Check for updates

The MIT Press

RESEARCH ARTICLE

Tracing the context in disciplinary classifications: A bibliometric pairwise comparison of five classifications of journals in the social sciences and humanities

Linda Sīle[1] iD, Raf Guns[1] iD, Frédéric Vandermoere[2] iD, Gunnar Sivertsen[3] iD, and Tim C. E. Engels[1] iD

[1]Centre for R&D Monitoring, Faculty of Social Sciences, University of Antwerp, Campus Middelheim—Middelheimlaan 1, 2020 Antwerp, Belgium
[2]Department of Sociology, Faculty of Social Sciences, University of Antwerp, City Campus—Sint-Jacobstraat 2-4, 2000 Antwerp, Belgium
[3]Nordic Institute for Studies in Innovation, Research and Education, Oslo, 0608 Toyen, Norway

ABSTRACT

Despite the centrality of disciplinary classifications in bibliometric analyses, it is not well known how the choice of disciplinary classification influences bibliometric representations of research in the social sciences and humanities (SSH). This is especially crucial when using data from national databases. Therefore, we examine the differences in the disciplinary profile of an article along with the absolute and relative number of articles across disciplines using five disciplinary classifications for journals. We use data on journal articles (2006–2015) from the national bibliographic databases VABB-SHW in Flanders (Belgium) and Cristin in Norway. Our study is based on pairwise comparisons of the local classifications used in these databases, the Web of Science subject categories, the Science-Metrix, and the ERIH PLUS journal classifications. For comparability, all classifications are mapped to the OECD Fields of Research and Development classification. The findings show that the choice of disciplinary classification can lead to over- or underestimation of the absolute number of publications per discipline. In contrast, if the focus is on the relative numbers, the choice of classification has practically no influence. These findings facilitate an informed choice of a disciplinary classification for journals in SSH when using data from national databases.

1. "TO COMMUNICATE INFORMATION IN THE AGGREGATE, WE MUST FIRST CLASSIFY"

The use of classifications is inevitable when communicating information aggregated in some way, as highlighted by the above quote from the seminal work on classifications "Sorting things out" (Bowker & Star, 2000) that sets the point of departure for the study presented here. Bibliometrics are not an exception. Disciplinary classifications permeate bibliometric practice in multiple ways: They are, for instance, used to delineate document sets to be analyzed, to acquire insights into the research profile of an institution, country, or other entity, and to normalize bibliometric indicators. Even though hardly anyone would doubt the centrality of disciplinary classifications in bibliometrics, the term *academic discipline* carries a great deal of

ambiguity and the operationalization of it is not straightforward (Hammarfelt, 2018; Sugimoto & Weingart, 2015). Here we explore how the choice of disciplinary classification of journals influences bibliometric analyses of research in the social sciences and humanities (SSH)[1].

Over time, multiple methods have been developed to identify academic disciplines to which publication sets belong (for an overview see Gläser, Glänzel, & Scharnhorst, 2017). First, one can follow librarianship practice and rely on content-based classifications such as the one employed in the Library of Congress (Bensman, 2007) or disciplinary databases (Eykens, Guns, & Engels, 2019). Using these classifications, the assumption is that a manual inquiry of the content of a publication (e.g., title, abstract, full-text) is the best approach to identify relevant disciplinary categories. Second, one can employ a citations-based classification approach (e.g., based on direct citations or bibliographic coupling) to identify clusters of publications of distinct academic disciplines or specialties (e.g., Carley, Porter et al., 2017; Klavans & Boyack, 2007; Leydesdorff, Bornmann, & Zhou, 2016). Third, one can use text-based algorithmic approaches (Eykens et al., 2019) or a hybrid text and citation-based approach (Janssens, Zhang et al., 2009). Finally, it is possible to use a more basic and pragmatic approach and rely on journal classifications. There are multiple journal classifications to choose from. One can use classifications from the well-known international multidisciplinary databases, such as Scopus and the citation indices in Web of Science (WoS; e.g., Leydesdorff & Rafols, 2009; Wang & Waltman, 2016) or those in use in national settings (e.g., the VABB cognitive classification from Flanders, the Dutch-speaking part of Belgium, described in Guns, Sīle et al., 2018). In each of these cases the implication is that all articles in a journal classified as belonging to a discipline X are treated as belonging to the discipline X. The journal-based approach is an efficient way to acquire approximate sense of disciplinary belonging for a set of journal articles. In addition, a journal-based approach is not methodologically demanding and therefore accessible for both bibliometricians and users of bibliometric knowledge. At the same time, journal-based methods tend to be less accurate when compared to manual or automated methods applied at the article level (Boyack & Klavans, 2020; Klavans & Boyack, 2017; Shu, Julien et al., 2019).

In our study the focus is on one of the methods—the use of disciplinary classifications of journals to infer the discipline of an article. The rationale behind this study is tied to a practical challenge encountered when pursuing bibliometrics on the basis of data from national databases for research output. In recent years, a considerable number of studies have employed national databases (e.g., Engels, Istenič Starčič et al., 2018; Kulczycki, Engels et al., 2018; Kulczycki, Guns et al., 2020; Pölönen, Engels, & Guns, 2019). This, to a great extent, results from the work within the European Network of Research Evaluation in the Social Sciences and Humanities (ENRESSH, www.enressh.eu) and the consequent awareness of the numerous underexplored data sources operated at the national level in Europe (Sīle, Pölönen et al., 2018).

National databases offer comprehensive coverage of publications; this is a crucial asset when studying research domains with limited coverage in international sources (e.g., SSH). Studies show that the coverage of SSH publications in WoS varies from around 10% to 65% (Kulczycki et al., 2018; Petr, Engels et al., 2020). In the recent decade a number of new data sources have emerged (e.g., Dimensions, Microsoft Academic, Crossref, and (earlier) Google Scholar). However, studies investigating their coverage indicate that coverage is still low for SSH, especially for publications in languages other than English (Harzing, 2019; Hug & Brändle, 2017). Thus, at the moment national databases remain the data sources that offer the most comprehensive coverage for research output in SSH.

---

[1] This article is a revised and extended version of a conference paper presented at ISSI 2019 (Sīle, Guns et al., 2019).

In addition, national databases are typically the sources that are used to evaluate and monitor research at the national level. Due to the fact that these sources have been developed for national purposes and with specific infrastructures in place, these sources are more suitable when the focus of bibliometric analysis is on research in a specific country.

The drawback of national databases, however, is that their local nature that becomes an obstacle in comparative settings; national databases often employ disciplinary classifications that have been developed and implemented in settings confined to a national context (Guns et al., 2018; Kulczycki et al., 2018; Ossenblok, Engels, & Sivertsen, 2012). We refer to such classifications as *local classifications*. Even though the national databases are more suitable for bibliometrics for SSH due to their coverage, the local classifications employed in the databases pose a challenge. We anticipate that such classifications are not (entirely) applicable to other contexts; if one set of publications is classified using a classification A, it may not be comparable to a set that is classified using a classification B. This is the main challenge we explore here.

The anticipation of this challenge stems from findings in the literature on classifications in Science, Technology, and Society studies (STS) and sociology (Bloor, 1982; Bowker & Star, 2000; Friese, 2010; Lampland, 2009; Penissat & Rowell, 2015; Zerubavel, 1996). From these studies we know that classifications are artefacts that carry traces of the social and cognitive contexts from which they originate. For example, the well-established Dewey Decimal Classification (DDC) contains a class that groups together psychology and philosophy. This is not surprising if we consider the history of academic disciplines: In the late nineteenth century, when DDC was developed, the two disciplines—psychology and philosophy—were placed under the same category ("Mental Faculties") because they were both concerned with "what the mind does" (Scott 1998 paraphrased in Bensman & Leydesdorff, 2009, p. 1109). Nowadays, it is unimaginable that disciplines as different as psychology and philosophy could be treated as a single category. This is an example of context being "imbricated" in a classification (Bowker & Star, 2000).

We expect such contextual features also in disciplinary classifications for SSH journals. From bibliometric analyses of SSH we know that in SSH one can observe diverse traditions in scholarly communication (Hicks, 2004; Nederhof, 2006). Aside from international scholarly literature, one can encounter, for example, communication more directed to a national audience. Also, in a more general sense, some SSH tends to be devoted to (more) locally relevant topics or, on the other hand, it can be institutionalized in a contextually specific manner (Depaepe, 2002; Kronegger, Mali et al., 2015; Small, 1999). For example, Small (1999) describes differences in the institutionalization of African-American studies in the United States. In one university, African-American studies is seen as an independent discipline closely bound to social activism. In another university, African-American studies is regarded as a multidisciplinary and purely academic research field associated with the more established disciplines such as sociology, history, and philosophy. This is an example of where researchers representing a research area with the same name do not share the same understanding of what the research area entails. A case like this indicates that also locally developed classifications (for SSH) potentially contain context-specific understandings of what a discipline is and how one can identify it. This consequently has implications for comparability in bibliometrics when using data that rely on different classifications: Do the results from bibliometric analyses remain comparable when using different classifications?

Here we explore this question in relation to the classification of articles on the basis of disciplinary classifications of journals. The main reason for this focus is the accessibility of this method as well as the gap in literature that has emerged due to the increasing number of bibliometric studies using data from national databases. In this context, journal-based

classification of articles is the most accessible method, which likely will remain in use also in the future—hence the need to better understand the relation between different journal classifications and bibliometric representations, despite the lower level of accuracy of this approach.

Two central objectives are pursued here. First, we seek to understand how the choice of a disciplinary classification of journals influences bibliometric analysis of articles in SSH. Second, our aim is to arrive at a number of empirically grounded practical recommendations concerning the choice of a classification of journals when studying SSH. To this end, we ask the following questions:

1. How is the disciplinary profile of articles and journals altered depending on the classification?
2. How do the total and the relative number of publications by discipline differ when using different classifications?
3. Are there differences in 1 and 2 across (a) national databases and (b) disciplines?

We seek answers to these questions by means of pairwise comparisons of five different classifications:

- the VABB cognitive classification (VABB)
- the classification used in the Norwegian list of Scholarly Journals (NPU)
- the WoS subject categories
- the Science-Metrix classification for journals (SM; Archambault, Beauchesne, & Caruso, 2011)
- the classification employed within ERIH PLUS (Lavik & Sivertsen, 2017).

The sample we use is drawn from two comprehensive databases: VABB[2] in Flanders and Cristin in Norway.

This paper is structured as follows: After a brief overview of related studies we continue with the main theoretical considerations that guide our study, a description of the disciplinary classifications, as well as the methods and data that are employed in the study. After that, we discuss the findings and their theoretical and practical implications for bibliometric research.

2. RELATED STUDIES

Multiple studies over the last decades have compared disciplinary classifications, classification methods, and their role in bibliometric analyses. Rafols and Leydesdorff (2009) explored two algorithm-based and two content-based disciplinary classifications for journals—WoS and the classification scheme developed in ECOOM-Leuven (Glänzel & Schubert, 2003)—and concluded that the agreement between the four classifications is in the range of 40% to 60%. In contrast, Bensman and Leydesdorff (2009) found "basic agreement" between WoS and the Library of Congress classification when comparing the disciplinary categories assigned to journals in behavioral sciences.

Agreement, however, does not mean that the classification is accurate. Wang and Waltman (2016) explored the accuracy of WoS and the disciplinary classification in Scopus. Using citation-based measures of journal relatedness they show that both classifications tend to

---

[2] In this study we treat the regional database VABB as a national database, as its setup is comparable to national databases (see also Verleysen, Ghesquière, & Engels, 2014).

assign too many categories to a journal. Bornmann (2018) highlighted that the accuracy of the machine-learning-based classification used in Dimensions is low, although it should be noted that the study was based on a small sample of publications (262 publications authored by Bornmann). Leydesdorff and Bornmann (2016) pointed out the ambiguities in disciplinary classifications of journals. Exploring the WoS category "information science & library science" (LIS) and the assignment of journals belonging to STS (without a designated category in WoS) it became evident that the accuracy of WoS is questionable. While journals from neighboring disciplines appear to be assigned to LIS, STS journals are spread across multiple categories (see also Vanderstraeten & Vandermoere, 2015). Similarly, Leydesdorff and Milojević (2015) showed that in one WoS category multiple communities can be found, as in the case of German sociology.

All these discrepancies foreground the role of disciplinary classifications in bibliometric analyses, like calculations of normalized citation scores (Haunschild, Marx et al., 2018) or journal rankings (Bensman & Leydesdorff, 2009). However, there are situations when differences due to the choice of disciplinary classification are minor. Klavans and Boyack (2007) compared reference-based science maps using WoS and Scopus data (and the respective classifications). Their analysis showed that the acquired maps are to a great extent convergent in terms of their representation of the structure of science. On a more fine-grained level, however, the increased coverage (as in Scopus) and consequently the broader pool of references, leads to a more accurate map of science. Similarly, Leydesdorff and Rafols (2009) note that despite the known modest accuracy of WoS, maps generated using this classification appear accurate (see also Rafols & Leydesdorff, 2009). The explanation for this, they argue, is the stochastic distribution of assignments on the lower levels, which does not appear when exploring the structure of science on higher levels of aggregation.

It is not known at the moment whether such findings can be related to the use of local (e.g., national) disciplinary classifications, especially when studying SSH. Studies exploring the contextual nature of local disciplinary classifications for SSH, to the best of our knowledge, do not exist. The use of comprehensive (national, regional, or institutional) databases where local classifications are employed is still an underexplored area. While a number of studies drawing on such data sources have been pursued (e.g., Hammarfelt & de Rijcke, 2015; Kulczycki et al., 2018; Ossenblok et al., 2012), understanding of the comparability of databases developed in different national contexts is at an early phase (Sivertsen, 2019). This applies also to disciplinary classifications for journals.

Dealing with local disciplinary classifications, it is possible to use crosswalks—concordance tables—for classifications whereby categories in one classification are mapped to categories from another classification (e.g., Engels et al., 2018; Kulczycki et al., 2018). However, the use of crosswalks is not always straightforward. For example, communication studies in the OECD Fields of Research and Development (henceforth FORD)[3] classification (OECD, 2007, 2015) are seen as a field in the social sciences, but in a study by Kousha et al. (2011) it belongs to the humanities. Similarly, psychology in FORD is assumed to belong to the social sciences, but in the relatively recent SM journal classification, psychology belongs to the health sciences (Archambault et al., 2011). Another example is anthropology, which by many is regarded as a distinct discipline with a long history and hence deserves its own category in a classification. In contrast, in FORD, anthropology is subsumed within the category "Sociology." Moving beyond SSH, an example can be found for biotechnology, which in the SM classification is

---

[3] In versions of the Frascati Manual that precede 2015, FORD classification is known as the Fields of Science (FOS) classification.

one category, while in FORD there are four categories spread across different knowledge domains: "Environmental Biotechnology" (2.8), "Industrial Biotechnology" (2.9), "Medical Biotechnology" (3.4), and "Agricultural Biotechnology" (4.4). These examples imply that the choice of a disciplinary classification may change the distribution of publications across disciplines or may lead to a number of publications that is larger or smaller depending on the employed disciplinary classification. Publications from some disciplines can eventually be spread across multiple disciplines; or, in contrast, articles from multiple disciplines could be identified as belonging to a single discipline.

3. CONTEXT IN CLASSIFICATIONS

For us, classification is "a spatial, temporal, or spatio-temporal segmentation of the world" (Bowker & Star, 2000, p. 10). Classifications order the society in which they are embedded (Durkheim & Mauss, 2010/1903). They "reproduce the pattern of social inclusions and exclusions" (Bloor, 1982, p. 267) and they can bear religious, political, and moral traces from the contexts in which they originated (Bowker & Star, 2000). This explains the example of a shared subject class for philosophy and psychology.

In addition, Bowker and Star highlight how classifications structure the way we see social phenomena. As the classifications become more engrained in social practices (e.g., reporting routines, rewarding mechanisms), a seemingly neutral category acts as a tool. Classifications as tools can make visible and prioritize a certain social practice or arrangement. Bowker and Star use the example of Nursing Interventions Classification (NIC) to show how classification has helped to claim autonomy and professional expertise for the nursing profession. Before the NIC classification was developed, there was no easily accessible and sufficiently detailed way to understand the nature of the nursing profession. NIC helped to increase awareness about the breadth of activities the nursing profession entails. However, a classification can also make certain social practices less valuable or simply invisible, as in the case of the residual categories (Star & Bowker, 2007). An example here is the mentioned field of STS: In WoS there is no category for this specialty. In a bibliometric analysis based on WoS, STS is not visible unless additional manual effort is done to reconfigure this classification (Leydesdorff & Bornmann, 2016). In the SM classification, in contrast, one can find a category "Science Studies" that includes the major STS journals. Thus, the visibility of STS depends on the choice of classification.

Transferring this reasoning to the use of local disciplinary classifications for journals in SSH, we anticipate such classifications to carry traces from their context. By "traces from the context" in this study we mean changes in the quantitative representations of SSH research as a result of the choice of journal classification.

With this conceptualization of classification we distance ourselves from the essentialist view that assumes that the disciplinary profile is a characteristic intrinsic to an article or a journal. Instead, we argue that, just like any other classifications, disciplinary classifications are social achievements resulting from local and context-specific considerations and negotiations.

We treat local classifications (VABB and NPU) as baseline; in doing so we assume that scholars working in a certain context are best positioned to determine the disciplinary profile of a journal. However, we do expect that the journal-category assignment, say, in Flanders will differ from that in Norway due to context-specific ways to understand and identify distinct academic disciplines, and consequently, also to recognize how features of particular academic disciplines manifest in a scholarly journal. We will not delve into the histories of these classifications; instead we will seek systematic quantitative deviations. If these deviations are explored

further, they may be traced back to the specific local conditions under which they were created. In this study, however, we limit ourselves to the quantitative manifestations of traces from context.

We expect that the three international classifications—WoS, SM, and ERIH PLUS—will differ in their similarity with the local classifications. These classifications, although perceived as international, have been developed in a specific context as well, which may or may not be similar to those where the two local classifications studied here are embedded. This could be yet another manifestation of the contextual nature of disciplinary classifications of journals. With these considerations and anticipations in mind, we proceed with a brief description of the classifications, data, and methods used in this study.

4. DISCIPLINARY CLASSIFICATIONS FOR COMPREHENSIVE SSH DATA

In our study we employ five different classifications: the Flemish VABB classification, the Norwegian NPU classification, the SM classification, the WoS subject categories, and the ERIH PLUS classification. All five classifications are used at the level of journals. In addition, the OECD FORD classification is used to acquire comparability of the different classifications by mapping the classifications to OECD FORD (see Table 1).

4.1. VABB

The VABB classification includes only fields in SSH. VABB is a cognitive classification (Daraio & Glänzel, 2016) developed at ECOOM-Antwerp in 2017 with the purpose of enriching bibliographic data in VABB-SHW, the database for research output in SSH in Flanders, with a cognitive classification (Guns et al. 2018). This is a two-level hierarchical classification based on OECD FORD classification (17 academic disciplines in total for SSH). For journal articles, the classification is applied at the journal level. The assignment of journals to academic disciplines has been carried out, first, by identifying the most often occurring academic discipline in multiple sources (WoS, Scopus, NPU, ISSN.org, and OCLC Classify) and, second, by manually assigning journals to disciplines. Each journal was assigned to up to five academic disciplines. Further information on this classification can be found in Guns et al. (2018).

4.2. NPU

NPU is a classification used in the Norwegian Register of Scholarly Journals, which is maintained for statistical purposes and for the performance-based research funding system (PRFS,

Table 1. Overview of categories for social sciences and humanities in five disciplinary classifications

	VABB	NPU	SM	WoS	ERIH PLUS (OECD)	OECD FORD
Number of levels	Two	Two	Three	One	Two	Two
Number of SSH categories at level 1	2	2	2	–	2	2
Number of SSH categories at level 2	17	35	6	–	14	14
Number of SSH categories at level 3	–	–	43	76	–	–
Maximum number of categories per journal	5	1	1	6*	14	–

* The maximum number of categories identified for the journals included in the data sets analyzed here.

the Norwegian Publication Indicator) in Norway[4]. This classification is applied to the subset of peer-reviewed scholarly publications (journal articles, articles in books, and books) in the national Norwegian database Cristin. While publications in books are classified individually when registered by their authors, NPU works as a journal classification determining the classification of all articles in each journal. The classification is based on the structure of disciplinary committees within Universities Norway (UHR), a cooperative body for 33 accredited universities and university colleges in the country. The committees are discipline specific, as they were created for interinstitutional collaboration within each academic discipline with a bachelor's and a master's level program in Norway. The main tasks of these committees related to educational activities were extended as the Norwegian PRFS was implemented in 2006. They now also advise the National Board of Scholarly Publishing. This board is responsible for the Norwegian Publication Indicator, including the contents of the Norwegian Register for Scientific Journals, Series and Publishers, which is maintained by NSD (Norwegian Centre for Research Data). The main purpose of the register is to identify publication channels that, first, meet minimum quality criteria and, second, are considered prestigious in the corresponding scientific community (Sivertsen, 2016). Because continuous expert advice from representative bodies is needed for both purposes, each journal is classified as belonging to only one disciplinary committee. The classification of a journal may change following an agreement between the relevant disciplinary committees. Journals may also be classified as interdisciplinary within each major area of research: Humanities, Social Sciences, Health Sciences, and Natural Sciences and Engineering.

For the analysis presented here, we developed a crosswalk that allows us to identify a corresponding OECD FORD category for each NPU category. It should be noted that according to NPU, the category "Psychology" belongs to health sciences, while the category "Media and communication" belongs to the humanities. Following OECD FORD, we consider both "Psychology" and "Media and communication" as social sciences (5.1 and 5.8 respectively).

4.3. SM

The SM classification was developed to "facilitate the production of bibliometric data" (Archambault et al., 2011, p. 1). This classification is based on the classification used in the Science and Engineering Indicators by the US National Science Foundation (NSF), the WoS classification, and the Australian Research Council (ARC) Evaluation of Research Excellence (ERA) classification. In addition, OECD FORD and the European Research Council classifications were consulted to fine-tune the classification. The classification was applied to journals using a citations-based algorithm in combination with the mentioned journal classifications (WoS, US NSF, ARC ERA). This classification has three levels: six categories at the highest level, 22 at the second level and 179 at the most fine-grained level. In total, 34,000 journals and conference proceedings that have at least one record in WoS and/or Scopus were classified using this scheme. Here we use the publicly available list of classified journals ($n = 15,007$). For the analysis presented here, we use a crosswalk that allows us to map SM classification to OECD FORD. Because the first and second levels of the SM classification contain categories that do not have a corresponding category in OECD FORD (e.g., "Applied Sciences" at the first

---

[4] We have chosen the abbreviation "NPU" for the classification as it is the abbreviation of the National Board for Scholarly Publishing in Norway (det Nasjonale Publiseringsutvalget, NPU). This is the context where the classification is used, maintained, and updated (see the overview here: https://npi.nsd.no/organisering). Here we use the journal assignment to disciplinary categories from the 2017 version of Norwegian register of scholarly journals maintained by the the Norwegian Centre for Research Data, available here: https://dbh .nsd.uib.no/publiseringskanaler/AlltidFerskListe.

level or "Communication & Textual Studies" at the second level), we map the lowest level: the 179 subfields.

4.4. WoS

WoS classification consists of 252 subject categories and is used to classify journals included in indices within the WoS Core Collection. For each journal, multiple categories can be employed. Pudovkin and Garfield (2002) wrote that the application of WoS to journals is based on an approach that involves journal's title and citation patterns and other criteria. At the same time, as has been acknowledged in the literature earlier, documentation of this classification and its underlying principles is limited (Leydesdorff & Rafols, 2009; Wang & Waltman, 2016). To map WoS to OECD FORD, we use the *OECD Category to Web of Science Category Mapping 2012 (2012)*. For recently added WoS categories that were not included in this mapping, we identified the corresponding OECD FORD category manually.

4.5. ERIH PLUS

The European Reference Index for the Humanities and the Social Sciences (ERIH PLUS)[5] is a dynamic register of journals and series in SSH (*n* = 8,009; Lavik & Sivertsen, 2017). In 2014, the responsibility to maintain and further develop this register was assigned to the Norwegian Centre for Research Data (NSD). This register employs a one-level classification scheme with 30 categories that is mapped by the NSD staff to OECD FORD. The procedure for the classification of journals is as follows: First, the person who suggests a journal for inclusion in ERIH PLUS (often, the journal editor) provides information on the disciplinary scope of the journal. This information is afterwards validated and corrected, if necessary, by the NSD staff (personal communication with the NSD staff, August 17, 2020). There are no restrictions on the number of categories to which a single journal can be assigned. For further information on ERIH PLUS, see Lavik and Sivertsen (2017).

4.6. OECD FORD

The OECD FORD classification, earlier known as the Fields of Science classification (FOS), was developed as part of the Frascati Manual for R&D measurement purposes. We use here the version that was published in 2015 (OECD, 2015) and the more detailed description of each category (OECD, 2007). This is a two-level classification scheme with six higher level categories and 42 lower level categories. As noted, we mapped the five explored classifications to OECD FORD for comparison. The Supplementary Material provides the complete list of journals and their OECD FORD classification codes used in this analysis.

5. METHODS

5.1. Data

We use data from two bibliographic databases (VABB-SHW in Flanders, the Dutch-speaking part of Belgium, and Cristin in Norway) and WoS. Data were retrieved from Cristin on March 24, 2018 and from VABB-SHW on May 23, 2017. WoS data were retrieved from the in-house WoS database maintained by ECOOM-Leuven on July 23, 2018, and from the Clarivate Analytics WoS database via the web interface on November 21, 2018. The choice to use data from these two countries is merely pragmatic. In addition, we assume that comparisons in

---

[5] Downloaded on March 4, 2020 from https://dbh.nsd.uib.no/publiseringskanaler/erihplus/periodical /listApprovedAsCsv.

relation to classifications are methodologically more robust if carried out using data from more than one national database.

The analysis is conducted using 10 data sets (five per country: Reference, WoS, SM, ERIH PLUS, VABB-NPU) delineated as follows (Table 2). The data sets WoS, SM, ERIH PLUS, and VABB-NPU are subsets of the reference data set. The reference data set is not used in analyses, but the summary of the number of publications by discipline is available in Table S1 in the Supplementary Material. The reference data set is retrieved from the national databases and consists of peer-reviewed journal articles in SSH (2006–2015) by authors affiliated to universities. The WoS data set is limited to articles from the reference data set that are indexed in the three main indices in WoS (SCI-E, SSCI, AHCI). While the SM data set is limited to articles in journals that are included in the SM classification, the ERIH PLUS data set is confined to the journals included in ERIH PLUS. Finally, the VABB-NPU data set contains articles in the subset of journals that were identified in the reference data sets both from Flanders and Norway; journals that were identified in only one of the data sets were not considered.

Criteria that are used in all data sets refer to the time frame 2006–2015, institutions to which authors of articles are affiliated (the five universities in Flanders and the eight universities in Norway), publication type (journal articles), and the peer-review status (only peer-reviewed publications are taken into account). Delineation on the basis of institutions was carried out using institution identifiers as they are recorded in the VABB-SHW and Cristin databases. In recent years, several higher education institutions in Norway have been merged: Here, identifiers refer to the present institution (if A is merged with B, then in the analysis all publications from A before the merger are regarded as belonging to B).

Peer-reviewed publications were identified as follows. For Flanders, "peer-reviewed" refers to publications in journals included in the list of peer-reviewed journals by the Authoritative Panel (VABB version 7) (for details see Verleysen et al., 2014) and/or that are identified as indexed in WoS for the WoS-based part of the Flemish PRFS. For Norway, "peer-reviewed" refers to Level 1 and Level 2 in the Norwegian Register of Scholarly Journals (Sivertsen, 2016).

Table 2. Samples and criteria used to delineate data subsets

Sample and criteria used to delineate data subsets	Data set				
	Reference	WoS	SM	ERIH PLUS	VABB-NPU
Flanders					
# of publications	29617	12549	11507	11061	18509
# of journals	5589	2377	2014	1957	2276
Norway					
# of publications	26054	9451	10436	9066	10653
# of journals	5193	2089	2004	1690	2276
Only articles indexed in Web of Science (SCIE, SSCI, AHCI)	–	Yes	–	–	–
Only articles in journals included in Science-Metrix list	–	–	Yes	–	–
Only articles in journals included in ERIH PLUS	–	–	–	Yes	–
Only articles in journals that are present in the data sets both from Flanders and Norway	–	–	–	–	Yes

Publications in SSH here means publications in journals belonging to SSH. The reference point for SSH are those disciplines that are listed under classes 5 (Social Sciences, SS) and 6 (Humanities, H) in the OECD FORD classification.

To delineate the WoS data sets, we identified articles that are indexed in WoS. The identification was carried out on article level. We used data sets retrieved from the ECOOM-Leuven in-house WoS database and the identification was carried out using a string-matching approach that allows for small differences in the matched references (Sīle & Guns, 2019, 2020). To acquire WoS classification for all articles in the WoS data sets, we searched ISI numbers via the Clarivate Analytics WoS user interface online. The SM and ERIH PLUS data sets were delineated by identifying articles that are included in, respectively, the SM classification and the ERIH PLUS list of journals on the basis of ISSN (both print and online if applicable). Finally, the VABB-NPU data sets were delineated using a list of unique ISSNs that can be identified in both the Flemish and Norwegian reference data sets.
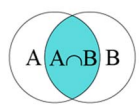
5.2. Analysis

Our analysis is based on seven pairwise comparisons (VABB and WoS, VABB and SM, VABB and ERIH PLUS, NPU and WoS, NPU and SM, NPU and ERIH PLUS, and VABB and NPU) at journal and article level. By article level we mean the analysis in terms of publication counts to which we apply the classification of journals.

Our analysis is structured in two parts. The first part of the analysis provides a general overview of the similarity of disciplinary profiles for journals and articles across the classifications. To do so, we use the Jaccard index calculated pairwise at the level of journals and articles. If d_A and d_B denote the set of disciplines for a journal or article according to, respectively, classifications A and B, the journal or article's Jaccard index associated with A and B is

the intersection of A and B

$$Jaccard_{A,B} = \frac{|d_A \cap d_B|}{|d_A \cup d_B|}.$$

$$\mathrm{Jaccard}_{A,B} = \frac{A\; A{\cap}B\; B}{A\; A{\cup}B\; B}$$

the union of A and B

The mean of the Jaccard index ($M_{Jaccard}$) is aggregated for each discipline, as well as for subtotals (SS, H) and totals (SSH) provide general insight into the similarity of classifications.

The second part of the analysis zooms into changes in absolute and relative numbers due to the choice of classification. To capture this, we use percentage error (*%error*) and arithmetic difference in share (*diff.share*). While *%error* captures relative differences in the absolute number of publications, the *diff.share* helps to identify alterations in the disciplinary structure—the distribution of publications across disciplines expressed as the share of publications by discipline. If we denote the number of publications according to classification A by V_A, *%error* is acquired as follows (taking classification B as the baseline):

$$\%error = \frac{V_A - V_B}{V_B} \times 100\%.$$

As noted in Section 3, we use the local classifications as the baseline, thus implying that we regard the local classifications to be more accurate. Hence the *%error* captures the extent to which the use of a different, other than the local, classification produces an error.

As our interest is in research performance, all analyses are carried out using fractionalized counts. Fractionalized counts are calculated as the number of local authors divided by the total number of authors. Detailed results in tabular format can be found in the Supplementary Material.

5.3. Limitations

The main limitation of this study is related to the comparative nature of the analysis. We use data from two different national databases. Even though both databases are assumed to be comprehensive databases for peer-reviewed scholarly publications (for Flanders, only for SSH), we are aware that there are some differences in database setups that might alter our results. For example, we know that in Norway, the category "journal article" also includes papers in conference proceedings that are published with an ISSN. Also, there might be discrepancies in what is considered a scholarly peer-reviewed publication, given the differences in inclusion criteria (Sīle et al., 2018). Due to the focus of this study we assume, however, that the influence of those discrepancies is of minor importance.

Another minor limitation is related to the WoS data sets, the data sets limited to the articles indexed in WoS. Our bibliographic data matching approach was developed taking VABB-SHW data as the point of reference. This means that the accuracy of the approach with respect to Cristin data might be lower due to differences in bibliographic data collection and processing practices. Consequently, the number of WoS-indexed articles for Norway might be underestimated. Given that the focus of this study is not WoS coverage, we assume that the acquired level of accuracy is acceptable.
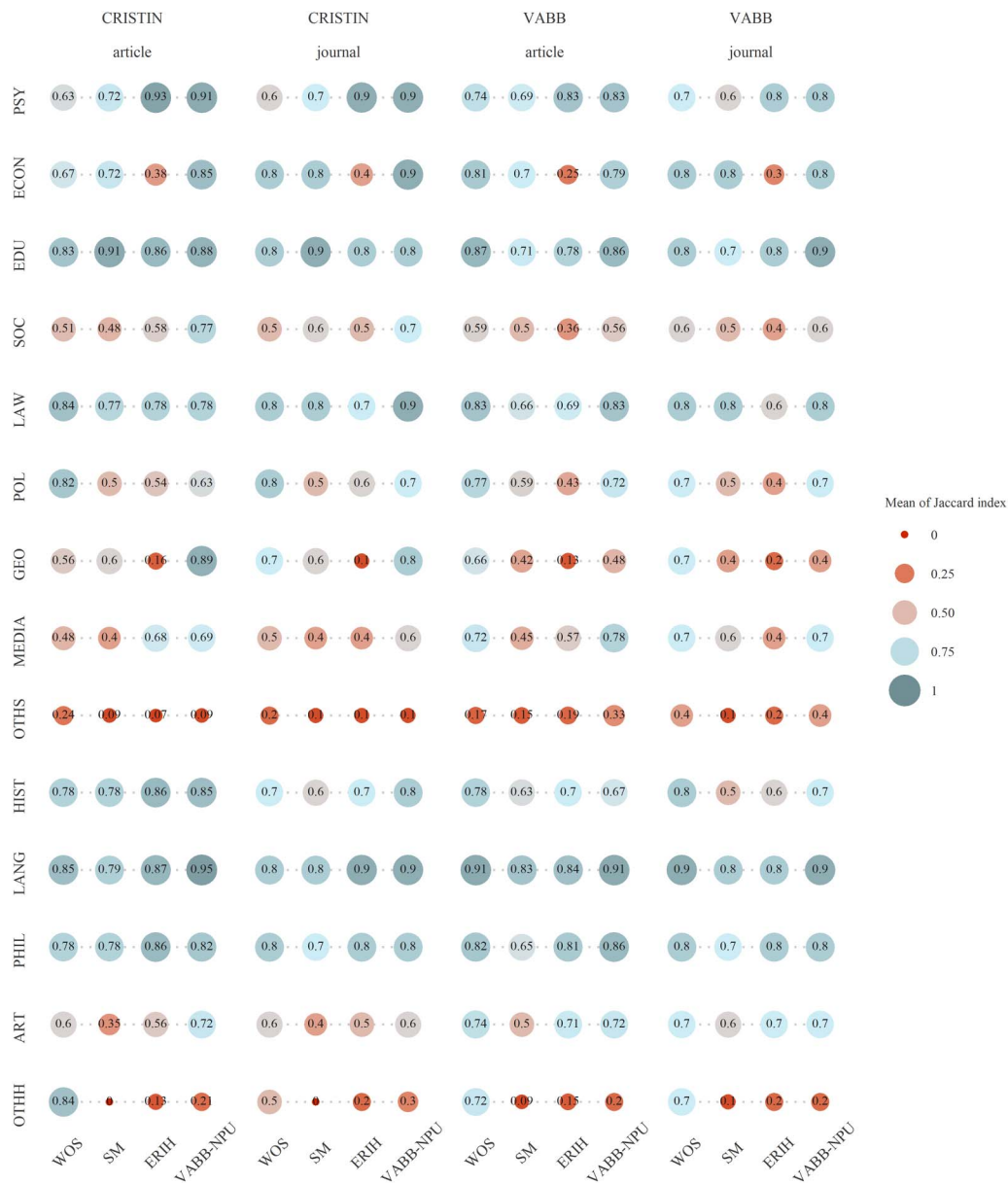
6. RESULTS

6.1. Similarity of Disciplinary Profiles Across Classifications

Figure 1 provides an overview of $M_{Jaccard}$ by discipline at the level of articles and journals (see also Tables S6 and S7 in the Supplementary Material). Although at the level of the articles there is a slightly larger similarity than at the level of the journals, the figures for both levels are comparable. Therefore, in what follows, we focus on the disciplinary profile at the level of articles.

A slight variation in $M_{Jaccard}$ can be observed across all the compared pairs of classifications (Figure 1); $M_{Jaccard}$ for SSH in total is on average 0.73 ($SD = 0.39$, $Md = 1$, range = 0.63–0.80). Across the compared classifications, the highest similarity can be identified for the pair of local classifications (NPU and VABB; $M_{Jaccard} = 0.80$). While VABB is quite similar to WoS (0.78), NPU is closer to the ERIH PLUS classification (0.77). The similarity between the two local classifications as well as between VABB and WoS might be explained by the fact that NPU and WoS were used when creating the VABB classification (Guns et al., 2018). The least similar pairs are the local classifications and SM (0.65 for VABB-SM and 0.63 for NPU-SM) as well as NPU-WoS (0.65). Comparing SSH, it is evident that in nearly all comparisons there is a slightly higher agreement for the humanities.

The variation across disciplines is noteworthy. For most disciplines, the similarity across classifications falls within the interquartile range. There are, however, two groups of disciplinary categories that tend to have either consistently higher or lower similarity than the other disciplines. High similarity across all classifications can be observed for "Economics and business," "Educational sciences," and "Law" as well as for "History and archaeology,"

Figure 1. Mean Jaccard index by discipline at the level of journals and articles. (PSY—Psychology, ECON—Economics and Business, EDU—Educational sciences, SOC—Sociology, LAW—Law, POL—Political science, GEO—Social and economic geography, MEDIA—Media and communications, OTHS—Other social sciences, HIST—History and archaeology, LANG—Languages and literature, PHIL—Philosophy, ethics, and religion, ART—Art, OTHH—Other humanities).

"Languages and literature," and "Philosophy, ethics, and religion." For these categories, the Jaccard index means are, on average, higher than 0.75 across all classifications.

Note, however, that a relatively low (<0.40) similarity can be observed for "Economics and business" for the comparisons with ERIH PLUS classification. A manual inquiry into the journals in this category shows that a number of journals that according to the other classifications belong to "Economics and business" are classified as journals in "Psychology," "Educational sciences," "History and archaeology," or other categories (e.g., *Academy of Management Review*, *Evaluation and Program Planning*, *European Journal of the History of Economic Thought*).

Among the categories that appear less similar across the different classifications, we find "Sociology" and "Social and economic geography," and the two residual categories "Other social sciences" and "Other humanities."

This analysis shows that the disciplinary profile of journal articles in Flanders and Norway can be determined with a considerable level of accuracy using the respective local classifications (the VABB for Norway and the NPU for Flanders). At the same time, this analysis also indicates that there is a risk of misrepresentation of a number of disciplines primarily in the social sciences (e.g., sociology).

6.2. Differences in the Absolute Number of Publications

In this section, we continue with the analysis of differences in the absolute number of publications captured with *%error*. Overall, the analysis in terms of the absolute number of SSH articles reaffirms findings from the analysis of the disciplinary profile of articles.

The total number of SSH publications is greater when using local disciplinary classifications rather than the WoS classification (3% for Flanders and 8% for Norway). Even greater differences can be observed for the comparisons with the SM classification (differences of 10% for Flanders and 16% for Norway). In the comparison with ERIH PLUS, however, the number of SSH articles remains the same. Comparing the two local classifications (VABB and NPU), the number of publications for Norway is identical using both classifications. For Flanders, in contrast, the number is 5% smaller when using the NPU classification.

Articles that are not identified within SSH using WoS, SM, or NPU, are most often associated with categories within the knowledge domain "Medicine and Health sciences" (Table 3, Figure 2). For both Flanders and Norway, these categories account for the majority of non-SSH articles. In addition to medical fields, non-SSH articles are associated with the knowledge domains "Civil engineering," "Earth and related environmental sciences," and "Computer and information sciences." This indicates, first, that boundaries between broad knowledge domains such as SSH, medicine, engineering, and natural sciences are somewhat blurred, as indicated in science mapping studies (Börner, Klavans et al., 2012). Second, this is an indication of a context-specific understanding of a particular discipline that leads to the assignment of journals such as *Urban Water Journal* to "Social and Economic Geography" (as in NPU) instead of "Earth and related environmental sciences" (as in WoS).

The findings for SS and H explored separately are slightly different, however. In most cases, the number of SS articles is greater when using the local classifications (1% for VABB-WoS, 13% for VABB-SM, 1% VABB-ERIH PLUS, 7% for VABB-NPU, 7% for NPU-WoS, and 19% for NPU-SM) with the exceptions for NPU-ERIH PLUS (2%) and NPU-VABB (1%), where the numbers are slightly lower when using the local classifications. For H, in most cases the number of articles is smaller when using the local classifications (4% for VABB-WoS, 12% for VABB-ERIH PLUS, 7% for NPU-WoS, and 15% for NPU-ERIH PLUS) with the exception for VABB-SM (10%) and NPU-SM (15%). Finally, a comparison of the two local classifications shows that the number of publications in SS and H is greater when using VABB (Flanders: SS 7%, H 7%; Norway: SS 1%, H 7%).

%error varies considerably by discipline similarly to what could be observed in the analysis of disciplinary profile (Figure 2, Tables S2–S5 in Supplementary Material). Figure 2 demonstrates changes in the number of articles by discipline using different classifications (see also Supplementary Material for comparisons with SM, and the two local classifications). For the comparisons with WoS, it is evident that a substantial number of publications in the category
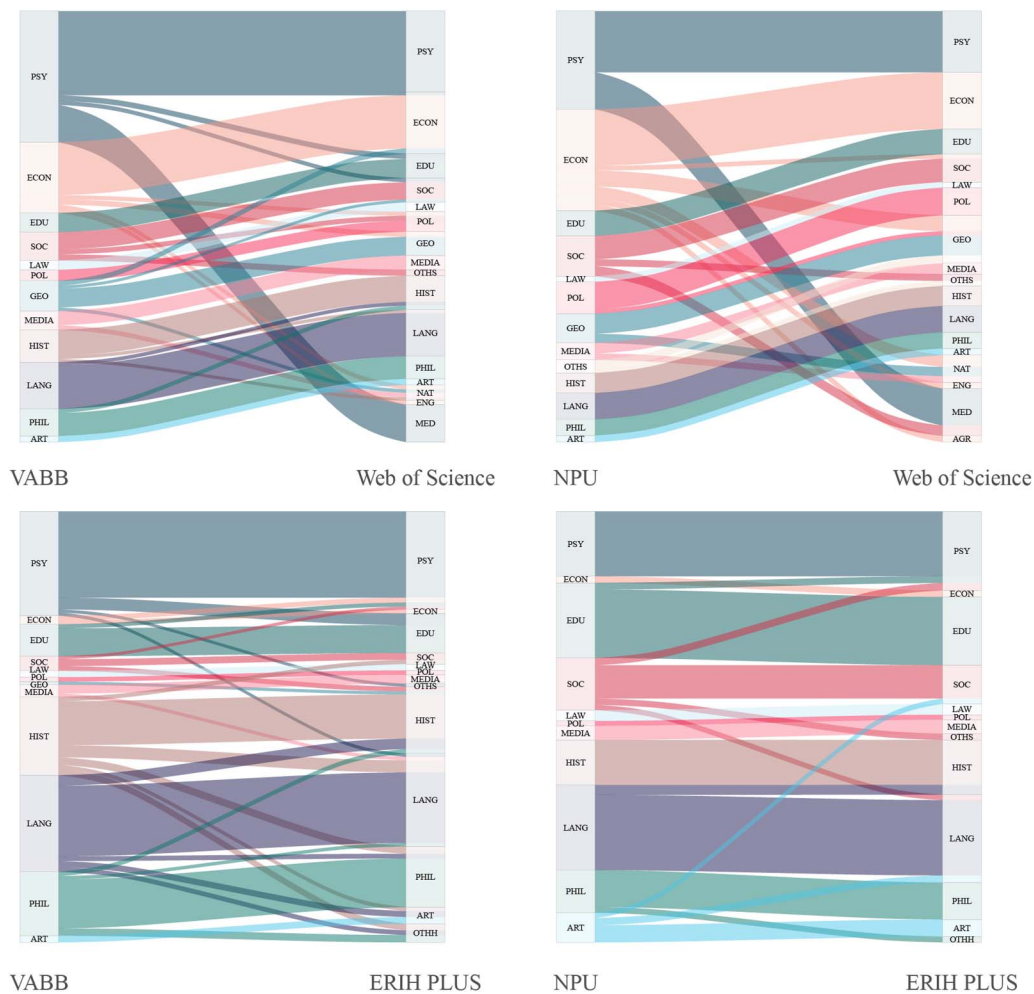
Table 3. The 10 most often occurring categories for articles not associated with social sciences and humanities*

Comparison with Web of Science subject categories						Comparison with Science-Metrix classification						Comparison of local classifications		
Flanders			Norway			Flanders			Norway			Flanders		
Category	#	%	Category	#	%	Category	#	%	Category	#	%	Category	#	%
BMED	199.9	54	HEAL	176.2	25	HEAL	268.1	30	HEAL	255.2	21	CMED	224.4	30
CMED	67.6	18	BMED	155.5	22	OTHE	238.0	26	CMED	167.3	14	HEAL	139.4	19
HEAL	37.7	10	EARTH	105.3	15	CMED	106.2	12	CIVE	155.4	13	OTHE	87.5	12
CIVE	18.3	5	CMED	93.2	13	COM	90.4	10	COM	139.5	12	COM	72.3	10
EARTH	12.3	3	COM	68.0	10	CIVE	65.8	7	ENVE	121.7	10	BMED	64.3	9
MAT	10.5	3	BIO	53.6	8	ENVE	60.8	7	EARTH	110.9	9	MATE	54.9	7
COM	6.9	2	AGR	16.8	2	MAT	39.5	4	AGR	101.4	8	MECH	34.0	5
OTHE	5.3	1	PHYS	9.8	1	EART	17.8	2	OTHE	93.2	8	BIO	33.0	4
AGR	4.0	1	MATE	6.7	1	BIO	13.7	2	BIO	47.3	4	CIVE	22.9	3
ENVE	3.8	1	ENVE	6.3	1	AGR	3.3	0	MAT	13.6	1	OTHHL	9.8	1
Total	371.9	100	Total	699.2	100	Total	908.8	100	Total	1212.6	100	Total	752.2	100

* For data sets ERIH PLUS (Flanders and Norway) and VABB-NPU (Norway), all SSH publications were classified as SSH also using the ERIH PLUS and the local classifications.

Academic discipline abbreviations. AGR: Agriculture, Forestry, and Fisheries; BIO: Biological sciences; BMED: Basic medicine; CIVE: Civil engineering; CMED: Clinical medicine; COM: Computer and information sciences; EARTH: Earth and related environmental sciences; ENVE: Environmental engineering; HEAL: Health sciences; MAT: Mathematics; MATE: Materials engineering; MECH: Mechanical engineering; OTHE: Other engineering; OTHHL: Other health sciences; PHYS: Physical sciences.

The third column (%) for each comparison refers to the share of all the articles that were not associated with social sciences and humanities.

Figure 2. Changes in the number of publications by discipline due to the choice of a classification. (for readability limited to combinations of disciplines with at least 75 fractionalized articles; PSY—Psychology, ECON—Economics and Business, EDU—Educational sciences, SOC—Sociology, LAW—Law, POL—Political science, GEO—Social and economic geography, MEDIA—Media and communications, OTHS—Other social sciences, HIST—History and archaeology, LANG—Languages and literature, PHIL—Philosophy, ethics, and religion, ART—Art, OTHH—Other humanities, NAT—Natural sciences, ENG—Engineering and technology, MED—Medical and health sciences, AGR—Agricultural sciences).

"Psychology" in the local classifications are seen as part of medical and health sciences when using WoS. Also, publications in "Economics and business" are spread across multiple categories in social and natural sciences. The comparisons with ERIH PLUS show that a relatively smaller number of publications is dispersed across multiple SSH categories. It is, however, striking that ERIH PLUS evidently carries a bias towards the humanities, presumably as a result of its historical focus on the humanities. As ERIH PLUS includes a relatively high number of journals in, for instance, "History and archaeology" and "Languages and literature," the number of publications appears much greater than when using WoS.

The *%error* is especially high for social and economic geography and the two residual categories "Other social sciences" and "Other humanities." These high differences might be an indication of a more conservative tendency in the journal-to-category assignment: Interdisciplinary research or research from new disciplines is perceived as belonging to one

of the more established disciplines. This interpretation is supported also by the very low differences for categories "Economics and business," "Educational sciences," "Law," and "Languages and literature." These low differences, however, are not consistent for the compared international classifications (WoS, SM, ERIH PLUS), thus indicating that these classifications also carry arbitrary assumptions on journal-discipline pairs that are not consistently aligned with the two local classifications.

Comparison of the two local classifications shows that at the level of individual disciplines, the number is typically greater using the VABB classification. A possible explanation for this could be the use of multiple categories per journal in VABB. However, contrary to the general trend, there are several categories ("Political science," "Media and communications," and "Other social sciences") where the number of publications is slightly greater when using the NPU classification.
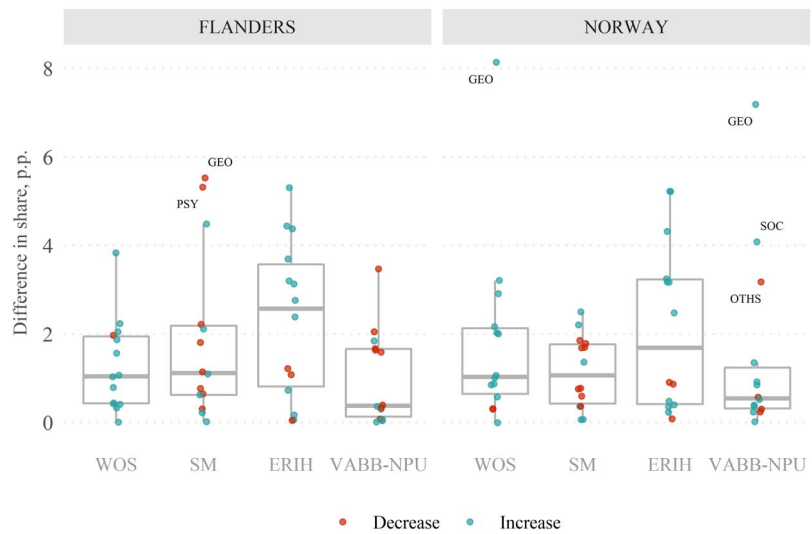
The variability in the differences at the article level can be to a certain extent explained by the unequal distribution of articles across disciplines: *%error* for categories with a lower number of articles will typically appear more substantial than for categories with many articles. For example, the *%error* for "Other social sciences" with only 54 fractionalized articles in the data set VABB-WoS is as high as 670%; using the WoS classification, 421 publications are assigned to the category "Other social sciences." Nevertheless, these findings indicate that there seems to be more agreement on journal-discipline pairs for some categories (e.g., "Economics and business") than others (e.g., "Social and economic geography"). Also, the distribution of differences is not equivalent for both countries. For example, the average *%error* for "Arts" for Flanders across all the comparisons is 45% (*SD* = 55%, *Md* = 30%, range = 3%–119%), while for Norway it is 25% (*SD* = 23%, *Md* = 17%, range = 10%–60%). The explanation for these differences might be sought in the tendency towards disciplinary homogenization (e.g., economics; Lee, Pham, & Gu, 2013) or, conversely, in the coexistence and persistence of pluralism with respect to the identity of a discipline (e.g., economic geography; Barnes & Sheppard, 2010).

6.3. Differences in the Relative Number of Publications and the Disciplinary Structure

In this section we continue with an exploration of changes in the disciplinary structure—the relative distribution of articles across disciplines—that result from the choice of a disciplinary classification. This representation of SSH research is of potential interest in policy settings when trying to gain an insight into the general structure (or change therein) of research in SSH.

From the considerable variation by discipline identified in the previous analyses we expect that the choice of the disciplinary classification may also affect the representations of the disciplinary structure of SSH research. Our analysis, however, shows that this is not the case. The disciplinary structure is influenced only to a minor extent by the choice of disciplinary classification. For all compared pairs of classifications, the average difference in share is very small (*M* = 1.6 p.p., *SD* = 0.9 p.p., *Md* = 1.6 p.p.; Figure 3).

For WoS and SM classifications, the mean differences are in the range 1.2–1.9 p.p. for both Flanders and Norway. For ERIH PLUS, the differences are slightly higher (*M* = 2.3 p.p. for Flanders and *M* = 2.2 p.p. for Norway). Larger differences can be identified for the following: The share of publications in "Social and economic geography" is 8.1 p.p. larger for Norway when using the NPU instead of the WoS classification. For Flanders it is smaller by 5.5 p.p. when using SM. Using ERIH PLUS, the shares of publications in "Languages and literature" are 4.4 p.p. for Flanders and 5.2 p.p. for Norway. Categories with practically no difference for both Flanders and Norway are "Law," "Political science," "Media and communications,"

Figure 3. Overview of difference in share by discipline (in absolute values, jitter (width 0.5) added to avoid overplotting). Each dot represents the difference in share for an individual discipline. Color code indicates whether the share is increasing (blue) or decreasing (orange) when replacing the local classification with another one. Labels of disciplinary categories added only to outliers (PSY—Psychology, SOC—Sociology, GEO—Social and economic geography, OTHS—Other social sciences).

and "Philosophy, ethics, and religion," High similarity only for Flanders can be observed for "Economics and Business" and for Norway for "Educational sciences."

The differences between the two local classifications are even smaller (Flanders: $M = 1.0$ p.p., $Md = 0.4$ p.p.; Norway: $M = 1.4$ p.p., $Md = 0.6$ p.p.). While the largest difference can be identified for the category "Social and economic geography" (3.5 p.p. for Flanders and 7.2 p.p. for Norway), categories with hardly any difference are "Educational sciences," "Law," "Political science," "Media and communications," "Languages and literature," "Philosophy, ethics, and religion," "Arts," and "Other humanities." However, also in the remaining categories the differences are minor. The complete set of results and alluvial diagrams for the remaining pairs of classifications are available as Supplementary Material.

7. DISCUSSION

7.1. Summary of Findings

In this study we have explored how the disciplinary profile—the absolute and the relative number of journal articles in SSH—varies depending on the choice of a disciplinary classification of journals. Specifically, we investigated variations pertaining to specific national databases and disciplines. To this end, we explored pairwise agreement between two local (NPU, VABB) and three international disciplinary classifications (WoS, SM, ERIH PLUS), as well as between the two local classifications. For comparability of the different disciplinary classifications, we made use of OECD FORD: all the classifications were mapped to categories within OECD FORD.

Overall, our study shows that the influence of a disciplinary classification is not equally pronounced across the three analyses pursued here. While the lowest differences were noted in the analysis of disciplinary structure, the highest were present in the analysis of changes in absolute numbers of publications by discipline (Figures 2 and 3, and Tables S1–S7 in the Supplementary Material). The analysis of disciplinary profiles showed modest differences with

a mean Jaccard index of 0.73, on average, for all compared pairs of classifications (SD = 0.39, Md = 1, range = 0.63–0.80).

At the same time, we identified a considerable variation by discipline across all analyses. Categories that tend to be more similar across all classifications and all levels of analysis are "Economics and business," "Educational sciences," "Law," "Media and Communication," and "Languages and Literature." For "Economics and business" and "Law," however, there are exceptions for some pairs of classifications (see Section 6.1). If we consider only the disciplinary profile of articles, high similarity can also be identified for "Psychology" and "Philosophy, ethics, and religion." Although "Psychology" has high similarity also in terms of absolute numbers (only for Flanders), neither this category nor the one for philosophy indicate high similarity in the relative number of publications. Categories that tend to have high differences across all classifications and analyses are "Social and economic geography" and the two residual categories "Other social sciences" and "Other humanities." Considering only the disciplinary profile, variation across pairs of classification can be noted also for "Sociology."

7.2. Implications for Bibliometric Analyses

Methodological requirements for bibliometric analyses vary depending on the focus of the analysis. For this reason, the importance of the choice of a disciplinary classification is not equally high in all analyses. If the focus is on the absolute number of publications per discipline, the choice of the disciplinary classification can lead to substantial over- or underestimation of the number of publications. In contrast, if the focus is on the relative numbers, the choice of classification seems to have practically no influence. This is the main conclusion that can be drawn from our findings.

For example, one might wish to describe and compare SSH research profiles for universities from different countries, say, for benchmarking purposes. If bibliometric data to be used are available only with differing disciplinary classifications, our findings suggest that different disciplinary classifications can be employed, as the influence on the disciplinary structure is minor. This finding resonates with a similar observation in the context of science mapping: different classifications lead to similar maps of science (Rafols & Leydesdorff, 2009). In contrast, when the focus is on publication counts in each discipline, our findings suggest that the disciplinary classification can lead to considerable differences. In such a context, a central question of interest is which classification, besides the local one, can best capture the disciplinary profile of journal articles in SSH, as well as the absolute and relative number of articles by discipline.

Here we used data from two databases developed in national settings (VABB and Cristin). For all analyses of articles in VABB, the two classifications that lead to results most similar to those acquired with the local classification are WoS and NPU ($M_{Jaccard}$ = 0.8 for VABB-WoS and VABB-NPU; $M_{\%error}$ = 24% with SD = 32% for VABB-NPU and $M_{\%error}$ = 64% with SD = 175% for VABB-WoS). The high average difference in *%error* for VABB-WoS results from the residual category "Other social sciences," where the *%error* is 670%. If this is disregarded, then the average *%error* reduces to 17% (SD = 18%) for VABB-WoS and 16% (SD = 17%) for VABB-NPU. WoS and NPU are the most similar to VABB also in terms of relative numbers ($M_{diff.share}$ = 1.4 for WoS-VABB and $M_{diff.share}$ = 1.0 for NPU-VABB). Overall, these findings indicate that for journal articles in VABB, the NPU and WoS can be used as alternatives if for any reason the VABB classification is not available. Of course, not all journals are available in WoS and NPU; this means that the analysis has to be restricted or the classification has to be extended, ideally in consultation with Flemish scholars.

For articles in Cristin, this study indicates differing classifications as the most similar to NPU, the classification used for a subset of articles in Cristin. For example, in terms of disciplinary profile of articles, ERIH PLUS and VABB appear as the most similar classifications ($M_{Jaccard}$ = 0.77 for NPU-ERIH PLUS and $M_{Jaccard}$ = 0.80 for NPU-VABB). In contrast, by absolute numbers of publications the two classifications that lead to the lowest average *%error* are WoS and SM ($M_{\%error}$ = 26% with SD = 26% for NPU-WoS and $M_{\%error}$ = 28% with SD = 22% for SM). Finally, the analysis of changes in disciplinary structure suggests that SM and VABB lead to results most similar to those acquired using NPU, although the variation across the different pairs of classifications is small with respect to this analysis. These somewhat contradictory findings indicate that using VABB and ERIH PLUS it is possible to identify the disciplinary profile of articles that is closest to NPU. However, these two classifications will lead to different numbers of publications by discipline, due to the multiple usage of categories and substantial discrepancies with respect to a number of categories (e.g., "Social and economic geography," "Other social sciences," and "Other humanities"). In other words, most articles that are categorized as "Social and economic geography" in NPU will also be recognized as such using VABB and ERIH PLUS. At the same time, ERIH PLUS and VABB will also identify other articles as belonging to this category, thus increasing the total number of articles in this category. Therefore, if it is important to capture articles in certain disciplines, but it is less important that the delineated data set will include potential false positives, one can use ERIH PLUS and VABB. At the same time, if the number of articles in each discipline is of importance, then it is safer to use WoS or SM.

To sum up, in a comparative context, for Flanders in Belgium and Norway, any of the local classifications can be used; alternatively, each data set can remain with the original classification. However, additional attention should be paid to the categories where larger differences were identified (e.g., "Sociology," "Social and economic geography," "Other social sciences," and "Other humanities"). Ideally, the classification of journals assigned to these categories should be validated by scholars from the respective countries. The assumption here would be that scholars recognize journals from their own discipline. Alternatively, limitations of the classifications in relation to these categories should be acknowledged, especially in contexts where some benchmark measure is calculated at the level of discipline.

Our findings cannot be directly transferred to other contexts without additional explorations that include more databases and classifications. However, this study can inform the choice of a disciplinary classification and the interpretation of comparative bibliometric analyses of SSH that rely on journal-based disciplinary classifications. First, one can keep in mind that changes in disciplinary structure due to the choice of classification tend to be minor. Second, changes in absolute numbers of publications by discipline can be considerable for some categories (e.g., "Social and economic geography"). Thus, if the focus of analysis is on this discipline or any other with the higher variation due to the classification, the journal classification should, ideally, first be validated for the given context to avoid over- or underestimation. This should be taken into account when comparing performance for different countries.

Our findings, alongside the centrality of disciplinary classifications in bibliometrics, indicate that there is a need to further explore the disciplinarity of SSH publications in a manner that accurately depicts research activities in SSH especially when studied comparatively. With this suggestion we continue with conceptual considerations.

7.3. Implications for the Conceptualization of Disciplinarity

All classifications, including disciplinary classifications for SSH journals, carry traces from the context. This means, as shown by our findings, that the assignment of journals to disciplinary

categories depends on context-specific understanding of each discipline and how one might identify journals belonging to the specific disciplines. A particularly interesting result is the difference in the total number of SSH publications; the *%error* for the total number can be as high as 16% (Norway: NPU and SM). This can be interpreted as an indication that in Norway SSH research is understood in a slightly broader sense than in, for example, the SM classification.

At the same time, our findings are in agreement with science mapping studies showing that the boundaries of academic disciplines are much more blurry than we tend to think (e.g., Börner et al., 2012; Klavans & Boyack, 2007). Certain research specialties can reside on the boundaries between two (or more disciplines). An example is the journal *Energy Policy*. While in NPU, it belongs to "Economics and business" and in VABB it is in "Social and economic geography," in SM it appears in "Environmental engineering." WoS assigns this journal to all of those three categories and an additional one: "Earth and related Environmental sciences."

It is tempting to interpret such variations as different levels of accuracy of the disciplinary classifications for journals. Consequently, one may wish to seek ways to improve approaches to classify publications or journal sets as has been done by, for example, Bornmann (2018), Leydesdorff and Bornmann (2016), and Wang and Waltman (2016). In such endeavors there is an implicit assumption that there is one correct category (or set of categories) for each journal. We, in contrast, foreground the contextual embeddedness of disciplinary classifications. This, first of all, means that scholars from the studied context are well placed to judge the accuracy of disciplinary classifications. Second, it means that a single classification that is equally fit for different contexts is unlikely to be achieved.

7.4. Future Research

This brings us to possible directions for future research. Here, we have focused on two local disciplinary classifications. Expanding the analyses with additional local and international classifications (e.g., Scopus, Dimensions) as well as data from other countries would enable more detailed insights.

A reviewer suggested using the principled methodology for comparing the accuracy of clustering solutions proposed by Waltman, Boyack et al. (2020). While we agree that it would be very interesting if this methodology could be applied to journal-based classifications, it poses some challenges that are beyond the scope of the present paper. Most importantly, the method of Waltman et al. (2020) requires that the compared clustering solutions have the same (or similar) granularity. However, our setting lacks a parameter to control for granularity, especially in the case of classifications without multiple categories per journal, and hence this granularity requirement cannot be satisfied in general. Hence, further research is needed before the methodology can be applied to the comparison of journal-based classifications.

The apparent absence of differences in the disciplinary structure might be related to our methodological approach. It is possible that a mixed methods approach would reveal additional contextual features of disciplinary classifications and consequently of metrics for which they are applied. Furthermore, recent studies have shown that article-level classifications can be considerably more accurate than classifications at journal level (Boyack & Klavans, 2020; Klavans & Boyack, 2017; Shu et al., 2019). One journal can contain articles from multiple academic disciplines. For this reason, more fine-grained and potentially more insightful findings could be acquired from comparisons where disciplinary classifications are applied at the article level. Finally, it is worthwhile to seek ways to incorporate in bibliometrics a pluralistic understanding of the studied phenomena (e.g., disciplines), thus acquiring insights that are more representative of the contexts being explored.

DATA AVAILABILITY

The list of journals with the disciplinary codes from all the classifications that were used in the analysis is available as Supplementary Material.

REFERENCES

Archambault, É., Beauchesne, O. H., & Caruso, J. (2011). Towards a Multilingual, Comprehensive and Open Scientific Journal Ontology. *Proceedings of the 13th International Conference of the International Society for Scientometrics and Informetrics*, 66–77. http://www.science-metrix.com/pdf/Towards_a _Multilingual_Comprehensive_and_Open.pdf

Barnes, T. J., & Sheppard, E. (2010). "Nothing includes everything": Towards engaged pluralism in Anglophone economic geography. *Progress in Human Geography*, *34*(2), 193–214. **DOI:** https://doi .org/10.1177/0309132509343728

Bensman, S. J. (2007). Garfield and the impact factor. *Annual Review of Information Science and Technology*, *41*(1), 93–155. **DOI:** https://doi.org/10.1002/aris.2007.1440410110

Bensman, S. J., & Leydesdorff, L. (2009). Definition and identification of journals as bibliographic and subject entities: Librarianship versus ISI *Journal Citation Reports* methods and their effect on citation measures. *Journal of the American Society for Information Science and Technology*, *60*(6), 1097–1117. **DOI:** https://doi.org /10.1002/asi.21020

Bloor, D. (1982). Durkheim and Mauss revisited: Classification and the sociology of knowledge. *Studies in History and Philosophy of Science Part A*, *13*(4), 267–297. **DOI:** https://doi.org/10.1016 /0039-3681(82)90012-7

Börner, K., Klavans, R., Patek, M., Zoss, A. M., Biberstine, J. R., … Boyack, K. W. (2012). Design and Update of a Classification System: The UCSD Map of Science. *PLOS ONE*, *7*(7), e39464. **DOI:** https://doi.org/10.1371/journal.pone.0039464, **PMID:** 22808037, **PMCID:** PMC3395643

Bornmann, L. (2018). Field classification of publications in Dimensions: A first case study testing its reliability and validity. *Scientometrics*, *117*(1), 637–640. **DOI:** https://doi.org/10.1007/s11192-018-2855-y

Bowker, G. C., & Star, S. L. (2000). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press. **DOI:** https:// doi.org/10.7551/mitpress/6352.001.0001

Boyack, K. W., & Klavans, R. (2020). A comparison of large-scale science models based on textual, direct citation and hybrid relatedness. *Quantitative Science Studies*, *1*(4), 1570–1585. **DOI:** https://doi.org/10.1162/qss_a_00085

Carley, S., Porter, A. L., Rafols, I., & Leydesdorff, L. (2017). Visualization of disciplinary profiles: Enhanced science overlay maps. *Journal of Data and Information Science*, *2*(3), 68–111. **DOI:** https://doi.org/10.1515/jdis-2017-0015

Daraio, C., & Glänzel, W. (2016). Grand challenges in data integration—state of the art and future perspectives: An introduction. *Scientometrics*, *108*(1), 391–400. **DOI:** https://doi.org/10.1007/s11192-016-1914-5

Depaepe, M. (2002). A comparative history of educational sciences: The comparability of the incomparable? *European Educational Research Journal*, *1*(1), 118–122. **DOI:** https://doi.org/10.2304/eerj.2002.1.1.2

Durkheim, É., & Mauss, M. (2010). *Primitive classification*. London: Routledge. **DOI:** https://doi.org/10.4324/9780203092828

Engels, T. C. E., Istenič Starčič, A., Kulczycki, E., Pölönen, J., & Sivertsen, G. (2018). Are book publications disappearing from scholarly communication in the social sciences and humanities? *Aslib Journal of Information Management*, *70*(6), 592–607. **DOI:** https://doi.org/10.1108/AJIM-05-2018-0127

Eykens, J., Guns, R., & Engels, T. C. E. (2019). Article level classification of publications in sociology: An experimental assessment of supervised machine learning approaches. *Proceedings of the 17th Conference of the International Society for Scientometrics and Informetrics*, *1*, 738–743. https://repository.uantwerpen.be/docman/iruaauth/5722dd/163024.pdf

Friese, C. (2010). Classification conundrums: Categorizing chimeras and enacting species preservation. *Theory and Society*, *39*(2), 145–172. **DOI:** https://doi.org/10.1007/s11186-009-9103-7

Glänzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, *56*(3), 357–367. **DOI:** https://doi.org/10.1023/A:1022378804087

Gläser, J., Glänzel, W., & Scharnhorst, A. (2017). Same data—different results? Towards a comparative approach to the identification of thematic structures in science. *Scientometrics*, *111*(2), 981–998. **DOI:** https://doi.org/10.1007/s11192-017-2296-z

Guns, R., Sīle, L., Eykens, J., Verleysen, F. T., & Engels, T. C. E. (2018). A comparison of cognitive and organizational classification of publications in the social sciences and humanities. *Scientometrics*, *116*(2), 1093–1111. **DOI:** https://doi.org/10.1007/s11192-018-2775-x

Hammarfelt, B. (2018). What is a discipline? The conceptualization of research areas and their operationalization in bibliometric research. *23rd International Conference on Science and Technology Indicators (STI 2018)*. Leiden, The Netherlands. https://www.researchgate.net/publication/327344547

Hammarfelt, B., & de Rijcke, S. (2015). Accountability in context: Effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the faculty of Arts at Uppsala University. *Research Evaluation*, *24*(1), 63–77. **DOI:** https://doi.org/10.1093/reseval/rvu029

Harzing, A.-W. (2019). Two new kids on the block: How do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science? *Scientometrics*, *120*(1), 341–349. **DOI:** https://doi.org/10.1007/s11192-019-03114-y

Haunschild, R., Marx, W., French, B., & Bornmann, L. (2018). Relationship between field-normalized indicators calculated with different approaches of field-categorization. *STI 2018 Conference Proceedings* (pp. 11–18). https://hdl.handle.net/1887/65267

Hicks, D. (2004). The four literatures of social science. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research* (pp. 473–496). Kluwer Academic. **DOI:** https://doi.org/10.1007/1-4020-2755-9_22

Hug, S. E., & Brändle, M. P. (2017). The coverage of Microsoft Academic: Analyzing the publication output of a university. *Scientometrics*, *113*(3), 1551–1571. **DOI:** https://doi.org/10.1007/s11192-017-2535-3

Janssens, F., Zhang, L., Moor, B. D., & Glänzel, W. (2009). Hybrid clustering for validation and improvement of subject-classification schemes. *Information Processing & Management*, *45*(6), 683–702. **DOI:** https://doi.org/10.1016/j.ipm.2009.06.003

Klavans, R., & Boyack, K. W. (2007). Is there a convergent structure of science? A comparison of maps using the ISI and Scopus databases. *Proceedings of ISSI 2007: 11th International Conference of the International Society for Scientometrics and Informetrics* (pp. 437–448). https://pdfs.semanticscholar.org/6c89/a7baef42f4f409f1a2c743530c980c40ec33.pdf?ga=2.230941528.658141258.1557390573-2087353322.1554993244

Klavans, R., & Boyack, K. W. (2017). Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? *Journal of the Association for Information Science and Technology*, *68*(4), 984–998. **DOI:** https://doi.org/10.1002/asi.23734

Kousha, K., Thelwall, M., & Rezaie, S. (2011). Assessing the citation impact of books: The role of Google Books, Google Scholar, and Scopus. *Journal of the American Society for Information Science and Technology*, *62*(11), 2147–2164. **DOI:** https://doi.org/10.1002/asi.21608

Kronegger, L., Mali, F., Ferligoj, A., & Doreian, P. (2015). Classifying scientific disciplines in Slovenia: A study of the evolution of collaboration structures. *Journal of the Association for Information Science and Technology*, *66*(2), 321–339. **DOI:** https://doi.org/10.1002/asi.23171

Kulczycki, E., Engels, T. C. E., Pölönen, J., Bruun, K., Dušková, M., … Zuccala, A. (2018). Publication patterns in the social sciences and humanities: Evidence from eight European countries. *Scientometrics*, *116*(1), 463–486. **DOI:** https://doi.org/10.1007/s11192-018-2711-0

Kulczycki, E., Guns, R., Pölönen, J., Engels, T. C. E., Rozkosz, E. A., … Sivertsen, G. (2020). Multilingual publishing in the social sciences and humanities: A seven-country European study. *Journal of the Association for Information Science and Technology*, asi.24336. **DOI:** https://doi.org/10.1002/asi.24336, **PMID:** 33288998, **PMCID:** PMC7687152

Lampland, M. (2009). Classifying laborers: Instinct, property, and the psychology of productivity in Hungary (1920–1956). In M. Lampland & S. L. Star (Eds.), *Standards and their stories: How quantifying, classifying, and formalizing practices shape everyday life* (pp. 123–142). Ithaca, NY: Cornell University Press.

Lavik, G. A. V., & Sivertsen, G. (2017). Erih Plus—Making the SSH visible, searchable and available. *Procedia Computer Science*, *106*, 61–65. **DOI:** https://doi.org/10.1016/j.procs.2017.03.035

Lee, F. S., Pham, X., & Gu, G. (2013). The UK Research Assessment Exercise and the narrowing of UK economics. *Cambridge Journal of Economics*, *37*(4), 693–717. **DOI:** https://doi.org/10.1093/cje/bet031

Leydesdorff, L., & Bornmann, L. (2016). The operationalization of "fields" as WoS subject categories (WCs) in evaluative bibliometrics: The cases of "library and information science" and "science & technology studies". *Journal of the Association for Information Science and Technology*, *67*(3), 707–714. **DOI:** https://doi.org/10.1002/asi.23408

Leydesdorff, L., Bornmann, L., & Zhou, P. (2016). Construction of a pragmatic base line for journal classifications and maps based on aggregated journal-journal citation relations. *Journal of Informetrics*, *10*(4), 902–918. **DOI:** https://doi.org/10.1016/j.joi.2016.07.008

Leydesdorff, L., & Milojević, S. (2015). The citation impact of German sociology journals: Some problems with the use of scientometric indicators in journal and research evaluations.

Soziale Welt, 66(2), 193–204. **DOI:** https://doi.org/10.5771/0038-6073-2015-2-193

Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology*, 60(2), 348–362. **DOI:** https://doi.org/10.1002/asi.20967

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*, 66(1), 81–100. **DOI:** https://doi.org/10.1007/s11192-006-0007-2

OECD. (2007). *Revised Field of Science and Technology (FOS) classification in the Frascati Manual*. http://www.oecd.org/science/inno/38235147.pdf

OECD. (2015). *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development. OECD Publishing*. **DOI:** https://doi.org/10.1787/9789264239012-en

OECD Category to Web of Science Category Mapping 2012. (2012). http://help.prod-incites.com/inCites2Live/5305-TRS/version/default/part/AttachmentData/data/OECD%20Category%20Mapping.xlsx

Ossenblok, T. L. B., Engels, T. C. E., & Sivertsen, G. (2012). The representation of the social sciences and humanities in the Web of Science—A comparison of publication patterns and incentive structures in Flanders and Norway (2005–9). *Research Evaluation*, 21(4), 280–290. **DOI:** https://doi.org/10.1093/reseval/rvs019

Penissat, E., & Rowell, J. (2015). The creation of a European socio-economic classification: Limits of expert-driven statistical integration. *Journal of European Integration*, 37(2), 281–297. **DOI:** https://doi.org/10.1080/07036337.2014.990140

Petr, M., Engels, T. C. E., Kulczycki, E., Duskova, M., Guns, R., Sieberova, M., & Sivertsen, G. (2020). Coverage of journal articles in social sciences and humanities in Web of Science and their representation in citation indexes: A comparison of five European countries. *ArXiv:2006.07103 [Cs]*. http://arxiv.org/abs/2006.07103

Pölönen, J., Engels, T., & Guns, R. (2019). Ambiguity in identification of peer-reviewed publications in the Finnish and Flemish performance-based research funding systems. *Science and Public Policy*, 47(1), 1–15. **DOI:** https://doi.org/10.1093/scipol/scz041

Pudovkin, A. I., & Garfield, E. (2002). Algorithmic procedure for finding semantically related journals. *Journal of the American Society for Information Science and Technology*, 53(13), 1113–1119. **DOI:** https://doi.org/10.1002/asi.10153

Rafols, I., & Leydesdorff, L. (2009). Content-based and algorithmic classifications of journals: Perspectives on the dynamics of scientific communication and indexer effects. *Journal of the American Society for Information Science and Technology*, 60(9), 1823–1835. **DOI:** https://doi.org/10.1002/asi.21086

Shu, F., Julien, C.-A., Zhang, L., Qiu, J., Zhang, J., & Larivière, V. (2019). Comparing journal and paper level classifications of science. *Journal of Informetrics*, 13(1), 202–225. **DOI:** https://doi.org/10.1016/j.joi.2018.12.005

Sīle, L., & Guns, R. (2019). Article-level matching of Web of Science to a local database in a comparative context. *Proceedings of the 17th Conference of the International Society for Scientometrics and Informetrics*, 2. http://issi-society.org/proceedings/issi_2019/ISSI%202019%20-%20Proceedings%20VOLUME%20II.pdf

Sīle, L., & Guns, R. (2020). *Article-level matching of Web of Science to a local database in a comparative context* (p. 7). ECOOM, University of Antwerp. **DOI:** https://doi.org/10.6084/m9.figshare.13013816

Sīle, L., Guns, R., Vandermoere, F., & Engels, T. C. E. (2019). Comparison of classification-related differences in the distribution of journal articles across academic disciplines: The case of social sciences and humanities in Flanders and Norway (2006–2015). *Proceedings of the 17th Conference of the International Society for Scientometrics and Informetrics*, 1, 1092–1097. http://issi-society.org/proceedings/issi_2019/ISSI%202019%20-%20Proceedings%20VOLUME%20I.pdf

Sīle, L., Pölönen, J., Sivertsen, G., Guns, R., Engels, T. C. E., Arefiev, P., Dušková, M., Faurbæk, L., Holl, A., Kulczycki, E., Macan, B., Nelhans, G., Petr, M., Pisk, M., Soós, S., Stojanovski, J., Stone, A., Šušol, J., & Teitelbaum, R. (2018). Comprehensiveness of national bibliographic databases for social sciences and humanities: Findings from a European survey. *Research Evaluation*, 27(4), 310–322. **DOI:** https://doi.org/10.1093/reseval/rvy016

Sivertsen, G. (2016). Publication-based funding: The Norwegian model. In M. Ochsner, S. E. Hug, & H.-D. Daniel (Eds.), *Research Assessment in the Humanities* (pp. 79–90). Cham: Springer. **DOI:** https://doi.org/10.1007/978-3-319-29016-4_7

Sivertsen, G. (2019). Developing Current Research Information Systems (CRIS) as data sources for studies of research. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 667–683). Cham: Springer. **DOI:** https://doi.org/10.1007/978-3-030-02511-3_25

Small, M. L. (1999). Departmental conditions and the emergence of new disciplines: Two cases in the legitimation of African-American studies. *Theory and Society*, 28(5), 659–707. **DOI:** https://doi.org/10.1023/A:1007034317576

Star, S. L., & Bowker, G. C. (2007). Enacting silence: Residual categories as a challenge for ethics, information systems, and communication. *Ethics and Information Technology*, 9(4), 273–280. **DOI:** https://doi.org/10.1007/s10676-007-9141-7

Sugimoto, C. R., & Weingart, S. (2015). The kaleidoscope of disciplinarity. *Journal of Documentation*, 71(4), 775–794. **DOI:** https://doi.org/10.1108/JD-06-2014-0082

Vanderstraeten, R., & Vandermoere, F. (2015). Disciplined by the discipline: A social-epistemic fingerprint of the history of science. *Science in Context*, 28(2), 195–214. **DOI:** https://doi.org/10.1017/S0269889715000058, **PMID:** 25996853

Verleysen, F. T., Ghesquière, P., & 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 W. Blockmans, L. Engwall, & D. Weaire (Eds.), *Bibliometrics: Use and abuse in the review of research performance* (pp. 115–125). London: Portland Press. http://www.portlandpress.com/pp/books/online/wg87/087/0117/0870117.pdf

Waltman, L., Boyack, K. W., Colavizza, G., & van Eck, N. J. (2020). A principled methodology for comparing relatedness measures for clustering publications. *Quantitative Science Studies*, 1(2), 691–713. **DOI:** https://doi.org/10.1162/qss_a_00035

Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics*, 10(2), 347–364. **DOI:** https://doi.org/10.1016/j.joi.2016.02.003

Zerubavel, E. (1996). Lumping and splitting: Notes on social classification. *Sociological Forum*, 11(3), 421–433. **DOI:** https://doi.org/10.1007/BF02408386