MTMT)	Yes	Yes	Yes
Iceland	–	–	–	–
Ireland	–	–	–	–
Israel	Database of Publications in the Social Sciences and Education	No	Yes	Yes
Israel	Index to Hebrew Periodicals (IHP)	Yes	No	–
Italy	LOGINMIUR	Yes	Yes	–
Latvia	Database is being implemented	–	–	–
Lithuania	Lituanistika	Yes	Yes	–
Luxembourg	–	–	–	–
Malta	–	–	–	–
Moldova	National Bibliometric Instrument (IBN)	Yes	No	–
Montenegro	–	–	–	–
The Netherlands	NARCIS	Yes	Yes	–
Norway	Current Research Information System in Norway (CRISTIN)	Yes	Yes	Yes
Poland	Polish Scholarly Bibliography (PBN)	Yes	Yes	Yes
Portugal	Database is being implemented	–	–	–
Romania	–	–	–	–
Russia	Russian Index of Science Citation (RINC)	Yes	Yes	Yes
Serbia	The Serbian Citation Index (SCIndeks)	Yes	No	–
Slovakia	Central registry of publication activity (CREPČ)	Yes	Yes	Yes
Slovenia	Co-operative online Bibliographic Systems and Services (COBISS)	Yes	Yes	Yes
Spain	–	–	–	–
Sweden	SwePub	Yes	Yes	Yes
Switzerland	–	–	–	–
Ukraine	–	–	–	–
UK	–	–	–	–
39 countries	21 databases	18 databases	17 databases	13 databases

[a]National bibliographic databases identified and described in the Stage 1.
[b]Database includes data on research output from any academic discipline.
[c]Database includes data on more than one research output type and in more than one language (for publications).

1. For which purpose(s) has the database been set up?
2. Which criteria are being used to decide upon the inclusion of data on publications within a database?
3. Which (implicit and explicit) exclusion criteria can be identified?
4. Who decides upon the inclusion criteria?

These narratives were written in collaboration with representatives of the 13 countries.

2.4. Limitations

Two main limitations concern the general approach of the study and the continuously evolving database setups. As noted earlier, in this study we adopted a rather broad definition of national bibliographic databases for SSH. We are aware that we list databases that have been designed specifically for the calculation of bibliometric indicators for research funding allocation purposes (such as The Danish Bibliometric Research Indicator (BFI) in Denmark) alongside

databases which have the main purpose of providing access to scholarly literature on a specific theme (e.g. Lithuanian studies in the database Lituanistika). We consider, however, that any well-structured bibliographic database of national scope is of value for bibliometric analyses of SSH, especially so when drawing upon multiple sources of data, thus addressing limitations posed when using a single data source that in some aspects may not be suited for bibliometrics of SSH.

The second limitation is linked to the observation that setups of databases are changing regularly. Data used in this study were collected and analysed from August 2016 to November 2017. During this period, some databases have been changed. Due to this, the reference point in time for data presented here is 1 July 2017.

3. Overview of national bibliographic databases for social sciences and humanities

In terms of the most common types of publications included in national bibliographic databases, the only publication type included in all the 21 databases is the journal article. Also, there are differences across countries in terms of the range of types of research output that can be reported to a national database. For example, there are databases that maintain an index of journal articles (e.g. The Serbian Citation Index (SCIndeks) in Serbia and the National Bibliometric Instrument (IBN) in Moldova). In contrast, there are also systems (e.g. CROSBI in Croatia, Registry of Information about Results (RIV) in the Czech Republic, Central registry of publication activity (CREPČ) in Slovakia, and Co-operative online Bibliographic Systems and Services (COBISS) in Slovenia) in which any type of research output can be reported. This is achieved, first of all, by using an extensive classification of research output types and, secondly, by introducing the open, unspecified category 'Other' in the classification that allows to report any other research output type.

Variations can be identified also in the timespan of research output included in the databases. While the timespan included in all databases is from 2011 onwards, half of the databases include research output beginning from 2001. Perhaps surprisingly, there are databases wherein a systematic collection of data goes back to the 1990s (e.g. CROSBI in Croatia, RIV in the Czech Republic, IBN in Moldova), 1980s (COBISS in Slovenia), or even 1970s (Index to Hebrew Periodicals (IHP) and Database of Publications in the Social Sciences and Education in Israel and LOGINMIUR in Italy).

The databases vary greatly in terms of the approach to data collection. The content for half of the databases (n = 11) is collected by means of data transfer. Most often (in eight databases), data are collected by means of data transfer from research organizations (e.g. universities, public research institutes, musea). This is an approach employed in, for example, VIRTA Publication Information Service (VIRTA) in Finland and CREPČ in Slovakia. The content of three databases is based on data transferred from publishers (Greek Reference Index for the Social Sciences and the Humanities (GRISSH) in Greece, IHP in Israel, and SCIndex in Serbia). In seven databases, data are reported manually. For three databases, manual reporting is done by authors or specialists within the reporting organizations (CROSBI in Croatia, Estonian Research Information System (ETIS) in Estonia, LOGINMIUR in Italy); in three other cases, data are entered in the database by staff maintaining the database (Database of Publications in Social Sciences and Education in

Israel, Lituanistika in Lithuania, IBN in Moldova). Finally, the content of four databases is collected by combining two or more methods. In one case, manual input by authors is combined with data transfer from Scopus (CRISTIN in Norway). Data in the Hungarian Scientific Bibliography (MTMT) (Hungary), RINC (Russia), and COBISS (Slovenia) are collected by combining manual data input (by authors or librarians) with data transfer from research organizations, publishers and other national or international databases (e.g. Web of Science and/or Scopus; for details see the Supplementary Table S2 and Sīle et al. 2017).

4. Comprehensiveness of 13 national bibliographic databases

In this section, we provide a summary of findings of the second stage of the study aimed at more detailed understanding of the content and comprehensiveness of 13 national bibliographic databases for SSH (see Table 1). First, we provide a summary of the findings concerning criteria that are used to decide upon the inclusion of data in a given database. Then, an overview of various aspects of comprehensiveness and procedures assuring the comprehensiveness of databases is presented (details on each database can be found in the Supplementary Table S3). Third, we provide narrative descriptions of the databases.

4.1. Summary of similarities and differences in inclusion criteria

It is possible to distinguish between two main approaches pertaining to inclusion criteria employed in national bibliographic databases for SSH. In databases like BFI (Denmark), VIRTA (Finland), and VABB-SHW (Flanders in Belgium), the inclusion of data on publications is based on a single general definition of publications. Such a definition typically specifies requirements which each publication (regardless of its type) needs to meet to be included in a database. For example, an often used requirement is that publications must be peer reviewed prior to their publishing.

Another approach to decide upon the inclusion of data in databases is to use a detailed classification of types of research output. In such classifications, each output type is defined specifying certain requirements that have to be met. This approach is followed by RIV (the Czech Republic), CROSBI (Croatia), COBISS (Slovenia), CREPČ (Slovakia), and CRISTIN (Norway). For the Polish database Polish Scholarly Bibliography (PBN), a combination of a general definition and a detailed classification is used.

A point to highlight is that many databases contain a subset of data that is used to calculate bibliometric indicators for research funding allocation purposes (e.g. BFI in Denmark; VIRTA in Finland; CREPČ in Slovakia; CRISTIN in Norway) or to transfer data to a national CRIS (COBISS in Slovenia). In such cases, inclusion criteria for that subset are stricter and differ from criteria applied to all records in a database (e.g. BFI in Denmark; VIRTA in Finland; CREPČ in Slovakia; COBISS in Slovenia; and CRISTIN in Norway).

In two databases, the approach to decide upon inclusion criteria differs from the above two approaches. In RINC (Russia), a large part of the data is collected directly from publishers. Up until 2017, the focus of that database was on scientific publications without explicit inclusion criteria. Similarly, the Swedish database SwePub is based on data harvested from bibliographic databases maintained

Table 2. Overview of procedures ensuring comprehensiveness in 13 national bibliographic databases

Database	Completeness is confirmed by authors	Completeness is confirmed by reporting organizations	Data are linked to national and/or institutional incentives		
			Data reporting[a]	Research evaluation[b]	Research funding allocation[c]
Belgium (Flanders): VABB-SHW	No	Yes	Yes	Yes	Yes
Croatia: CROSBI	No	Not applicable	Yes	Yes	Yes
Czech Republic: RIV	No	Yes	Yes	Yes	Yes
Denmark: BFI	No	No	No	No	Yes
Finland: VIRTA	No	Yes	Yes	No	Yes
Hungary: MTMT	Yes	Yes	Yes	Yes	Yes
Israel: Database of Publications in the Social Sciences and Education	Not applicable	Not applicable	No	Yes	No
Norway: CRISTIN (NVI)	Yes	Yes	Yes	No	Yes
Poland: PBN	No	Yes	Yes	Yes	Yes
Russia: RINC	Yes	Yes	Yes	Yes	Yes
Slovakia: CREPČ	No	No	Yes	Yes	Yes
Slovenia: COBISS	Yes	Yes	Yes[d]	Yes	Yes
Sweden: SwePub	No	No	Yes[e]	No	No

[a] Data reporting is mandated on national or institutional level.

[b] Data are used for research evaluation purposes.

[c] Data are used to calculate bibliometric indicators for research funding allocation purposes.

[d] Applicable only to certain publication types.

[e] Data reporting is mandated on institutional level (in some institutions). There is no national-level mandate.

by Swedish research organizations. Consequently, inclusion criteria in SwePub are dependent on the criteria used across the various organizations.

4.2. Aspects of comprehensiveness

We inquired how comprehensive databases are in terms of specific aspects of comprehensiveness. We found out that, in relation to different job positions of authors, all databases collect data authored by academic staff (in CREPČ in Slovakia, only those in full-time positions) and doctoral students. Of 13 databases, 11 databases include also publications by administrative and technical staff as well as master level or other students. Concerning academic disciplines within SSH, all but one database collect publications from any SSH disciplines, the exception being the Database of Publications in the Social Sciences and Education in Israel which is focused on publications in the social sciences. Noteworthy, nearly all (n = 11) databases are general databases that collect data on research output from any academic discipline, exceptions being VABB-SHW in Flanders, Belgium, and the aforementioned database in Israel.

The language of publications is used as an inclusion criterion in one database: the Database of Publications in the Social Sciences and Education in Israel collects data on publications in English or Hebrew. Finally, we identified that the intended audience of publications is not used as a criterion in any of the databases. However, as we will show in the narrative descriptions of the 13 databases, criteria that are used to delineate subsets of data for research evaluation and funding allocation purposes, sometimes implicitly emphasize those publications that address a scholarly audience.

In terms of research organization types, we find that there are databases such as RIV in the Czech Republic, RINC in Russia, and COBISS in Slovenia where all research organizations are included, regardless of the sector they belong to and/or the source of funding. In contrast, the databases in Flanders (Belgium, VABB-SHW) and

Denmark (BFI) include data primarily from universities – with additional data on publications from higher education institutions (in VABB-SHW from 2000 to 2012) and from university hospitals (in BFI). Concerning organizational units, all databases collect research output by authors affiliated to academic units. Output linked to administrative units, however, is collected in 11 of 13 databases.

4.3. Procedures ensuring comprehensiveness

Table 2 presents an overview of the procedures implemented to ensure comprehensiveness. Most often, the completeness of data is confirmed by organizations reporting data (8 of 13 databases). Less often the completeness is confirmed by authors; this approach is used in four databases and only in combination with confirmation of completeness by reporting organizations. For the Database of Publications in the Social Sciences and Education in Israel, this question does not apply since the database is created as an information source for the general public. Completeness in this case is understood as a result of systematic work in the collection expansion carried out by people maintaining the database.

We identified that the data within all databases are linked to national or institutional incentives. For 11 databases, data reporting is mandated on national or institutional level. Similarly, data from nine databases are used to calculate bibliometric indicators for research evaluation purposes. Further, data from nearly all (11 out of 13) databases are used to calculate bibliometric indicators for research funding allocation purposes.

To sum up, all databases employ at least one procedure that ensures comprehensiveness of a database. Whereas completeness confirmation most often is asked from reporting research organizations, in terms of incentives, data within databases are typically used to calculate bibliometric indicators for research funding allocation purposes.

4.4. Principles guiding the data collection in 13 national bibliographic databases for social sciences and humanities

4.4.1. *VABB-SHW in Belgium (Flanders)*

VABB-SHW was created for the purpose of research funding allocation across universities in Flanders, the Dutch-speaking part of Belgium. For VABB-SHW, general principles concerning research output to be included have been decided by the Flemish Government (Verleysen et al. 2014; Vlaamse Overheid 2012). In addition, decisions upon the inclusion of particular publishing channels, publishers, and publications are taken annually by the Authoritative Panel (AP, *Gezaghebbend Panel* in Dutch)—a panel of 18 professors in SSH disciplines affiliated to a Flemish university.

VABB-SHW stores data on publications authored by the university employees or doctoral students affiliated to an organizational unit in SSH within any of the five universities in Flanders. Until 2012, the database included also data on a small number of publications from non-university higher education institutions. In terms of publication types, VABB-SHW collects data on journal articles, articles in books, monographs, edited books, and articles in conference proceedings. The inclusion is based on a general definition of publications. For VABB-SHW, a publication must:

- 'Be publicly accessible
- Be unambiguously identifiable by an ISBN or an ISSN number
- Make a contribution to the development of new insights or to applications resulting from these insights
- Have been subjected, prior to publication, to a demonstrable peer-review process by scholars who are experts in the (sub)field to which the publication belongs. Peer review should be carried out by an editorial board, a permanent reading committee, external referees, or by a combination of these. The review should contain input from outside the author(s)'s research team and should be independent from the author(s). The author cannot organize the peer review of her or his own draft manuscript' (Verleysen et al. 2014 p. 119).

In VABB-SHW, the main exclusion criteria result from the kinds of research organizations included in the database and the general definition of publications that is used to decide upon the inclusion of data on publications. VABB-SHW does not include publications from non-universities. However, in Flanders, most SSH research is conducted within universities. Considering the criteria specified in the definition of publications, the database does not include publications by authors affiliated to organizational units in research fields other than SSH, or publications by Bachelor or Master students.

Aside from the data on publications that are recognized as peer reviewed by AP, there is also a broader data set, not publicly accessible, containing all publications within the five publication types that have been reported by the five universities in Flanders.

4.4.2. *CROSBI in Croatia*

CROSBI was created primarily for the purpose of reporting to the research funders and later on for research funding allocation purposes and research evaluation at institutional, project, or individual researcher level (see also Stojanovski 1999).

In CROSBI, there are two main inclusion criteria: CROSBI stores data on publications authored by employees or students affiliated to an organizational unit registered in the Register of research entities of the Ministry of Science and Education (universities, polytechnics, colleges, research institutes, etc.) or researchers registered in the Register of researchers of the Ministry of Science and Education (affiliated at HE, research organization or organization from business sector). Apart from this, CROSBI is intentionally designed to be as inclusive as possible.

Data are reported to CROSBI using a detailed classification of research outputs. In addition, it is possible to include also publications and research outputs that fall beyond the classification approach currently employed. In certain publication types, there are also formal criteria for inclusion; for example, ISBN is mandatory for books (but not for textbooks), and ISSN is mandatory for research articles published in the journals indexed by WoS and other online bibliographic databases or citation indices. When it comes to research evaluation or funding allocation, only the selected, scholarly types of publications are reported to the Ministry of Science and Education, the Croatian Agency of Science and Education, and other institutions doing evaluation. For these reports, specific inclusion criteria apply; however, these criteria do not influence the actual content of CROSBI.

In CROSBI, the main exclusion criteria result from the requirement for organizational units or authors to be registered by the Ministry of Science and Education. Due to this, researchers who are not employees (or students) in legal entities with a registered research activity in Croatia or researchers who are not registered by the Ministry of Science and Education do not report data on their research output to CROSBI.

4.4.3. *RIV in the Czech Republic*

RIV is a general bibliographic database (a module in the national research information system IS VaVaI), the main purpose of which is collecting, processing, and providing information about all research activities in the Czech Republic. A subset of RIV is linked to the national research evaluation system.

For RIV, inclusion criteria are decided by the government of the Czech Republic. In this database, a general definition of publications is not used. Any publication can be reported using a detailed classification of research output types. However, specific additional criteria apply due to the usage of RIV for research evaluation purposes. Only 'research organisations' as defined in national legislation (The Act on the Support of Research and Development 2002) can participate in the research evaluation system and apply for state research funding. In general terms, research organizations here mean organizations of any kind, in both public and private sector, that are recognized as pursuing research. However, also those organizations that are not officially recognized as 'research organisations' can report their research output into RIV on a condition that they participate in a publicly funded research project or have an agreement with one of the national bodies distributing research funding (e.g. The Ministry of Education, Youth and Sports).

Consequently, data on any research output can be reported to RIV assuming that a record is assigned to the correct research output type; inaccurate assignment may sometimes lead to records being deleted from the database. Such cases, however, are rare since the accuracy of research output types, typically, is checked within the organizations reporting the data. Other possible exclusion criteria may be applied at the level of funders. This applies (in most cases) to those research organizations that participate in the national research evaluation system. For example, a funder may decide about adding or deleting certain data within the database RIV: if there is a number of publications explicitly listed in a final report as outputs of a research project, the funder can decide about deleting all other publications not listed in the final report and reported to RIV in relation to that project.

4.4.4. *BFI in Denmark*
BFI was created for the purpose of research funding allocation. For BFI, inclusion criteria are decided by the Academic Committee (six recognized researchers representing every main research area) and the Steering Committee (three university rectors and the Deputy Director General of the Danish Agency for Science and Higher Education as Chairman). BFI stores data on 15 types of publications by authors affiliated to the eight Danish universities or university hospitals. The inclusion is based on a general definition of publications. For BFI, a scientific publication must:

- 'Present new knowledge,
- Be the product of research activity that complies with academic quality within the field and contributes to development of the research field,
- Be reviewed by at least one peer who evaluates the quality of the publication and the scientific contribution and who meets BFI requirements for peer reviewers' (Ministry of Higher Education and Science 2017).

Although inclusion of data on research output in BFI is based on a general definition and guidelines for its implementation (Ministry of Higher Education and Science 2017), the database that underpins BFI collects all data from the relevant research organizations. This is a consequence of the technical solution for BFI: first, all data on research output from institutional research information systems are collected. Then, the whole data set is processed using an algorithm that applies certain criteria on a step-by-step basis. In principle, all data reported on the local level are in the database, yet those records that do not meet requirements lack some metadata categories. For example, mapping of research output type classification that is used in the reporting organizations takes place after the peer review status is checked. If a record has not passed a certain step (e.g. identified as not peer reviewed), then the publication type is not matched.

The main exclusion criterion in BFI concerns research organizations. Research output by authors that are not affiliated to universities is excluded. However, after relatively recent reforms, the majority of research institutes were either merged with universities or linked by affiliation of authors, (i.e. an author affiliated to a research institute is, typically, also affiliated to a university). Consequently, the only research organizations beyond BFI are public research organizations, private organizations, and where research is not the main activity (e.g. university colleges and musea).

4.4.5. *VIRTA in Finland*
VIRTA was launched in 2016 as an advanced solution to integrate publication data and to make them available for a range of services. Since 2011, bibliographic data are collected for the purpose of monitoring research output and allocating part of the research funding at the national level. For VIRTA, the idea was to broaden the use of data beyond the research funding allocation system.

Inclusion criteria are decided by the Ministry of Education and Culture. In VIRTA, a general definition of publications specifies the following requirements:

- 'The publication must be publicly available to anyone,
- The publication channel must have an editorial board or a publisher independent of the author, who makes decisions on publications published on the channel,
- The publication has not been previously published in a format which can be reported on in the data collection system,

- The publication is based on research or expert activities carried out by the author' (Ministry of Education and Culture 2015 p. 3).

In VIRTA, 30 different research output types can be reported. In terms of research organizations, the inclusion is determined, mainly, by the Universities Act and the Polytechnics Act which requires all higher education institutions (14 universities and 23 universities of applied sciences) to supply publication information to the Ministry. In addition, five university hospital districts, each consisting of several hospitals, and six public research institutes have agreed to provide publication data to VIRTA.

Applying the definition of publications, the following publications are excluded: publications made available to only a limited audience (e.g. conference participants), self-published material, and translations, as well as new editions with only minor changes. In the same way, extracurricular, third sector, or business-related publications are excluded. Also, there can be formal criteria for inclusion of publications to certain publication types. For example, a publication without ISSN or ISBN cannot be registered into the category of peer-reviewed publications, but can be included in some of the non-refereed categories. This is a requirement of the performance-based research funding system. The requirement of ISSN (or ISBN) does not exclude publications entirely from VIRTA but affects to some extent the accuracy of publication categories: some peer-reviewed publications have to be placed in the category of non-refereed outputs because of a missing ISSN/ISBN.

4.4.6. *MTMT in Hungary*
The main purpose for MTMT is bibliometrics-based research evaluation; in addition MTMT is intended as a general source of information on research in Hungary.

For MTMT, inclusion criteria are specified in the law on the Hungarian Academy of Sciences. In general, all research output is included in MTMT using a detailed classification of output types. However, further criteria apply for subsets that are used in research evaluation and/or funding allocation. For those subsets, all output from research funded by public funds should be included in MTMT. Nevertheless, it is possible to include unaffiliated publications (e.g. researchers in retirement or without an affiliation) and publications resulting from research funded by other sources.

The main focus is on journal articles, books, book chapters, and conference proceedings. However, MTMT stores data also on research data, engineering and artistic products, and other types of research output (see Holl et al. 2014).

4.4.7. *The Database of Publications in the Social Sciences and Education in Israel*
The main purpose of this database is to systematically collect scholarly publications written by Israeli researchers in the social sciences and education and to develop a bibliographic database open to the general public. Consequently, the content of publications guides the inclusion of data in this database.

In the Database of Publications in the Social Sciences and Education, inclusion criteria are decided by The Henrietta Szold Institute. The Database of Publications in the Social Sciences and Education stores data on scholarly publications by researchers in Israel and Israeli researchers overseas on the condition that at least one author is affiliated to an academic Israeli institution (higher

education institutions, research institutes, and non-governmental organizations).

In the Database of Publications in the Social Sciences and Education, a general definition of publications is used. Publications must meet the following requirements:

- the actual publications have been identified as existing prior to entering a record in the database;
- in subjects of education, psychology, sociology, demography, social welfare, labour, communication, criminology, management and political science; and
- in Hebrew or English.

The database stores data on the following publication types: books, journal articles, reports, theses, and dissertations. Due to the focus of this database, the only exclusion criteria result from the specifications of publications outlined above.

4.4.8. *CRISTIN in Norway*
CRISTIN is a general research information system with a bibliographic database. A subset of this bibliographic database is the Norwegian Science Index (NVI). CRISTIN was set up as a multipurpose system: collected data were thought to be useful, first of all, for calculation of bibliometric indicators in the Norwegian research funding allocation system (known as the Norwegian model or the NPI). In addition, data were deemed useful also for research evaluation purposes and reporting on institutional or individual level.

In CRISTIN, inclusion criteria are decided by the staff maintaining the database. For NVI, inclusion criteria are decided by the National Board of Scholarly Publishing representing the scholarly community in Norway. CRISTIN stores data on research output by authors affiliated to research organizations participating in CRISTIN (all Norwegian higher education institutions, all research-active hospitals, and most independent research institutes).

In CRISTIN, a general definition of publications or output is not used. However, such a definition is used in NVI:

'a scholarly publication must:
1. present new insight
2. in a scholarly format that allows the research findings to be verified and/or used in new research activity
3. in a language and with a distribution that makes the publication accessible for a relevant audience of researchers
4. in a publication channel (journal, series, book publisher) which represents authors from several institutions and organizes independent peer review of manuscripts before publication' (Sivertsen 2016 p. 81).

In CRISTIN, any research output type can be reported; NVI, however, includes data on journal articles, articles in a book or conference proceeding, and monographs. In addition, publications should present new insights. The latter requirement is decided upon within the reporting organizations. The next two criteria are addressed by means of a dynamic register of approved publication channels maintained by the National Publishing Board.

Concerning exclusion criteria, there are differences in the approach to data collection in CRISTIN (the general system), NVI (the subset of scholarly publications), and the subset of NVI that is used to calculate bibliometric indicators for NPI. Consequently, exclusion criteria for the different subsets vary. Here, we describe exclusion criteria for CRISTIN and NVI resulting from the kinds of research

organizations that are included in the database. The database does not include publications by authors affiliated to organizations that receive funding from sources other than the Ministry of Education and Research, the Ministry of Health, and the Research Council of Norway. This means that the following organizations are excluded (with some exceptions): private companies, non-governmental organizations and public research organizations that receive funding from other ministries or non-governmental sources. For NVI, additional exclusion criteria result from the general definition of publications and, in addition, publications without ISSN/ISBN, publications that have not been peer reviewed and publications that do not contain new insights are excluded. All but three publication types are excluded, and also local publishing channels are excluded, local being defined as two-thirds of authors affiliated with the same organization.

4.4.9. *PBN in Poland*
PBN was set up for research evaluation and research funding allocation purposes. For PBN, criteria for inclusion of publications are linked to the Polish performance-based research funding system entitled 'Comprehensive Evaluation of Scientific Units'. Concerning the most recent design of this system, decisions on inclusion criteria were made by two advisory groups appointed by the Ministry of Science and Higher Education (Kulczycki 2017).

PBN stores data on publications by authors affiliated to scientific units. 'Scientific units' here refer to units within organizations of any sector where research is carried out: within higher education institutions, research institutes, and other research organizations in the public or private sector.

Further, inclusion criteria are defined for individual types of publications. The criteria that apply to all publication types are as follows:

- Publications have been peer reviewed prior to their publishing;
- Publications present new insights; and
- Publications have an ISSN and/or ISBN.

These inclusion criteria result in the exclusion of authors that are affiliated to administrative units or who are not affiliated to any scientific unit as specified above. Publications which do not meet the criteria for the corresponding publication type are also excluded. Similarly, given that mandate to report was introduced in 2013, it may be that before 2013 not all publications are reported.

4.4.10. *RINC in Russia*
RINC has been set up for research evaluation purposes and also, more generally, to collect information on all publications (and their citations) by Russian authors (Arefiev et al. 2012). At the beginning, the focus of RINC was on Russian scientific journals. Currently, any research output type produced by authors affiliated to Russian research organizations can be included in RINC. Data on journals and journal publications, however, is the most developed subset of RINC.

So far inclusion criteria in RINC have been decided by the staff maintaining the database. However, the current RINC Procedure Rules and Regulations (RINC 2008, unpublished internal document) were originally developed in line with national legislation that describes criteria for journals to be included in a register of Russian scholarly outlets. It states that, for example, a publication contains results of theoretical and/or experimental research or represents cultural monuments and historical documents and is meant to be

disseminated to a broad audience (Vysshaja attestacionnaja komissija pri Ministerstve obrazovanija i nauki Rossijskoj Federacii 2007).

Until 2017, there were no exclusion criteria in RINC. From time to time, some journals were rejected to be included in the list of indexed periodicals, typically in cases of publishing malpractices. As from 2018, a new version of the RINC Procedure Rules and Regulations, approved by the RINC Expert Advisory Board (a body representing the academic community), will include new inclusion and exclusion criteria.

Moreover, the RINC Expert Advisory Board also authorizes the working regulations and methodological basis of a subset of RINC, the Russian Science Citation Index (RSCI) which includes over 650 of the most prestigious journals. RSCI is a proprietary citation database (a joint project of RINC and Clarivate Analytics) of Russian scholarly journals maintained through a highly selective process. Now RSCI is a part of WoS (see also Moskaleva et al. 2017).

Exclusion criteria result from data collection practices. In some cases, it may be that publications that have not been identified in the data from main data providers are missing. Such publications are reported manually by Russian higher education institutions, but the completeness of this depends on how data input is organized. Consequently, it may be that some publications, especially in SSH, are not reported to RINC.

4.4.11. CREPČ in Slovakia

CREPČ was set up for a broad range of purposes: for reporting on institutional and individual level, for research monitoring and management, and two other specific purposes, namely, for the supplementation of the Slovak national bibliography with the so-called grey literature from academic institutions and for biographical research (Ministerstvo školstva Slovenskej republiky 2008).

For CREPČ, the inclusion criteria are decided by the Government (Ministerstvo školstva, vedy, výskumu a športu 2012). CREPČ stores data on publications by authors with a full-time academic position and by internal doctoral students in public or private higher education institutions in Slovakia. In this database, publications should meet the following criteria:

- the actual publications have been identified as existing prior to entering a record in the database; and
- publications are publicly accessible (printed or online access).

In CREPČ, any publication can be reported using a detailed classification of publication types that provides definitions and specifies inclusion criteria for each type. When it comes to research evaluation or funding allocation, only the selected, scholarly types of publications are taken into account. Principal aspects of their definition are similar to those used in other countries, namely, peer review before publishing, new insights into the topic, ISBN or ISSN.

Exclusion criteria for CREPČ result from the kinds of research organizations and the position of authors included in the database. CREPČ neither includes publications from institutes within the Academy of Sciences nor from other public research organizations. Similarly, the database does not store data on publications by external doctoral as well as other students or by authors with an administrative and/or technical position or a part-time academic position.

4.4.12. COBISS in Slovenia

COBISS is a Slovenian national shared bibliographic system established in 1990s (Seljak and Seljak 2002). It contains bibliographic metadata of practically all Slovenian outputs (all Slovenian production published in Slovenian or other languages either at home or abroad). For COBISS, there is no single responsibility for the inclusion or exclusion criteria; any publication can be reported using a detailed classification of publication types that provides definitions and specifies inclusion criteria for each type. However, similarly to CROSBI (Croatia), CRISTIN (Norway) and other databases, further inclusion criteria apply depending on the use of data. These inclusion criteria do not directly alter the content of the database; it concerns only the delineation of a subset of certain publication types. For example, a subset of COBISS is linked to SICRIS, the Current Research Information System in Slovenia.

4.4.13. SwePub in Sweden

SwePub was set up primarily to provide access to research carried out within Swedish higher education institutions and other research organizations (Kungliga biblioteket 2015). The National Library of Sweden, in coordination with the Association of Swedish Higher Education, maintains SwePub and is responsible for setting inclusion criteria. Data in this database are collected by retrieving data from institutional databases within higher education institutions and some public research institutes each of which have their own inclusion criteria. All the SwePub data should comply with the SwePub metadata format and follow national practices (e.g. in classification of research outputs and academic disciplines). In principle, the above design should not lead to exclusion of any output. However, at this point, not all research organizations beyond the higher education sector in Sweden make their data available to SwePub.

5. Discussion and conclusion

5.1. Identifying national databases

As shown in the section on findings from the first stage of the study, there are (at least) 21 national bibliographic databases in Europe and Israel collecting data on publications within SSH. However, we do wish to highlight the ambiguity with the term 'national database' that surfaced during this study. Some may consider a database as national database only if it is aimed to be comprehensive. Others may see national databases as such if they are maintained by national governmental bodies regardless of the scope. As noted earlier, we adopted a rather broad definition of a national database and relied on the knowledge of the study participants. This approach, on the one hand, helped to acquire an overview that spans a considerable number of national contexts. On the other hand, this may have led to some inconsistencies in terms of the kinds of databases that are included in (or excluded from) this overview. This latter aspect is especially crucial, given that, using the findings of this study, we have to conclude that there are no national databases in France, UK, or Spain. As noted earlier, from other sources we know that in these countries databases with a broad scope do exist and using our definition they may well be seen as national databases. Here, we have stayed close to the data we collected. Similarly, we would like to highlight that the very existence (or absence) of a national database is a theme that can be explored further in its own right. Occasionally, colleagues from some countries (e.g. Switzerland) report strong opposition to implementing a national database. In other countries (as this study shows), a national database has been maintained for several decades. This raises a question: what explains the support or resistance to bibliographic data collection initiatives? Such a question as well as the ambiguity with the very term

'national database' points to a need to continue studies of national databases. To that end, we believe, this overview serves as an informative point of departure.

5.2. Similarities and differences

Concerning database designs and principles guiding research output data collection, it is noticeable that some databases are rather restricted in scope (e.g. VABB-SHW in Flanders, NVI in Norway) while in others, data on any research output can be reported (e.g. CROSBI in Croatia, RIV in the Czech Republic, CREPČ in Slovakia, COBISS in Slovenia).

A common feature shared across the 13 databases is the inclusion of data on research output authored by academic staff affiliated to universities (though with some more detailed variations, e.g. CREPČ in Slovakia). Considerable differences, however, exist in the kinds of other research organizations that are represented in the studied databases. In some databases, the focus is on universities (e.g. VABB-SHW in Flanders, Belgium), while in others, any researcher, regardless of affiliation, can report her output to a national database (e.g. COBISS in Slovenia). Such differences as well as findings on principles guiding data collection processes indicate that the design of the databases as well as the organization of the data collection is closely linked with country-specific practices. For example, the range of research organizations included in RIV (the Czech Republic) is greater than in CRISTIN (Norway) and in VABB-SHW (Flanders, Belgium). For the Czech Republic, it is known that, historically, a prominent role in the national science system was played by research institutes within the Academy of Sciences (Arnold 2011); similarly, in Norway a significant share of SSH is carried out in public research institutes (Solberg 2016). In contrast, in Flanders (Belgium), research activities in institutes are minor compared to universities (Geerts et al. 2016). Thus, the differences in the range of research organizations included in national bibliographic databases, on the one hand, help to understand the content of the databases. On the other hand, these differences highlight that without an in-depth knowledge of the detailed context of databases, it is challenging to draw conclusions concerning the comprehensiveness of databases.

Further, often a subset of a database is either linked to national research evaluation or research funding allocation systems. First, these subsets typically have additional criteria such as the requirement for publications to be peer reviewed or stricter rules concerning bibliographic data (e.g. ISSN is required for journal publications). Second, research output beyond these subsets tends to be reported to a lesser extent. Hence, even though the 13 databases we explored here seem to be comprehensive bibliographic databases, some variation in comprehensiveness may be present for output types that are not relevant for research evaluation. Consequently, even though the 13 databases include at least one procedure ensuring comprehensiveness, it is not known which procedures lead to the most comprehensive results and what variations exist across the different databases.

5.3. Implications for bibliometrics-supported research evaluation of SSH using data from national bibliographic databases

The acquired insights into the national bibliographic databases reaffirm their value in bibliometrics-supported evaluation of research in SSH. The range of data of research output that are collected enables exploration of SSH that may lead to insights quite different from those we have had so far from citation databases such as WoS

and Scopus. The challenge, however, is the observed variation across the database setups. Hence, before considering the use of data from multiple national databases for research evaluation purposes, we suggest to pursue explorative analyses aimed to identify the extent to which specific features in database design influence bibliometric indicators and implications thereof for research evaluation. Similarly, one has to take into account the types of research output that are used for research funding allocation or research evaluation purposes. These subsets of data seem to be reported more systematically and, hence, are likely to be more comprehensive. This, however, is a non-systematic observation that could be explored empirically.

5.4. Suggestions for development of national bibliographic databases for SSH

In this study, we identified several features that may be taken into account when developing existing databases or designing new ones. First of all, it is informative if setups of databases are documented (preferably in English). This seems to be a straightforward requirement; however, documentation, if any, often turned out to be written for internal use and/or in a national language. In responses to answers on, e.g. inclusion criteria, often references to specific research organizations or registers are used without awareness that in other countries such institutions may not exist or, even more challenging, a different kind of organization may be referred to using the same name (e.g. research institutes). This flexibility of terminology is tied to country-specific social and historical trajectories, yet, for the purposes of comparative studies, it would be useful if the databases were documented by linking country-specific terms to some international framework. In this study, when describing the range of research organizations, we adapted the terminology from the OECD Frascati Manual (OECD 2015). The use of this standard may have limitations, but, for the purposes of this overview, this standard lends itself as a common ground from which to start a conversation on country-specific characteristics (and also mismatch with the terminology proposed).

Finally, we noted that some databases are broader and others more restrictive in terms of the types of research output that can be included in a database. If one would aim for a broad database, one may first design a detailed classification of research output types taking into account practices within the diverse SSH disciplines. Second, one can introduce the category 'Other'. This is relatively easy to do, and it can introduce considerable flexibility in databases. Typically, however, a broader range of research output types limits the comprehensiveness of databases. Data on research output types that do not play a role in country-specific rewarding or accountability structures tend to be less comprehensive. This then raises a question whether it is worth creating an elaborate classification of research output types if part of the data will not be reported, thus decreasing the validity of bibliometric indicators based on such data. The answer to such a question is beyond the scope of this study, yet we hope that the insights we have provided here into the national bibliographic databases for SSH will lead to, first of all, more valid and accurate bibliometric analyses of SSH, secondly, reflections and discussions of how national bibliographic databases are designed to appropriately address the needs and specifics of SSH in general and within a particular country, and thirdly, informed discussions on integration of data drawn from different national databases.

On the latter, we wish to highlight that ENRESSH envisions a European database created by integrating existing databases and information systems in Europe. Recently, ENRESSH has carried out a pilot project integrating institutional publication data from Finland, Flanders (Belgium), Norway and Spain (Puuska et al. 2018). The overview presented here in combination with insights generated in the pilot project serves as a source of information on possibilities and challenges for a European database as well as other data integration initiatives more broadly.

Supplementary data

Supplementary data is available at *Research Evaluation Journal* online.

Acknowledgements

Funding

References

Arefiev, P. G., Eremenko, G. O., and Gluhov, V. A. (2012) 'Rossijskij Indeks Nauchnogo Citirovanija—Instrument Dlja Analiza Nauki', *Bibliosfera*, 5: 66–71.

Arnold, E. (2011) International Audit of Research, Development & Innovation in the Czech Republic. Final Report. Synthesis Report. p. 57. Brighton: Technopolis Group. <https://rio.jrc.ec.europa.eu/en/file/8082/download? token=l5tMSCU7> accessed 31 Oct 2016.

Biesenbender, S., and Hornbostel, S. (2016) 'The Research Core Dataset for the German Science System: Developing Standards for an Integrated Management of Research Information', *Scientometrics*, 108/1: 401–12.

COST Association (2015) Memorandum of Understanding for the implementation of the COST Action 'European Network for Research Evaluation in the Social Sciences and the Humanities' (ENRESSH) CA15137. p. 19. Brussels. <https://e-services.cost.eu/files/domain_files/CA/Action_CA15137/mou/CA15137-e.pdf> accessed 15 May 2016.

DORA (2012) San Francisco Declaration on Research Assessment. <http://www.ascb.org/wp-content/uploads/2017/07/sfdora.pdf> accessed 16 Jan 2018.

Geerts, N. et al. (2016) STI in Flanders: Science, Technology & Innovation Policy & Key Figures—2015. No. D/2016/3241/060. p. 98. Brussels: Department of Economy, Science and Innovation. Flemish Government. <http://ebl.vlaanderen.be/publications/documents/84052> accessed 4 Aug 2016.

Hicks, D. (2004) 'The Four Literatures of Social Science' in Moed H. F., Glänzel W., and Schmoch U. (eds) *Handbook of Quantitative Science and Technology Research*, pp. 473–96. Dordrecht: Kluwer Academic Publishers.

Hicks, D. et al. (2015) 'The Leiden Manifesto for Research Metrics', *Nature*, 520: 429–31.

Holl, A. et al. (2014) 'MTMT: The Hungarian Scientific Bibliography', *Uses of Open Data Within Government for Innovation and Efficiency*. Samos, Greece.

Kulczycki, E. (2017) 'Assessing Publications through a Bibliometric Indicator: The Case of Comprehensive Evaluation of Scientific Units in Poland', *Research Evaluation*, 26/1: 41–52.

Kulczycki, E. et al. (2018) 'Publication Patterns in the Social Sciences and Humanities: Evidence from Eight European Countries', *Scientometrics*. DOI: 10.1007/s11192-018-2711-0

Kungliga biblioteket (2015) SwePub—en nationell tjänst för analys och bibliometri. No. 1.4.1-2015-444. <http://librisbloggen.kb.se/wp-content/uploads/2015/05/SwePub-en-nationell-tj%C3%A4nst-f%C3%B6r-analys-och-bibliometri.pdf> accessed 24 March 2017.

Mahieu, B. et al. (2014) *Measuring Scientific Performance for Improved Policy Making. Summary of a Study*. p. 16. Brussels: European Parliament; Directorate-General for Parliamentary Research Services. <http://www.europarl.europa.eu/RegData/etudes/etudes/join/2014/527383/IPOL-JOIN_ET(2014)527383_EN.pdf> accessed 4 Jun 2017.

Martin, B. et al. (2010) Towards a Bibliometric Database for the Social Sciences and Humanities—A European Scoping Project. p. 55. <https://globalhighered.files.wordpress.com/2010/07/esf_report_final_100309.pdf> accessed 3 July 2016.

Michels, C., and Schmoch, U. (2012) 'The Growth of Science and Database Coverage', *Scientometrics*, 93/3: 831–46.

Ministerstvo školstva Slovenskej republiky (2008) Smernica č. 13/2008-R zo 16. októbra 2008 o bibliografickej registrácii a kategorizácii publikačnej činnosti, umeleckej činnosti a ohlasov. CD-2008-19745/42317-1: 11. <http://cms.CREPČ.sk/Data/Sites/1/pdfsubory/smernica_2008.pdf> accessed 23 Nov 2017.

Ministerstvo školstva, vedy, výskumu a športu (2012) Vyhláška Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky z 18. decembra 2012 o centrálnom registri evidencie publikačnej činnosti a centrálnom registri evidencie umeleckej činnosti, 456/2012. <http://cms.crepc.sk/Data/Sites/1/pdf/vyhlaska-crepc, creuc.pdf> accessed 17 Jan 2018.

Ministry of Education and Culture (2015) *Publication Data Collection Instructions for Researchers*. <https://confluence.csc.fi/display/suorat/Julkaisutiedonkeruun+tutkijaohjeistukset? preview=/62049905/62049919/Publication%20data%20collection%20instructions%20for%20researchers_23-4-2015_en.pdf> accessed 2 Feb 2017.

Ministry of Higher Education and Science (2017) Guidelines for *Registering Research* for the Danish Bibliometric Research Indicator. <https://ufm.dk/forskning-og-innovation/statistik-og-analyser/den-bibliometriske-forskningsindikator/BFIs-regler/guidelines_for_registering_research_for_the_danish_bibliometric_research_indicator.pdf> accessed 11 Nov 2017.

Moed, H. F. et al. (2009) Options for a Comprehensive Database of Research Outputs in Social Sciences and Humanities. p. 66.

Moskaleva, O. et al. (2017) 'Russian Index of Science Citation: Overview and Review' in 16th International Conference on Scientometrics & Informetrics. Conference proceedings. pp. 1076–86. Wuhan: Wuhan University.

OECD (2015) *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, the Measurement of Scientific, Technological and Innovation Activities*. Paris: OECD Publishing.

Puuska, H.-M. et al. (2018) Proof of *Concept* of a European *Database* for *Social Sciences* and *Humanities Publications*: Description of the VIRTA-ENRESSH *Pilot*. Helsinki: CSC & ENRESSH. <https://doi.org/10.6084/m9.figshare.5993506> accessed 2 May 2018.

RINC (2008) *Reglament vkljuchenija nauchnyh zhurnalov v Rossijskij indeks nauchnogo citirovanija (RINC)*. Unpublished internal document.

Seljak, M., and Seljak, T. (2002) 'The Development of the COBISS System and Services in Slovenia', *Program*, 36/2: 89–98.

Sīle, L. et al. (2017) European Databases and Repositories for Social Sciences and Humanities Research Output. p. 25. Antwerp: ECOOM & ENRESSH. <https://doi.org/10.6084/m9.figshare.5172322.v2> accessed 9 July 2017.

Sivertsen, G. (2016) 'Publication-Based Funding: The Norwegian Model', in Ochsner M., Hug S. E., and Daniel H.-D. (eds) *Research Assessment in the Humanities*, pp. 79–90. Cham: Springer International Publishing.

Solberg, E. (2016) RIO country report 2015: Norway. No. EUR 27858 EN. p. 58. Norway. <http://publications.jrc.ec.europa.eu/repository/bitstream/JRC101206/norway_rio_cr2015.pdf> accessed 26 Sept 2017.

Stojanovski, J. (1999) 'Bibliography in the Network Environment: Croatian Scientific Bibliography (CROSBI)' in V. Hljuz-Dobrić and D. Kalpić (eds) Proceedings of the 21st International Conference on Information Technology Interfaces, pp. 359–64. Zagreb.

The Act on the Support of Research and Development (2002) 130/2002 . <http://www.vyzkum.cz/FrontClanek.aspx? idsekce=15607&ad=1&attid=706285> accessed 22 Nov 2017.

Verleysen, F. T., Ghesquière, P., and Engels, T. C. E. (2014) 'The Objectives, Design and Selection Process of the Flemish Academic Bibliographic Database for the Social Sciences and Humanities (VABB-SHW)'. In: Blockmans, W., Engwall, L., & Weaire, D. (Eds). *Bibliometrics: Use and Abuse in the Review of Research Performance*, pp. 115–25. Portland Press: London.

Vlaamse Overheid (2012) Besluit van de Vlaamse Regering betreffende de financiering van de Bijzondere Onderzoeksfondsen aan de universiteiten in de Vlaamse Gemeenschap. <https://data-onderwijs.vlaanderen.be/edulex/document.aspx? docid=14492> accessed 12 Nov 2017.

Vysshaja attestacionnaja komissija pri Ministerstve obrazovanija i nauki Rossijskoj Federacii (2007) *Porjadok registracii jelektronnyh nauchnyh izdanij, publikacii v kotoryh priravnivajutsja k opublikovannym rabotam, otrazhajushhim osnovnye nauchnye rezul'taty dissertacii.* <http://vak1.ed.gov.ru/ru/dissovet/dissovet_stat/index.php? id54=135&from54=4> accessed 17 Oct 2017.