Universiteit Antwerpen

This item is the archived peer-reviewed author-version of:

A comparison of cognitive and organizational classification of publications in the social sciences and humanities

Reference:
Guns Raf, Sīle Linda, Eykens Joshua, Verleysen Frederik, Engels Tim.- A comparison of cognitive and organizational classification of publications in the social sciences and humanities
Scientometrics: an international journal for all quantitative aspects of the science of science and science policy - ISSN 0138-9130 - 116:2(2018), p. 1093-1111
Full text (Publisher's DOI): https://doi.org/10.1007/S11192-018-2775-X
To cite this reference: https://hdl.handle.net/10067/1512450151162165141

uantwerpen.be
Institutional repository IRUA

A comparison of cognitive and organizational classification of publications in the Social Sciences and Humanities

Raf Guns, Linda Sīle, Joshua Eykens, Frederik T. Verleysen & Tim C.E. Engels

raf.guns@uantwerpen.be, linda.sile@uantwerpen.be, joshua.eykens@uantwerpen.be, frederik.verleysen@uantwerpen.be, tim.engels@uantwerpen.be

Centre for R&D Monitoring (ECOOM), Faculty of Social Sciences, University of Antwerp, Antwerp (Belgium)

Abstract

We study the discrepancy between two ways of classifying publications in the social sciences and humanities (SSH): on the basis of the contents of publications and publication channels (cognitive classification) or on the basis of the organizational structure of departments, faculties etc. (organizational classification). Using data for the period 2000-2015 from 76,076 peer reviewed publications (co-)authored by SSH researchers affiliated to a Flemish university, we compare the organizational classification of the publications with a cognitive classification based on OECD Fields Of Science. In all fields except one, the cognitive discipline with the most publications is the one that most closely matches the organizational discipline, although there are substantial differences between disciplines in terms of overlap and concentration. At a higher aggregation level, we find that 73% of publications from the humanities are published in humanities publication channels, while this is only the case for 59% of publications from the social sciences. Social sciences are shown to have important links to medicine and health sciences. The spread of publications over cognitive disciplines, including non-SSH disciplines, is only partially related to the assignment of some publication channels to multiple FOS fields as well as multidisciplinary collaboration. Our study shows that it is quite common for researchers affiliated to an organizational unit to also carry out research in other domains.

Keywords: Social sciences, Humanities, Cognitive classification, Organizational classification

JEL Classification I23

Mathematics Subject Classification 62P25

Introduction

Publications, publication channels, and authors can be assigned to disciplines on the basis of different criteria. The most intuitive criterion is presumably the contents of the publications, but in practice other criteria are frequently used as well. Daraio and Glänzel (2016) distinguish between four types of classifications:

- *administrative*: based on responsibility from the point of view of a government or research policy;
- *cognitive*: based on content of publications or publication channels;
- *organizational*: based on internal organizational structure;
- *qualification-based*: based on education and competencies.

However, little is known on the relations between these types. In this paper, we focus on the relation between organizational and cognitive classifications. Whitley (2000) has pointed out the "degree of separation of control" between the praxis of research and "administrative structures of work organizations." In the case of Flanders, Belgium, for instance, the organizational structure of universities – departments, faculties, etc. – traditionally results mainly from educational needs rather than research.

One might expect that in practice some or all of these classifications tend to coincide, that is, that the four classifications will generally lead to a publication or its authors being classified in the same discipline(s). At the same time, it is clear that this is not absolute: for instance, a researcher with formal training in economics may carry out computer science research while working in a physics department. Relatedly, it has been shown that researchers from the same organizational discipline may be quite different in terms of publication patterns (Verleysen & Weeren, 2016a, 2016b).

Daraio and Glänzel (2016) discuss the case of the Flemish Academic Bibliographic Database for Social Sciences and Humanities (VABB-SHW, see https://www.ecoom.be/en/vabb), a comprehensive database of social sciences and humanities (SSH) output in Flanders (Verleysen, Ghesquière, & Engels, 2014): they point out that the Flemish performance-based research funding system uses a mixture of cognitive (Web of Science) and organizational (VABB-SHW) classification. Since its launch in 2010 the VABB-SHW classifies publications according to the authors' departments or research groups and is hence based on organizational criteria. This way of classifying publications is also used in other performance-based funding systems (e.g. the Polish PBN, see Kulczycki, Engels, & Nowotniak, 2017). Several national comprehensive coverage databases of research publications, however, contain a cognitive classification of publications (Kulczycki et al., 2018). Thus far, to the best of our knowledge, no European comprehensive database of (SSH) publications classifies publications according to two or more types of classifications. In order to fill this gap, we present in this paper the results of the elaboration of the VABB-SHW with a cognitive classification.

The basic motivation for the addition of a cognitive classification to the VABB-SHW emerges from the need to further our understanding of disciplinary boundaries and cross-disciplinary collaborations in the SSH. Often, the SSH have appeared as fragmented, and rather different from the biomedical, natural and technical sciences in terms of research and publication patterns (Hicks, 2004; Nederhof, 2006). Many scholars, both within and outside the SSH, seem to take this view of the SSH as a distinct realm for granted. At the same time, however, advocacy groups, funding bodies and governments alike have pushed for the integration of SSH into interdisciplinary research programs (Pedersen, 2016). These demands stem, to some extent, from the need to approach complex societal challenges in a concerted and integrated way from diverse angles at the same time, from the need to adopt new analytic and technological possibilities into research processed in all fields, and from the need to sustain the broader, societal impact of research. Indeed, several SSH fields have undergone and still are undergoing a transformation as a result of integration (e.g. of economists and sociologists into the field of environmental sciences), of interdisciplinary collaboration (e.g. in archaeology with the adoption of analytical tools developed in the chemical and engineering sciences), and of addressing pressing societal issues (e.g. in philosophy and psychology through intense collaboration with medical fields). Moreover, research into interdisciplinarity in the SSH has shown that in terms of organizational structures disciplines and interdisciplines are often closely intertwined (Abbott, 2001; Frickel & Ilhan, 2017), sometimes to the extent that disciplinary boundaries almost entirely disappear (the idea of post-disciplinarity, see Hay, 2010). In sum, we expect a comparison of a cognitive and an organizational classification of publications to reflect the crossing of boundaries that has been observed both in terms of research questions tackled and approaches adopted, as well as in terms of organizational frictions.

On a more practical level, the availability of both a cognitive and an organizational classification within the same system allows for the use of the data in diverse contexts. Whereas organizational classification of publications may be sufficient for resource allocation at the national level, it may not be for fine-grained evaluation exercises (Daraio and Glänzel, 2016).

In addition, a cognitive classification facilitates the comparison of publication patterns across countries (Ossenblok, Engels, & Sivertsen, 2012; Pölönen, Engels, Guns, Sivertsen, & Verleysen, 2017).

The aim of this exploratory paper is to investigate the commonalities and differences between cognitive and organizational classifications using empirical data from the VABB-SHW. In the next section, we describe the data set in detail and discuss our methods, including the steps taken to obtain a cognitive classification of all peer-reviewed publications in the VABB-SHW. Since a publication can be classified into more than one category in both the organizational and the cognitive classification, the Results section starts by describing the structure of both co-occurrence networks. Afterwards we explore to what extent the cognitive and organizational classification agree at the level of broad areas of science and at the discipline level. A substantial portion of research by SSH scholars turns out to be in other areas of research, according to a cognitive classification. We show that the same data can be used to characterize the concentration of one organizational discipline over cognitive disciplines. In the Discussion section, we discuss the implications and limitations of our work. Finally, we present our conclusions and some suggestions for future research.

This study is an extended version of a conference paper presented during the 2017 ISSI conference in Wuhan, China (Guns, Engels, & Verleysen, 2017). Apart from our own previous work, we have not been able to find other research that empirically studies the relation between cognitive and organizational classifications.

Data and methods

VABB-SHW and organizational classification

Data were obtained from the VABB-SHW. All 76,076 publications that were classified as SSH, published in the period 2000–2015 and classified as peer-reviewed in the context of the Flemish funding system were taken into account. Both publications that are indexed in Web of Science (WoS) and publications that have been selected by the Authoritative Panel (GP), an independent panel of senior professors from different SSH disciplines, are taken into account. Table 1 provides an overview of the numbers per publication type. It is clear that journal articles are the most prevalent type; in total these articles have appeared in 10,038 journals.

The organizational classification in VABB-SHW assigns each publication to one or more VABB disciplines according to the authors' departments or research groups. The mapping of an organizational unit to a VABB discipline is carried out by the University of Antwerp branch of the Flemish Centre for R&D Monitoring (ECOOM-Antwerp), which maintains the VABB-SHW database, and approved by the research administration offices of the Flemish universities. In this organizational classification we distinguish 16 VABB disciplines, of which 7 belong to the Humanities and 9 to the Social Sciences. In addition, two broader categories *Social sciences general* and *Humanities general* are used. Most discipline names clearly label the discipline in question. It is worth noting, however, that *History of arts* also encompasses architecture and arts.

Table 1. Overview of publication types in the data set

Publication type	**Selection by**	***N***
Journal articles	GP	27,749
	WoS	30,569
Monographs	GP	1,199

Edited books	GP	2,057
Book chapters	GP	11,799
Conference proceedings	GP	1,380
	WoS	1,323

Cognitive classification

Now we turn to the cognitive classification of this set of publications. We assigned each publication to one or more disciplines from a modified version of the OECD Fields Of Science or FOS (OECD, 2007).[1] The FOS classification was chosen for mainly pragmatic reasons. First, several mappings from other well-known science classifications exist, including the ones in use by Web of Science and Scopus. Second, OECD FOS is used in several other countries and its implementation can hence facilitate cross-country comparisons. Finally, it is fairly broad with just 42 fields at the second level, thereby making classification somewhat easier.

A first modification to the FOS classification scheme was the addition of a separate field *Multidisciplinary* to cater for publication outlets like *PLOS ONE*. This field is similar to the category *Multidisciplinary* in Scopus or *Multidisciplinary sciences* in Web of Science. Second, the following three refinements were made to the FOS classification:

- superclass *History and archaeology* was split into subclasses *History* and *Archaeology*;
- superclass *Languages and literature* was split into subclasses *Languages and linguistics* and *Literature*;
- superclass *Philosophy, ethics and religion* was split into subclasses *Philosophy and ethics* and *Religion*.

The rationale behind these refinements follows from the observed need for a more fine-grained classification of humanities disciplines in other national publication databases that have implemented a cognitive classification based on FOS (e.g. VIRTA in Finland). Apart from the groupings *Humanities general* and *Social sciences general*, only the VABB discipline *Criminology* does not have a direct counterpart in the adapted FOS scheme, where this subject is classified under *Law*. From the perspective of FOS, only *Social and economic geography* (besides *Other humanities* and *Other social sciences*) does not have a clear counterpart in VABB. Although we use a slightly refined version of OECD FOS, we will refer to it as 'FOS' where there is no danger of confusion. Moreover it should be stressed that the refined scheme can easily be translated to the original Fields Of Science by leaving out *Multidisciplinary* and mapping each refinement to its superclass.

Automatic classification

The assignment of publications to FOS fields was done as much as possible by relying on the classification of publications and publication channels in external data sources. More specifically, where possible, journal articles were assigned to FOS fields on the basis of how the journal was classified in external sources. Similarly, publications with an ISBN (books, proceedings and book chapters) were – where possible – classified based on how the book was classified externally.

[1] An updated version of the Frascati Manual (OECD, 2015) refers to these as the Fields of Research and Development (FORD), – apart from a few superficial name changes the fields themselves remain unchanged in the newer version.

Table 2 summarizes which external data sources were used, which classification scheme(s) they use and how many VABB-SHW publications, or the publication channels thereof, were found in each. For every classification scheme, a concordance table was used to translate to the FOS classification. FOS fields were assigned to journals as follows: for a given journal, the FOS fields that result from translating the classifications in WoS, Scopus and the Norwegian Register for Scientific Journals (using data from Ossenblok et al., 2012) were taken. Then, only the field(s) that occur(s) most often were assigned to this journal. This way, we try to partially account for idiosyncrasies that may occur in these databases. To give an example, the journal *Scientometrics* can be found in WoS, Scopus, and the Norwegian Register for Scientific Journals. All three sources classify it as *Media and communications* (under which library and information science is grouped), whereas only WoS and Scopus classify it as *Computer and information sciences*, and only Scopus as *Other social sciences* and *Law*. Hence, only the classification as *Media and communications* is retained.

Only when none of these three sources classified a given journal, the classification from ISSN.org was used. The latter data source was given lower priority since it often only provides a top-level classification for journals that are in fact more specific. For example, the *Proceedings of the Royal Society B: Biological Sciences* was classified as DDC 500 (Natural Sciences and Mathematics), even though the journal only publishes biological research.

Table 2. Overview of data sources and number of VABB-SHW publication channels and publications classified using them.

Data source	Used for	Classification scheme(s)	URL	*N*
Web of Science	6,474 journals	Web of Science Subject Categories	http://www.webofknowledge.com	37,376
Scopus	7,208 journals	All Science Journal Classification (ASJC)	https://www.elsevier.com/solutions/scopus/content	46,785
Norwegian Register for Scientific Journals	4,030 journals	NPI scientific fields	https://dbh.nsd.uib.no/publiseringskanaler/Om	41,570
ISSN.org	945 journals	- Dewey Decimal Classification (DDC) - Universal Decimal Classification (UDC)	https://portal.issn.org/	2,123
OCLC Classify	7,935 books and proceed-ings	- Dewey Decimal Classification (DDC) - Library of Congress Classification (LCC)	http://classify.oclc.org/classify2/	14,149

For publications with an ISBN, the OCLC Classify web service was used, which provides a RESTful API to look up the DDC and LCC codes assigned to a given ISBN in WorldCat. For each ISBN (7,935 in total) we used the DDC and/or LCC code that was most frequently assigned based on holdings counts. These were in turn translated to one or more FOS fields.

Manual classification

A limited number of publications (*N*=6,072) did not have a subject indication in any of the aforementioned data sources and was classified manually and independently by two of the co-authors of this paper. The main considerations guiding the manual assignment of records to academic disciplines were derived from the list of disciplines and research specialties associated with each Field Of Science (OECD, 2007). For example, categories Anthropology and Social Work were included in the more general category *Sociology*, while the parts of Information science addressing social aspects were assigned to the category *Media and communications*. In addition to these considerations, further assignment decisions were documented and developed in ECOOM-Antwerp. For example, the discipline Philology was classified as *Languages and linguistics*, but Graphic design as *Arts* unless a publication was addressing issues related to media and/or communication (in that case, *Media and communications*) or marketing (*Economics and business*). All cases where there was disagreement were resolved by rechecking the classification and making a final decision, where necessary through discussion among both raters and/or other colleagues. To assess the inter-rater reliability, we use the joint probability of agreement, i.e. the fraction of publications that were classified in the same way. Because a publication can be assigned to more than one field, we report an upper and a lower bound. The upper bound is obtained by assuming that if at least one field matches for a publication, there is agreement. The lower bound is obtained by only counting those cases where the exact same fields have been assigned. The upper and lower bound are 0.64 and 0.58 respectively. The assignment of records to disciplines was carried out by all five authors and the work was split into subsets of records (linked to different ongoing studies), each subset being classified by two people. Of these subsets, the one with the highest joint probability of agreement the upper and lower bound are the same, namely 0.69. The one with the lowest probability had an upper and lower bound of 0.55 and 0.48 respectively.

After all peer-reviewed publications had been classified either automatically or manually, the following corrections were made manually:

- There were 24 publications with 6 or more (up to 8) FOS fields. These were reviewed. Since it appeared that they could be classified as at most two fields, they were reclassified manually into one or two FOS fields.
- We made sure that the class *Multidisciplinary* was not used together with other labels. 426 publications remain as multidisciplinary.
- Several publications were assigned to both a top-level class and one of its subclasses. In that case, only the subclass was retained.
- For the remaining 2,002 cases in which a top-level class was assigned, the top-level class was replaced with its last subclass (*Other…*). E.g., *Medical and health sciences* was replaced with *Other medical science*.
- As explained earlier, we implemented three refinements of FOS. If a class like *History and archaeology* was used, it was replaced with the combination of its subclasses.

As a result, publications were classified into, at most, 5 FOS fields (Table 3). An example is the journal *Agriculture and human values*, publications in which are classified as *Biological sciences*; *Agriculture, forestry, and fisheries*; *Agricultural biotechnology*; *Sociology*; and *History*.

Analysis

We study the co-occurrence among VABB disciplines and among FOS fields using bibliometric maps visualized with VOSviewer (Van Eck & Waltman, 2010). The VABB map is based on a

weighted network in which the nodes represent VABB disciplines and the weight of an edge is the number of times the two disciplines are assigned to the same publication. This happens in the case of co-authorship between researchers from different administrative units, but also in the case where one researcher has multiple organizational affiliations. The FOS map is also based on a weighted network. Here, nodes represent FOS fields and the weight of an edge is the number of publications to which both FOS fields are co-assigned.

Next, we directly compare the cognitive and organizational classification using a heatmap visual representation. For each pair of VABB discipline and FOS field, we determine the number of publications assigned to both. Because the absolute numbers of publications vary between VABB disciplines and FOS fields, we use the share of publications, ranging between 0 and 1. This can be done from the perspective of a VABB discipline – share of publications from that discipline assigned to a FOS field compared to all publications from that discipline – or from the perspective of a FOS field. For the purposes of the present paper, we will mostly consider things from the perspective of VABB disciplines, i.e. the organizational classification of the publications.

As mentioned earlier, publications can be assigned to more than one VABB discipline (23.9%) and/or to more than one FOS field (7.8%). Indeed, it seems plausible that interdisciplinary collaboration is one of the driving forces enabling researchers to go beyond their own organizational discipline (van Rijnsoever & Hessels, 2011). To assess the strength of this effect, we separately consider the set of publications assigned to only one VABB discipline and one FOS field (N=53,568 or 70.4%).

In addition to visual representations using heatmaps, we quantify the concentration of a VABB discipline in FOS fields using the Gini inequality index (Gini, 1921). This index is equal to two times the area between the Lorenz curve and the diagonal (Rousseau, 2011) and varies between 0 (perfect evenness) and 1 (perfect unevenness, i.e. concentration). In the situation where a VABB discipline perfectly corresponds to one FOS field, we would obtain a Gini index equal to 1.

Analysis files are publicly available from https://github.com/rafguns/cogn-org-classification.

Limitations

Our approach to cognitive classification has two main limitations. Journal articles are classified by assigning each journal to one or several academic disciplines. Most journals are classified automatically using classifications in four different sources as described above. Even though the schemes used in all these sources are known and can be mapped onto the FOS classification, the principles guiding assignment of journals in each source are largely non-published and may not always be the same. For instance, Pudovkin and Garfield (2002, footnote 1) report that journal categories are assigned using a combination of manual, "heuristic" methods and a citation-based algorithm (see also Leydesdorff & Bornmann, 2016), whereas the ISSN database cannot take citation information into account. The validity of the final classification drawing on these various sources thus hinges on the validity of the classification in those sources.

Further, the manual assignment of records to disciplines was carried out in different stages. Although general principles guiding the assignment were agreed upon from the beginning, more fine-grained assignment rules were developed and documented during the process. Consequently, there may be some variation over time in the principles guiding the assignment and among the people carrying out the classification. Due to the relatively small share of records

classified manually, it is reasonable to assume that the results of analysis presented here are not greatly altered in case the manual approach to classification has led to some inaccuracies.

Results

Co-occurrences among VABB disciplines

We find up to 8 VABB disciplines for one publication (Table 3). Figure 1 presents the co-occurrence among VABB disciplines (organizational classification). We observe a clear distinction between the Humanities and the Social Sciences, with an interesting middle position for Philosophy. Indeed Philosophy is known for its capacity to encompass many different topics, crossing disciplinary boundaries (Lamont, 2009). Further we observe strong links of sociology and economics and business with several of the disciplines, especially other social sciences disciplines. Psychology appears more at the outside of the map, probably because of its relation to medical fields; law for its part is sometimes distinguished from SSH (van Leeuwen, 2013). Among the humanities we observe a cluster around linguistics, perhaps due to the leading role of this discipline in the area of digital humanities (Bod, 2012).

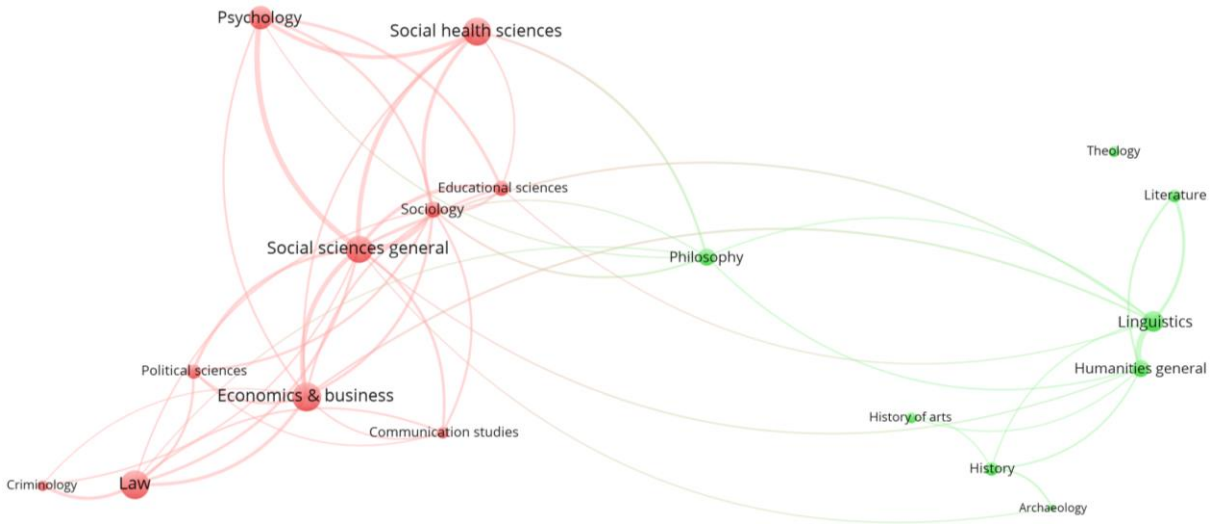

Figure 1. Map of co-occurrences among VABB disciplines. Only 50 most prominent links are shown. This map can be web-started from http://bit.ly/vabb-coocc.

Co-occurrences among FOS fields

If two FOS fields occur together on a single publication, they are related in that this publication or publication channel addresses topics that concern both fields. Up to 5 FOS fields can be assigned to a single publication (Table 3). We visualize the co-occurrences of FOS fields in the VABB-SHW data set with a bibliometric map (Figure 2). In the map, nodes are scaled according to the number of publications in VABB-SHW. FOS fields belonging to the same top-level field have the same colour.

Table 3. Numbers of publications with a given number of VABB disciplines and FOS fields

	VABB disciplines							
	1	2	3	4	5	6	7	8

FOS fields	1	53,568	12,809	2,887	717	120	19	5	0
	2	3,804	1,035	240	120	8	0	0	1
	3	425	125	23	3	1	0	0	0
	4	109	37	8	2	1	0	0	0
	5	7	1	1	0	0	0	0	0
		57,913	14,007	3,159	842	130	19	5	1

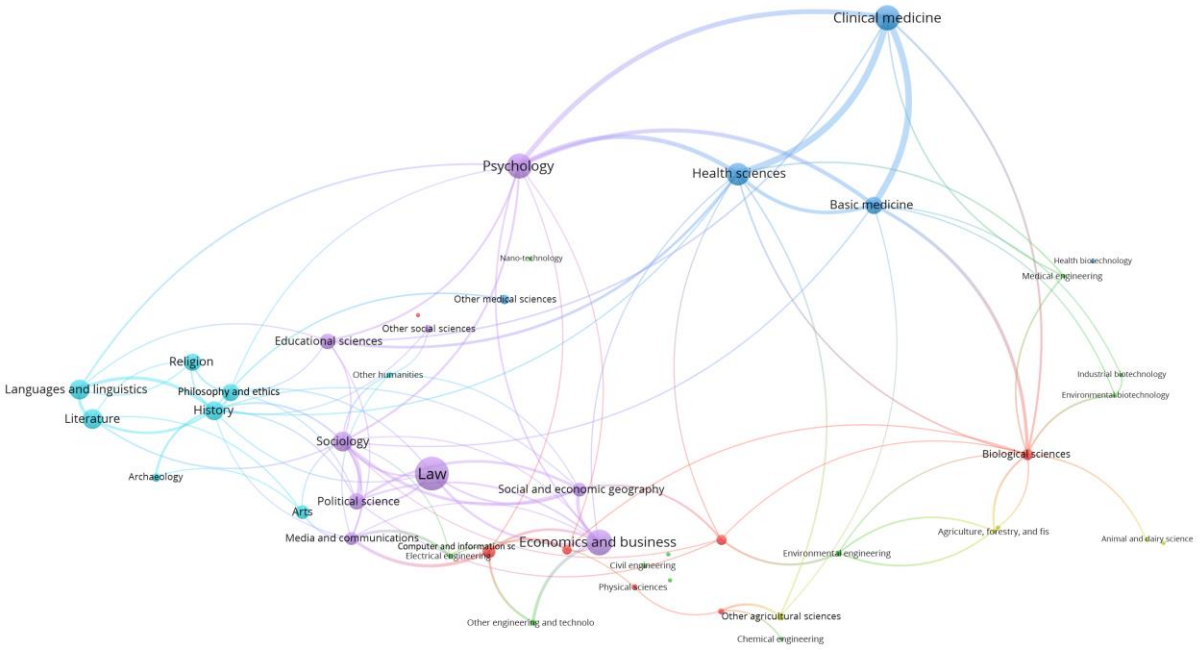

Figure 2. Map of co-occurrences among FOS fields in VABB-SHW. Only 100 most prominent links are shown. This map can be web-started from http://bit.ly/fos-coocc.

Figure 2 bears some resemblance to the overall structure of other bibliometric maps of science (Börner et al., 2012; Klavans & Boyack, 2009): it has a roughly circular shape, in which the SSH are linked on one side to medicine and on the other to computer science. At the same time, several peculiarities can be observed as well, such as the relative unimportance of most natural sciences. This is mainly due to the fact that the data set only contains publications by researchers affiliated to SSH units. This map is based on publications of different types, including book publications, which renders it different from most other maps of science; one exception is the work by Boyack and Klavans (2014), who explore the effect of including non-source items in a map based on citation relations. We further note that several top-level fields appear to have one central hub (*History* for the humanities, *Biological sciences* for the natural sciences) and that in our map environmental sciences also play a bridging role between social sciences and biological sciences. Finally, the prevalence of health science and medicine – and the strength of the links between them – is remarkable, given the SSH-centric nature of the data set.

Cross-tabulation of cognitive and organizational classification

First, we compare the share of publications from the six top-level FOS fields in the aggregated organizational fields of *Social sciences* and *Humanities*. This is summarized in Figure 3, in which only the 426 publications that were classified as multidisciplinary are not represented. In this figure as well as the following ones, the shade of a cell reflects its value. For both social sciences and humanities, we see that the majority of publications is published in the corresponding FOS field, although the share in humanities (73%) is considerably larger than in social sciences (59%). Moreover, publications authored by researchers from the humanities appear in social science channels in 15% of the cases, whereas this is only 5% for the reverse case. The top-level FOS field of *Medical and health sciences* is the second largest recipient of publications from scholars affiliated to social science units (23%).

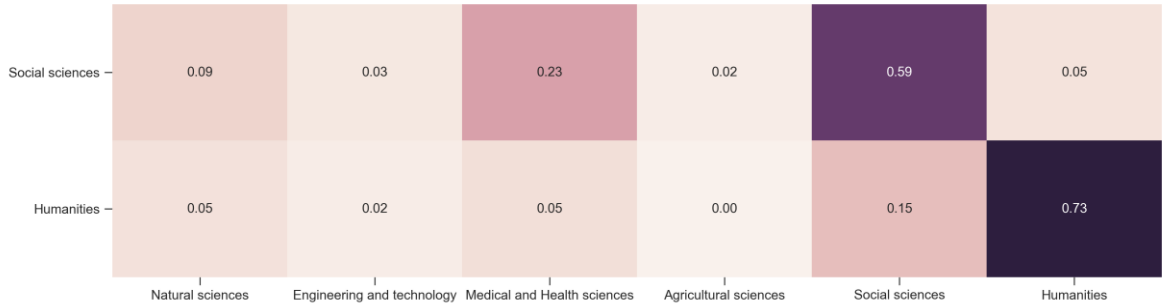

Figure 3. Heatmap of publications classified according to organizational unit (rows) and content (columns). Data are reported for top-level classifications and normalized per row.

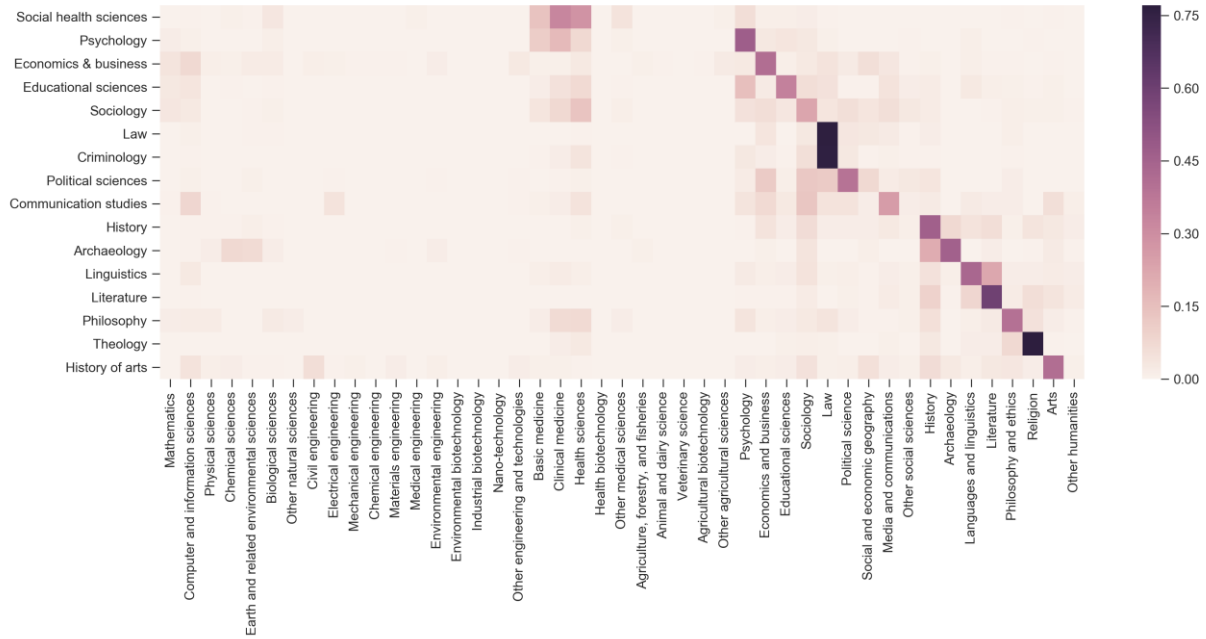

Figure 4. Heatmap of publications classified according to organizational unit (rows) and content (columns). Data are normalized per row. The underlying data table is available as supplementary material 1.

Figure 4 provides a more detailed view. In this figure the order of disciplines is based on the FOS classification (OECD, 2007, 2015); VABB disciplines mirror the order of the closest corresponding FOS fields. The rough 'diagonal' line in the right-hand side of the picture indicates that, overall, the organizational classification is at least an approximation of the intellectual content of publications. Gaps in the diagonal are explained by the fact that some FOS fields – *Social and economic geography*, *Other social sciences*, and *Other humanities* – have no direct counterpart in VABB disciplines. Because the FOS field of *Health sciences* is classified under *Medical and health sciences*, the first row (*Social health sciences*) does not show up as part of the diagonal. We also point out that some pairs of disciplines, like linguistics and literature or history and archaeology, are shown to be closely related. Because the 'general' VABB disciplines *Social sciences general* and *Humanities general* are far more heterogeneous than the others, they are not shown; the figure is, in other words, based on the 71,640 publications that could be assigned to at least one of the 16 VABB disciplines.

The intensity of cells in the diagonal indicates the extent to which a VABB discipline corresponds to a FOS field. The disciplines with the highest correspondence are *Theology* (77.1%), *Law* (76.5%), and *Criminology* (76.4%, mapped to FOS field *Law*). On the other hand it can be seen that some disciplines have a much weaker disciplinary profile, such as *Sociology* (23.1%), *Communication studies* (25.4%), and *Social health sciences* (28.5%).

Social health sciences is the only organizational discipline that has a higher share of publications in another FOS field than the one we considered to be the closest corresponding: 32.8% of its publications appear in *Clinical medicine* journals. Research from other VABB disciplines also appears regularly in medical journals, like *Psychology*, *Sociology* (sociology of health), and *Philosophy* (e.g., ethical issues).

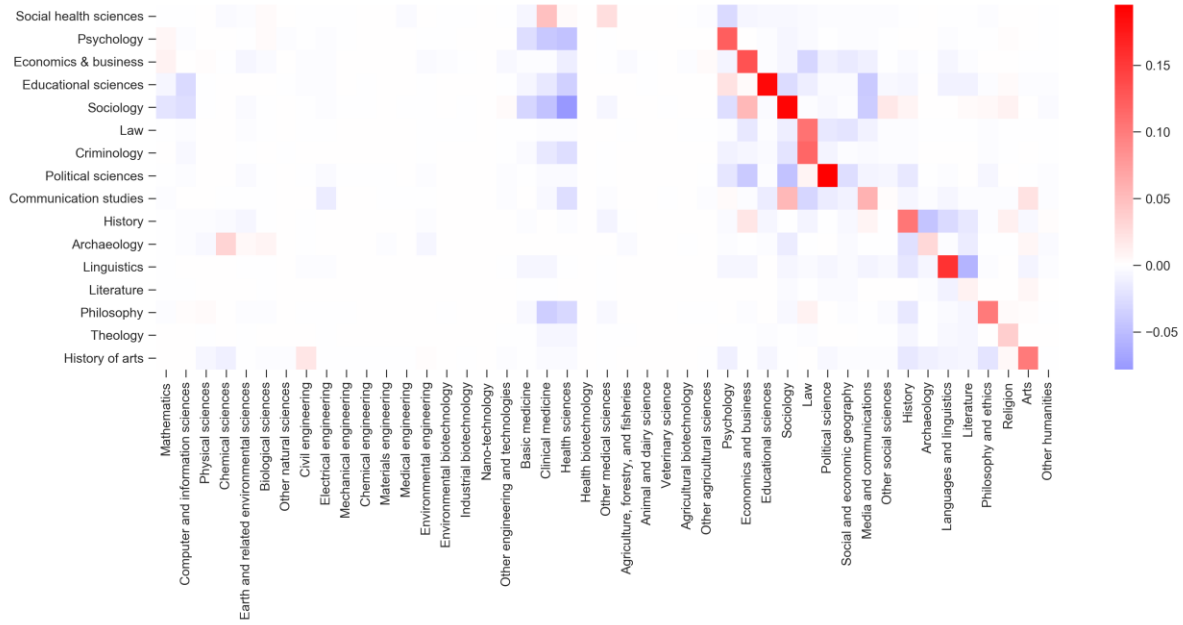

Figure 5. Difference in percentage points between shares based on single-assignment publications and shares based on all publications. The underlying data table is available as supplementary material 2.

A possible explanation for the non-matches between cognitive and organizational classification is the fact that a substantial group of publications are assigned to multiple VABB disciplines and/or multiple FOS fields. To study the influence of this factor we also calculate the share of publications from FOS fields in each VABB discipline, taking only single-assignment publications into account, i.e. publications that are assigned to one VABB discipline and one FOS field. At the aggregate level, we find relatively small changes if we only take single-assignment publications into account: 63.3% (resp. 78.7%) of those publications from VABB *Social sciences* (resp. *Humanities*) is published in the corresponding FOS field.

Figure 5 visualizes the difference between shares based on single-assignment publications and shares based on all publications that were counted for Figure 4. Red shades indicate that the percentage calculated using only single-assignment publications is higher than the one using all publications, whereas blue shades indicate that the percentage for single-assignment publications is lower. White means that no difference was found. In general, the share based on single-assignment publications increases for the FOS field that corresponds best to a given VABB discipline and decreases for most of the other FOS fields. The main exception is, again, *Social health sciences*, where the largest increase occurs for FOS field *Clinical medicine*. The strongest increase in percentage points is observed for *Political sciences* (+19.5 p.p.), while the strongest decrease is found for *Sociology*'s involvement in FOS field *Health sciences* (−7.8 p.p.). The increase for *Literature* is very small (+1.0 p.p.).

Concentration

These findings clearly indicate that some organizational disciplines are relatively concentrated in terms of in which cognitive fields articles are classified, whereas others are more spread out. This can be quantified using a concentration measure like the Gini index (Gini, 1921).

The results (Table 4) are related to, but not interchangeable with, the results regarding the match between FOS field and corresponding VABB discipline. While all disciplines are relatively concentrated over FOS fields, the difference between the most and the least concentrated ones – *Theology* and *Economics & business* respectively – is considerable. As one would expect, the Gini index that is based on single-assignment publications only is higher in all cases than the one based on all publications.

Overall, the *Humanities* (0.67) are slightly more concentrated than *Social sciences* (0.59). It may seem paradoxical that the aggregations of VABB disciplines are less concentrated than the disciplines themselves. This is, however, logical: given a VABB discipline, most FOS fields have zero or just a few publications. By aggregating, such 'holes' largely disappear.

Table 4. Gini index per VABB discipline, based on concentration in FOS fields

VABB discipline	Gini index (all publications)	Gini index (single-assignment publications only)
Theology	0.93	0.94
Criminology	0.93	0.96
Law	0.92	0.94
Literature	0.90	0.91
Psychology	0.88	0.90
Archaeology	0.87	0.89
Social health sciences	0.87	0.90
Political sciences	0.85	0.91
Linguistics	0.84	0.89
History	0.81	0.86
Educational sciences	0.80	0.87
Philosophy	0.76	0.80
Communication studies	0.75	0.81
Sociology	0.74	0.81
History of arts	0.74	0.81
Economics & business	0.72	0.79

Discussion

Summary of findings

This exploratory study highlights the ambiguity of the notion 'academic discipline' and classifications that are used to assign publications to disciplines. Figures 1 and 2 illustrate that the work in most academic disciplines within SSH is inter-linked (Figure 1) and has ties to knowledge domains outside the SSH (Figure 2). Figure 1 attributes a central position to philosophy. Indeed, philosophy may be quite different from other disciplines in the SSH, because basic philosophy investigates the claims made by other disciplines (Lamont, 2009). In the Flemish context, we also observe that philosophy is closely involved in addressing actual ethical issues such as those related to biomedical innovations. Figure 2 points to a circular shape in disciplinary ties similar to what has been shown by (Börner et al., 2012). Even though their method differs from the one we use here, we also find that social sciences are linked with health sciences (through psychology) and with computer and information science (through economics and business as well as social and economic geography). The order of the other non-SSH disciplines, follows the same sequence (clockwise): social sciences, computer and information science, physical sciences and chemistry (here, with rather minor representation), earth sciences, biological sciences, medicine, and back again to social sciences. As in the UCSD

map, the humanities appear to represent a rather distinct knowledge domain that is less connected with the other fields and mostly branches off from the social sciences. Contrary to the UCSD map, we find that environmental science, as an established interdisciplinary field, is linked to both biological sciences and social sciences.

On a more fine-grained level, we find that a substantial number of publications by researchers affiliated to a unit belonging to a particular academic discipline, turns out to be publications contributing knowledge to other knowledge domains (too). Sometimes this may be an indication of collaborations across disciplines, yet other cases seem to be knowledge pursuit across the boundaries made up by one's organizational affiliation (as shown by comparison that takes into account only single-assignment publications, Figure 5).

We also find differences between the humanities and the social sciences (Figure 3). Researchers with affiliation to humanities more often stay within the domain of humanities in their pursuit of knowledge. In social sciences, in contrast, publications are much more dispersed across different knowledge domains. Only 59% of publications by researchers in social sciences are associated with this field of research in terms of the content of publications. However, leaving out social health sciences (which are classified under social sciences in our organizational VABB classification but under *Medical and health sciences* in the cognitive FOS classification) changes this percentage to 68%, which is much closer to the 73% for the humanities. Furthermore, we identify differences across disciplines. For example, we find that some organizational disciplines tend to be very closely linked to a single field of research (e.g., *Theology*, *Criminology* and *Law*; see Figure 4 and Table 4). Others appear as more loosely tied to a single discipline (e.g., *Sociology*, *Communication studies*; see Figure 4).

Implications

The findings presented in this paper indicate that the choice of classification type (cognitive or organizational) plays a crucial role in representations of research activities. This has further implications for bibliometric analyses and science policy. Bibliometric analyses often rely, primarily out of pragmatic considerations, on classifications that are already used to assign records to academic disciplines. Given the discrepancies, stemming from the two classification types explored here, it is evident that using different classification types may lead to different conclusions. This situation may apply in analyses where data are drawn from different sources, e.g. in international comparisons of research output registered in national databases.

This raises the question which classification type to use in a given situation. Somewhat analogously to the choice of a bibliometric indicator, we believe that this largely depends on the purpose of the classification. For most purposes (e.g. retrieval, getting an overview of a given knowledge domain, normalization of bibliometric indicators) cognitive classifications are probably to be preferred. An organizational classification is an option in cases where such a cognitive classification is not easily available, and/or one wants to sidestep the inherent subjectivity of cognitive classification. Further research may clarify in which contexts administrative and qualification-based classifications are preferable.

Similar considerations apply in a science policy context, in particular, in research evaluation activities. Often research evaluation is carried out by knowledge domains identified on the basis of organizational units. As we show in this paper, publications affiliated to an organizational unit may in some cases be widely dispersed across different disciplines or even knowledge domains. Consequently, such cases in research evaluation, especially if pursued with quantitative means, may mischaracterize the contribution of a particular organizational unit to

the development of knowledge, as discussed in the literature on evaluation of interdisciplinary research more broadly (e.g., Huutoniemi & Rafols, 2016).

On a more conceptual level, this study raises further questions on the legitimacy of disciplinary classifications, which also can be seen as a limitation of this study. In this study we used the OECD FOS classification, which is a rather crude classification considering the SSH. While there are, for instance, 75 WoS categories for SSH, the FOS classification distinguishes only 14 SSH fields (refined to 17 in our study). Similarly, we are aware that classifications carry different social and cultural considerations (e.g., Bowker & Star, 2000). Consequently, it is a matter of debate whether some of the choices in classification we have made following OECD FOS guidelines adequately represent SSH in Flanders. For example, one may argue that it is not accurate to equate Anthropology with Sociology as it occurs here. As explained earlier, the choice in favour of FOS was for reasons of feasibility and comparability. Yet, the usage of an international standard classification scheme also holds limitations, e.g. in the sense that it is not geared towards the Flemish context and that it may not be that responsive to new developments. In this context it is worth nothing that within the Flemish Centre for R&D Monitoring, the ECOOM-Hasselt branch has developed a semantically documented research discipline list based on the OECD FOS classification. For some disciplines outside the SSH, the structure of the second level in this list has been updated according to the current research performed in Flanders. Furthermore, two additional levels of classification are distinguished and some fields (e.g. architecture) are positioned as crossing the boundaries of the six broad domains at the top level of the original OECD FOS classification (Poelmans & Vancauwenbergh, 2017).

The results presented in this study may also guide future work in social studies of science. One can interpret the overlap between VABB-SHW's organizational classification with FOS as the degree to which scholars affiliated to organizational units (like departments or research groups) also conduct and publish research in these domains. Put differently: to what extent do scholars affiliated to an organizational unit perform cognitive work in the same scholarly area? Our results tend to confirm the theorizing on 'new disciplinarity' by authors like Marcovich and Shinn (2011, 2012). These authors stand in stark contrast with assumptions raised by postmodernist theorists who claim that traditional disciplines are disappearing, and hold that disciplines will remain the main referents. The analysis shows a rather stable but elastic map for the SSH, wherein each discipline has its own peculiarities.

The dispersed overlapping of sociology (both as a VABB and a FOS category) with other fields, for example, could be studied in light of Abbott's prognoses that sociology shares topics with many other cognitive fields (Abbott, 2001, p. 6). Moreover, sociology seems to have a strong connection with publications classified as *Other social sciences*. This class is devoted to interdisciplinary publications and 'other social sciences'. The latter can be understood as emerging or 'peculiar' fields within the social sciences. Marcovich and Shinn (2012) conclude that whereas the system of knowledge production is becoming more complex and dense than ever before, and new fields of knowledge production are emerging at a painstakingly high rate, the main referents will not as quickly cease to exist (p. 601). Further socio-historical contextualisation of particular cases in the SSH, however, is necessary to verify their claims.

Although quite crude, the expansion of the OECD Fields of Science coding scheme can be considered as evidence for the relative persistence of 'traditional' disciplines, the increasing complexity of scientific knowledge, and the spawning of new disciplines. The first edition of the list was published in 2002, but became subject to change rather shortly after (2007). Due to rapid changes with regard to fields in science and technology, for example, the team working on the Frascati Manual soon decided that it was appropriate to further expand the list (OECD,

2007). The elasticity and dynamic nature of knowledge domains is rather explicitly acknowledged, as an 'other' subclass for each top-level field was introduced. This was done '… to keep the classification flexible and allow for appearance of new areas of studies' (ibid., p. 5). Further explorations of publications classified as 'other', or assigned to multiple categories might provide insights in these future horizons, and shed light on interdisciplinary working grounds and projects as well. There remains a great deal of potential, not only in clarifying the differences between classification schemes, but also in analysing the different results these exercises yield, and putting them in context. This can be done by using more qualitative techniques in tandem with quantitative methods like the one deployed for this study or the unsupervised learning approach by Suominen and Toivanen (2016).

Conclusions

Different types of classification do not always yield the same results. In this paper, we have examined the discrepancy between cognitive and organizational classification of publications in social sciences and humanities, using the example of the Flemish VABB-SHW database. Our results show that all organizational disciplines are to some extent prone to publishing 'outside' one's own discipline but there are great differences between disciplines: while 77% of theology publications end up in publication channels belonging to theology, this is only 23% for sociology. At a higher aggregation level, we find that 73% of humanities publications are published in humanities journals, while this is only the case for 59% of social sciences publications (68% without social health sciences). Social sciences are shown to be closely linked to medicine and health sciences; this is partially due to the inclusion of *Social health sciences* as a social science in the organizational classification but also to other social sciences like sociology and psychology. The spread of publications over cognitive fields, including non-SSH fields, can be partially explained by the assignment of some journals to multiple FOS fields as well as multidisciplinary collaboration: leaving out such 'multi-assignment' publications, we find an average increase of 10.4 percentage points in correspondence between VABB discipline and matching FOS field. However, although e.g. collaboration may strengthen the occurrence of research beyond the boundaries of one's (organizational) discipline, it is not a necessary condition. Indeed, our results show that it is quite common for researchers affiliated to an organizational unit to also contribute to the advancement of knowledge in other domains.

In future research, we intend to expand the cognitive classification of publications presented in this paper with an automated approach (cf. Suominen & Toivanen, 2016) whereby the metadata of publications in the VABB-SHW are matched to the refined version of the OECD FOS classification that has been developed by ECOOM-Hasselt (Poelmans & Vancauwenbergh, 2017). Future work could also focus on more qualitative approaches that further contextualize cases of specific (inter)disciplines and subdisciplines in the SSH. Finally, our results have highlighted the variability across classifications and types of classifications. It would be interesting to further explore this aspect taking also administrative and/or qualification-based classifications into account.

Acknowledgments

The present study is an extended version of an article presented at the 16th International Conference on Scientometrics and Informetrics, Wuhan (China), 16 - 20 October 2017). This investigation has been made possible by the financial support of the Flemish government to ECOOM. We thank Emanuel Kulczycki and Sadia Vancauwenbergh for feedback on specific issues.

References

Abbott, A. (2001). *Chaos of Disciplines*. Chicago: University Of Chicago Press.

Bod, R. (2012). *Het Einde van de Geesteswetenschappen 1.0* (pp. 1–14).

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

Bowker, G. C., & Star, S. L. (2000). *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Boyack, K. W., & Klavans, R. (2014). Including cited non-source items in a large-scale map of science: What difference does it make? *Journal of Informetrics*, *8*(3), 569–580. https://doi.org/10.1016/j.joi.2014.04.001

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

Frickel, S., & Ilhan, A. O. (2017). Disciplinary and interdisciplinary change in six social sciences: A longitudinal comparison. In S. Frickel, M. Albert, & B. Prainsack (Eds.), *Investigating interdisciplinary collaboration: Theory and practice across disciplines* (pp. 148–169). New Brunswick, NJ: Rutgers University Press.

Gini, C. (1921). Measurement of inequality of incomes. *The Economic Journal*, *31*(121), 124–126.

Guns, R., Engels, T. C. E., & Verleysen, F. T. (2017). Cognitive and organizational classification of publications in the Social Sciences and Humanities. In *Proceedings of the 16th International Conference on Scientometrics and Informetrics* (pp. 112–117). China: Wuhan University.

Hay, C. (2010). The Changing Nature of European Political Science: The Discipline in an Age of Acknowledged Interdependence. *European Political Science*, *9*(1), S121–S131. https://doi.org/10.1057/eps.2010.41

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: The use of publication and patent statistics in studies of S&T systems* (pp. 473–496). Dordrecht: Kluwer Academic.

Huutoniemi, K., & Rafols, I. (2016). *Interdisciplinarity in Research Evaluation* (SSRN Scholarly Paper No. ID 2818321). Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=2818321

Klavans, R., & Boyack, K. W. (2009). Toward a consensus map of science. *Journal of the Association for Information Science and Technology*, *60*(3), 455–476. https://doi.org/10.1002/asi.20991

Kulczycki, E., Engels, T. C. E., & Nowotniak, R. (2017). Publication patterns in the social sciences and humanities in Flanders and Poland. In *Proceedings of the 16th International Conference on Scientometrics and Informetrics* (pp. 95–104). Wuhan.

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

Lamont, M. (2009). *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge, MA: Harvard University Press.

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. https://doi.org/10.1002/asi.23408

Marcovich, A., & Shinn, T. (2011). Where is disciplinarity going? Meeting on the borderland. *Social Science Information*, *50*(3–4), 582–606. https://doi.org/10.1177/0539018411411036

Marcovich, A., & Shinn, T. (2012). Regimes of science production and diffusion: towards a transverse organization of knowledge. *Scientiae Studia*, *10*(SPE), 33–64. https://doi.org/10.1590/S1678-31662012000500003

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the Social Sciences and the Humanities: A Review. *Scientometrics*, *66*(1), 81–100. https://doi.org/10.1007/s11192-006-0007-2

Organisation for Economic Co-operation and Development. (2007). *Revised field of science and technology (FOS) classification in the Frascati manual*. Paris: Working Party of National Experts on Science and Technology Indicators, Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development. (2015). *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development*. Paris: OECD Publishing. Retrieved from https://doi.org/10.1787/9789264239012-en

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.

Pedersen, D. B. (2016). Integrating social sciences and humanities in interdisciplinary research. *Palgrave Communications*, *2*, 16036. https://doi.org/10.1057/palcomms.2016.36

Poelmans, H., & Vancauwenbergh, S. (2017). Development of the Flemish research discipline list: state of affairs. Presented at the ECOOM Conference, Hasselt.

Pölönen, J., Engels, T. C. E., Guns, R., Sivertsen, G., & Verleysen, F. T. (2017). SSH journal publishing in Flanders and Finland. Presented at the 2nd Conference on Research Evaluation in the Social Sciences and Humanities, Antwerp, Belgium.

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. https://doi.org/10.1002/asi.10153

Rousseau, R. (2011). Lorenz curves determine partial orders for comparing network structures. *DESIDOC Journal of Library & Information Technology*, *31*(5), 340–347.

Suominen, A., & Toivanen, H. (2016). Map of science with topic modeling: Comparison of unsupervised learning and human-assigned subject classification. *Journal of the Association for Information Science and Technology*, *67*(10), 2464–2476. https://doi.org/10.1002/asi.23596

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

van Leeuwen, T. (2013). Bibliometric research evaluations, Web of Science and the Social Sciences and Humanities: a problematic relationship? *Bibliometrie-Praxis Und Forschung*, *2*, 1–18. https://doi.org/10.5283/bpf.173

van Rijnsoever, F. J., & Hessels, L. K. (2011). Factors associated with disciplinary and interdisciplinary research collaboration. *Research Policy*, *40*(3), 463–472. https://doi.org/10.1016/j.respol.2010.11.001

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. 117–127). London: Portland Press. Retrieved from http://www.portlandpress.com/pp/books/online/wg87/087/0117/0870117.pdf

Verleysen, F. T., & Weeren, A. (2016a). Clustering by publication patterns of senior authors in the social sciences and humanities. *Journal of Informetrics*, *10*, 254–272. https://doi.org/10.1016/j.joi.2016.01.004

Verleysen, F. T., & Weeren, A. (2016b). Mapping diversity of publication patterns in the social sciences and humanities. An approach making use of fuzzy cluster analysis. *Journal of Data and Information Science*, *3*(1), 1–27.

Whitley, R. (2000). *The intellectual and social organization of the sciences*. Oxford: Oxford University Press.