Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 146 (2019) 265–278

Procedia Computer Science

www.elsevier.com/locate/procedia

14th International Conference on Current Research Information Systems, CRIS2018

The creation of the Flemish research discipline list, an important step forward in harmonising research information (systems)

Sadia Vancauwenbergh[a,b]1, Hanne Poelmans[a,b]

[a]*Centre for R&D Monitoring, Division ECOOM-Hasselt, Martelarenlaan 42, BE-3500 Hasselt (Belgium)*
[b] *Hasselt University, Research Coordination Office, Martelarenlaan 42, BE-3500 Hasselt (Belgium)*

Abstract

In 2011, Peters et al. [1] wrote a report on the administrative simplification of research reporting in Flanders. Next to the description of several data flows that could be harmonised, validated and merged into CRIS systems, the report also contained important recommendations on the use of common standards and classifications that could lead to more efficient and qualitative research information (systems) and thus more efficient research reporting. As such, in Flanders several classifications are used in research reporting. In the context of reporting on research output and expertise, data is clustered based on research disciplines. Currently, a variety of research discipline classification lists are utilised in Flanders, each depending on the authority to whom a researcher is obliged to report. Because the existence of several research discipline lists for a region as small as Flanders is redundant and inefficient, one of the most important recommendations of the report by Peters et al. [1] was the creation of a single research discipline code list that could be used to tag information inside CRIS systems for reporting, dissemination and visualisation purposes. The uniform definition and management of the semantics of this research discipline code list for all research actors in Flanders could have an immediate impact on the accuracy of reporting and the policy pursued on the basis thereof. In addition, such managed classification systems can be used in dynamic research information systems that drastically reduce the administrative burden of the research population, which automatically entails an important investment in research and innovation.

Peer-review under responsibility of the scientific committee of the 14th International Conference on Current Research Information Systems, CRIS2018.

Keywords: Semantics; data and classification governance; research discipline; research information (systems)

1. Introduction

Researchers and research institutions worldwide have to report on their research activities throughout the entire

* Corresponding author. Tel.: +3211269109.
E-mail address: sadia.vancauwenbergh@uhasselt.be

1877-0509
Peer-review under responsibility of the scientific committee of the 14th International Conference on Current Research Information Systems, CRIS2018.
10.1016/j.procs.2019.01.075

research lifecycle to governments, financers and third parties for granting and evaluation purposes. These reports describe the same research activities but vary with regards to the required format and the requested content, depending on the stakeholders' context and goals. Moreover, next to the description of the research itself, these reports often contain a set of research output measures, such as e.g. publications, projects, financial data, patents, spin-offs, staff, etc., that quantify the research performance of a (group of) researcher(s).

At first glance, these figures seem to provide an easily comprehensible and straightforward means to evaluate research performance. However, research performance as described in these reports should be processed, evaluated and compared with caution. First, the research output measures reported by researchers from different institutions are hardly ever gathered from a single, central information system but are collected from institutional databases. These databases are designed or adopted based on the requirements of the research institution itself, causing variability in which information is available. Second, the semantics underlying the registered data is often overlooked by the research community, both when gathering the information as well as during research evaluation (Fig. 1). That is, the terms used in the institutional databases are often not explicitly defined and the registration of information occurs according to informal rules and definitions used by the data administrators themselves, thereby creating an informal semantic layer. If information is then exchanged between systems, without explicitly consulting the underlying semantics, different data can be combined erroneously leading to inconclusive data. The problem aggravates when indirect information consumers, i.e. evaluators that make use of metrics without having an in-depth knowledge of the research environment, evaluate the resulting metrics without consulting the underlying (in)formal semantics, if already present, and take the numbers for granted. Depending on the evaluator's knowledge on the concepts and indicators used in the metrics, the problem can become even worse.

In an optimal scenario, report writers are perfectly aware of how their institutional databases are designed, how database administrators register information in these databases, and how these data are used in metrics and reported on to different stakeholders. In the context of the research evaluation, direct evaluators, i.e. evaluators that are familiar with the data and metrics as used by the individual institutions, play a crucial role in interpreting the reported metrics and informing the indirect information consumers about how the reported information can and cannot be used. In addition, direct evaluators play a decisive role in aggregating information from different data providers to a level transcending the individual institutions. That is, information from different data providers can only be aggregated in a meaningful manner when all data elements contain exactly the same information and thus relay on the same underlying semantics.

Fig 1. Use of research classifications by information providers and consumers according to contextual semantics.

In 2011, the Flemish Government (Department of Economy, Science and Innovation) requested the Flemish universities and higher education institutions in collaboration with the Flemish Interuniversity Council (VLIR), to analyse the quality and efficiency of research reporting in Flanders. This study, published in the report on the

'Simplification of Research Reporting' [1], pointed out that one of the main factors derogating the quality and efficiency of research reporting in Flanders was the use of a different classification system to register research information for every information requester in Flanders (i.e. the Flemish government, funders and other institutional research administrations). One of the most important recommendations of this study was the development and implementation of a Flemish research classification standard, that is semantically enriched and serves all different use purposes in research.

2. Research classifications

Research classifications are mostly hierarchical systems that are used to cluster research information into predetermined categories. Although a wide variety of research classifications exists, they can basically be classified into 3 categories describing either the identity of a researcher, research input or research output. The first category contains classifications that are generally used for describing the characteristics of the research population in terms of age, nationality, gender, professional activities (researcher, technical and supporting staff), or research discipline to which a researcher belongs. Research input is often covered by classifications that describe the origin of financial resources provided to research (i.e. financer, geographical region, …). The largest variety of classifications, however, can be found in the context of research output where almost every piece of output is captured in a classification, mainly for monitoring and evaluation purposes. These classifications are typically multidimensional in that they cover a variety of output ranging from publications, patents to even artistic design and simultaneously try to depict the innovative nature of the findings, their corresponding technological sector,... Unfortunately, over the years many typologically similar classifications have evolved at different organisational levels, ranging from single institutions, to regional and (inter)- and (supra)national governments, and developed according to the specific operational needs of each organisation(al) entity [2]. Obviously the terms used in these classifications have a specific, connotative meaning as intended by its developers in relation to the goals set out. Yet, hardly any classification contains these explicit semantic descriptions thereby opening the door for interpretative use by the different stakeholders ranging from researchers, to managers, policy analysts, social analysts and R&D executives. Moreover, different interpretations can also be given when classifications are used in different contexts and environments. Not surprisingly, the lack of semantic descriptions can result in inconclusive data.

This paper focusses on the development of the Flemish research discipline list, i.e. a single, harmonised and semantically enriched classification system for registering and reporting on research in specific disciplines.

2.1 The Flemish landscape of research disciplines lists

In 2014, Flemish research administrations use 4 different research discipline lists, i.e. 1) the revised OECD Fields of Science and Technology Classification list (FOS) as published in the Frascati Manual (annex to OECD [3]), 2) the former Inventory of the Scientific and Technological Research in Flanders' list (IWETO/FRIS), 3) the Fund for Scientific Research Flanders' list (FWO) and 4) the Flemish Interuniversity Council science discipline list (VLIR).

In brief, the FOS classification list has been drafted under the supervision of the Working Party of National Experts on Science and Technology Indicators (NESTI). In 2006, the list was revised by an OECD Task Force due to the dynamics in the research field as well as critiques from the research community. The FOS classification list consists of 2 hierarchical levels, of which the first level comprises 6 major disciplines, followed by 40 subdisciplines on the second level each described by a limited number of keywords. The list was originally drafted in English and is used in Flanders for reporting on research staff in the frame of the OESO O&O questionnaire to the European government. In 2015, the revised FOS classification list was updated and renamed into the Fields of Research and Development (FORD) classification list [4].

In the early nineties, the IWETO/FRIS science discipline list[2] was created in agreement with international trends by the Fund for Scientific Research Flanders. Later on, science disciplines have been added to the list on an *ad hoc* basis by the Flemish government and its name changed to the Flanders Research Information Space (FRIS) science disciplines list. Currently, the list consists of 2 hierarchical levels of respectively 5 and 389 codes, each reflecting

[2] Available via https://www.ewi-vlaanderen.be/sites/default/files/iwdisciplinecodes.pdf

research disciplines to a different level of granularity. The IWETO/FRIS list was originally composed in Dutch, and has been translated to English afterwards. The list is used by research administrators for reporting on researchers and research projects to the Flemish government.

The Fund for Scientific Research (FWO) scientific disciplines code list[3] originates from the same list, as the IWETO/FRIS science discipline list, but was updated separately, based on *ad hoc* requests submitted to the FWO. In October 2010, the list underwent a profound update under supervision of a scientific discipline committee appointed by the FWO. This resulted in a list consisting of 3 hierarchical levels, of respectively 5, 41 and 1.029 codes. The third hierarchical level was sometimes even further categorised. The list was written in Dutch and was later on translated to English. This disciplines list is used by the FWO for monitoring and reporting purposes as well as for finding experts for evaluating research proposals.

The VLIR science discipline list[4] is based on the scientific disciplines described in the Flemish decree on the annual reporting obligations of the Flemish universities. Three disciplines (historical sciences/art sciences, law/criminology, and medicines/dentistry) were deduplicated in the VLIR code list and two codes were added: 'other technical sciences' and 'general and logistic services', resulting in a total of 34 scientific disciplines. Despite its name, the list is more oriented towards education. The VLIR discipline list is only available in Dutch and is used for reporting on scientific personnel to the Flemish government. The list corresponds to the IWETO/FRIS science domains list, that is not to be confused with the IWETO/FRIS science disciplines list.

3. Towards a draft of the Flemish research discipline list

Based on the report by Peters et al. [1], the Flemish government assigned a project to the Expert Centre for Research & Development Monitoring (ECOOM), termed 'Classification Management in Data Governance Centre' with the intent to harmonise Flemish research-related classifications and to provide a semantic description for the terms used.

3.1 Information collection

As a first step, we collected information on existing research discipline classifications used for research reporting both in Flanders and beyond. We specifically focussed on classifications that cover the full spectrum of research and gave special attention to the inclusion of lists used at the regional, national, European and international level, with at a minimum of one list per continent. This was done in order to get a full overview of all research disciplines regardless of their geographical origin [5]. Classification lists that are used for bibliometric purposes, e.g. Clarivate's Web of Science and Elsevier's Scopus were omitted in this process because these classification lists are developed based on the content of their respective databases and in the context of calculating and normalising citation behaviour of publications within certain disciplines. This means that the coverage as well as the granularity of the disciplines in these classification lists is influenced by differences in the publication and citation behaviour between disciplines. That is, disciplines that typically produce output in unwritten forms (e.g. the arts) or in non-English languages are underrepresented in these classification lists, whereas broad disciplines such as physics are represented by many subfields.

In total, we considered 12 research discipline lists next to the 3 Flemish lists mentioned above (Table 1). All lists were analysed with regards to the hierarchical structure used and the level of detail contained. In total, 3 lists did not use any form of hierarchy, with a varying degree of terms used (ranging from 5 to 67 terms). The majority of lists made use of a hierarchical structure. Out of these, 6 lists used a 2-level structure, while the remaining 5 lists used a 3-level hierarchy. One list used a 4-level hierarchy. In general, the third level added more granularity to the levels above, albeit to a different extent when comparing all research discipline lists. Furthermore, the level of detail was sometimes even different within a hierarchical level of a single list. This can be explained on the one hand by the problem of defining research disciplines, as they contain various perspectives such as philosophical, anthropological, sociological, historical and managerial aspects that should be handled. On the other hand this can be explained by the implications

[3] Available via http://www.fwo.be/media/236854/Disciplinecodes-ENG.pdf
[4] Available on request

that distinctions in research disciplines have on higher education research, policy and practice [6].

Table 1. Overview of research discipline lists used with indication of the geographical region and the number of hierarchical levels.

Name	*Geographical region*	*# Hierarchical levels*
ANZSRC - FOR[5]	Australia & New Zealand	3
CASRAI[4]	America	3
CSC[6]	China	2
ERC[7]	Europe	2
OECD[8]	Europe	2
FAPESP[9]	Brazil	2
RSA-RF[4]	Africa	2
ANVUR[10]	Italy	3
SSD[11]	Italy	4
ÖFOS[12]	Austria	3
RAE[13]	UK	1
REF[14]	UK	1
NWO[15]	The Netherlands	2
FWO[16]	Flanders	4
FRIS	Flanders	3
VLIR	Flanders	1

3.2 Architecture of the harmonised Flemish research discipline list

Next, we analysed the aims for which the research classifications were originally designed as these largely determine the degree of detail, and thus hierarchy needed for the new harmonised Flemish research discipline list. In brief, we found that research discipline lists used for reporting to governments generally use 1 or 2 hierarchical levels, while discipline lists used for reporting to scientific experts in the field generally tend to use more hierarchical levels. As the new, harmonised Flemish research discipline list has to serve both audiences, more hierarchical levels seemed to be desirable. The specific amount of levels was determined by looking into the concrete use purposes of the new harmonised research discipline list in Flanders (Fig. 2). In agreement with the report on reduction of research reporting [1], the new list would be used for reporting to the European government (OESO O&O Questionnaire) and the Flemish government (R&D personnel) in order to fulfil legal reporting obligations. In addition, the list would be used to disseminate on Flanders's academic research potential to companies and research organisations with the intent to stimulate the creation of Flanders as an innovation hub as well as to tax payers as a means to inform then on the money spent by the government. Finally, the list would also be used by Flemish funders in order to organise the evaluation of research proposals by experts in the field. In agreement with the amount of detail needed, it was decided to create 4 hierarchical levels.

---

[5] Available via https://www.arc.gov.au/grants/grant-application/classification-codes-rfcd-seo-and-anzsic-codes or http://www.abs.gov.au/Ausstats/abs@.nsf/Latestproducts/4AE1B46AE2048A28CA25741800044242?opendocument
[6] Available via http://old.moe.gov.cn//publicfiles/business/htmlfiles/moe/moe_834/201104/116439.html
[7] Available via https://www.agreenskills.eu/Media/files/Disciplines-Subdisciplines-ERC or https://erc.europa.eu/sites/default/files/document/file/erc%20peer%20review%20evaluation%20panels.pdf
[8] Available via https://read.oecd-ilibrary.org/science-and-technology/frascati-manual-2015_9789264239012-en#page61
[9] Available via http://ipscience-help.thomsonreuters.com/inCites2Live/filterValuesGroup/researchAreaSchema/fapespBrazil/fapespSchema.html
[10] Available via https://www.cun.it/documentazione/academic-fields-and-disciplines-list/
[11] Available via http://www.miur.it/UserFiles/115.htm
[12] Available via http://bartoc.org/de/node/1094
[13] Available via http://www.rae.ac.uk/aboutus/uoa.asp
[14] Available via https://www.ref.ac.uk/2014/panels/unitsofassessment/
[15] Available via https://www.nwo.nl/en/funding/funding+process+explained/research+fields
[16] Available via http://www.fwo.be/media/236854/Disciplinecodes-ENG.pdf

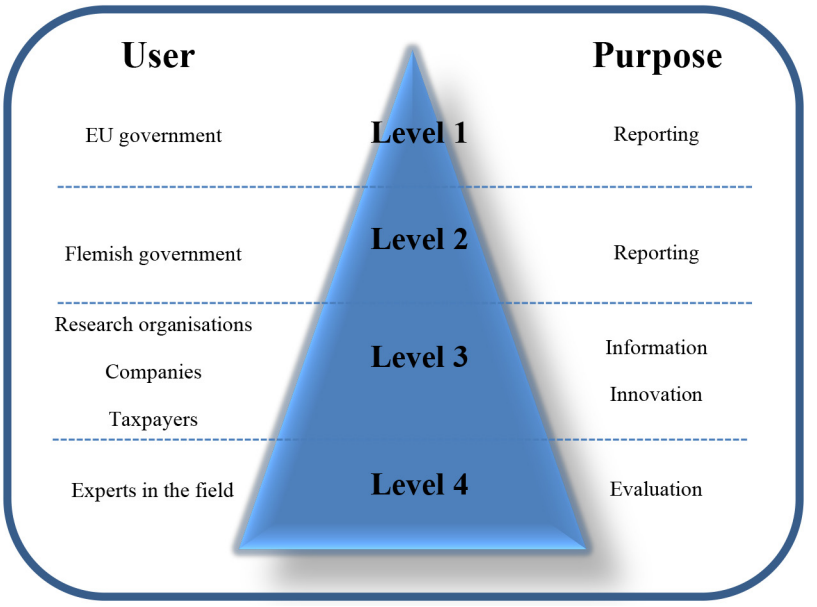

Fig 2. Use purposes of the new, harmonised research disciplines list

3.3 Drafting the content of the new, harmonised Flemish research discipline list

To populate the research discipline list with terms representing research disciplines, we started with analysing the terms used on the first hierarchical level of all discipline lists in relation to the Fields of Research and Development (FORD) classification list of OECD [4] as this list is used for reporting purposes to the European government. As not too many differences were observed, we proposed to adopt the first level of the OECD Fields of Research and Development list [4] completely, as this would facilitate the mapping of research disciplines, and thus reduce the administrative burden on the research community later on. As such, the first level of the Flemish research discipline list consists of 6 research disciplines, i.e. natural sciences, engineering and technology, medical & health sciences, agricultural and veterinary sciences, social sciences and humanities and the arts. Furthermore, a seventh code was added in order to ensure a continuity in research reporting from the previous VLIR science discipline codes to the new Flemish research discipline list. The second level was largely based on the FORD classification list [4], and the 2 remaining hierarchical levels were prefilled based on the FWO science discipline list, as this was considered to be the most complete Flemish research discipline list at the time. However, comparison with the other 15 research discipline lists resulted in the identification of several new terms. These terms either reflected new and emerging research disciplines, or sometimes seemingly denoted alternative terms for already included research disciplines. Furthermore archaic terms were seen as well as terms that did not correspond to research disciplines. All these new terms were colour-marked and added to the draft list.

Next, a semantic definition had to be defined for each recognised research discipline. As a starting point we populated the draft research discipline list with definitions from the Scope Notes from the Web of Science (Thomson Reuters)[17], professional literature and Wikipedia.

3.4 Inclusion of a detection method for emerging research disciplines

Next, we evaluated practices and methods used in the past to keep research discipline lists in line with new and emerging research disciplines. In Flanders (and in most places), it has been a general practice to address these

[17] http://mjl.clarivate.com/scope/

shortcomings to the agency that 'owns' the research discipline list. These agencies mostly consult a group of expert scientists, who advise, based upon their knowledge and expertise, whether the new term can be considered to be included in the research discipline list. The ANZSRC-FOR research discipline list however, uses a semi-automated strategy to detect new emerging research disciplines. Moreover, the second and third level of the ANZSRC-FOR list contain respectively so-called 'other xx sciences' and 'xx sciences, not elsewhere classified' categories. When the amount of projects within these categories in relation to the total amount of projects within the same level crosses a threshold of 5%, these categories are marked as possibly containing new research fields. These categories are then passed to expert groups who decide upon their possible inclusion. This automatic detection method has been working quite well for the ANZSRC-FOR list, as the list can be considered among one of the most up-to-date research discipline list that we analysed. We therefore decided to include a similar semi-automated detection method in the harmonised Flemish research discipline list and consider the projects sent by the Flemish universities to the Flanders Research Information Space (FRIS)-portal as a basis for our calculations. As not yet all research performing organisations (higher education and other research institutions) provide information to FRIS, we reasoned that the inclusion of a manual registration would ensure the coverage of all research disciplines.

	Sector (Level 1)		Disciplinary Field L1		Disciplinary Subfield L2		Disciplinary Subfield L3
06	Humanities and the arts	0604	Arts	060401	Architectural design		
06	Humanities and the arts	0604	Arts	060401	Architectural design	06040101	Design innovation
06	Humanities and the arts	0604	Arts	060401	Architectural design	06040102	Design management
06	Humanities and the arts	0604	Arts	060401	Architectural design	06040103	Design practice
06	Humanities and the arts	0604	Arts	060401	Architectural design	06040104	Design research
06	Humanities and the arts	0604	Arts	060401	Architectural design	06040105	Digital and interaction design
06	Humanities and the arts	0604	Arts	060401	Architectural design	06040106	Inclusive design
06	Humanities and the arts	0604	Arts	060401	Architectural design	06040107	Sustainable design
06	Humanities and the arts	0604	Arts	060401	Architectural design	06040199	Architectural design not elsewhere classified
06	Humanities and the arts	0604	Arts	060402	Architecture		
06	Humanities and the arts	0604	Arts	060402	Architecture	06040201	Architectural heritage and conservation
06	Humanities and the arts	0604	Arts	060402	Architecture	06040202	Architectural practice
06	Humanities and the arts	0604	Arts	060402	Architecture	06040203	Architecture management
06	Humanities and the arts	0604	Arts	060402	Architecture	06040299	Architecture not elsewhere classified
06	Humanities and the arts	0604	Arts	060403	Art studies and sciences		
06	Humanities and the arts	0604	Arts	060403	Art studies and sciences	06040301	Architectural history and theory
06	Humanities and the arts	0604	Arts	060403	Art studies and sciences	06040302	Criticism and theory
06	Humanities and the arts	0604	Arts	060403	Art studies and sciences	06040303	Curatorship
06	Humanities and the arts	0604	Arts	060403	Art studies and sciences	06040304	Architectural design history and theory
06	Humanities and the arts	0604	Arts	060403	Art studies and sciences	06040305	Film studies
06	Humanities and the arts	0604	Arts	060403	Art studies and sciences	06040306	History of art
06	Humanities and the arts	0604	Arts	060403	Art studies and sciences	06040307	History of music
06	Humanities and the arts	0604	Arts	060403	Art studies and sciences	06040308	History of performing arts
06	Humanities and the arts	0604	Arts	060403	Art studies and sciences	06040309	History of stage craft
06	Humanities and the arts	0604	Arts	060403	Art studies and sciences	06040310	Iconology
06	Humanities and the arts	0604	Arts	060403	Art studies and sciences	06040311	Interior architecture history and theory
06	Humanities and the arts	0604	Arts	060403	Art studies and sciences	06040312	Landscape architecture history and theory
06	Humanities and the arts	0604	Arts	060403	Art studies and sciences	06040313	Musicology and ethnomusicology
06	Humanities and the arts	0604	Arts	060403	Art studies and sciences	06040314	Performance studies
06	Humanities and the arts	0604	Arts	060403	Art studies and sciences	06040315	Theatre science
06	Humanities and the arts	0604	Arts	060403	Art studies and sciences	06040316	Visual cultures
06	Humanities and the arts	0604	Arts	060403	Art studies and sciences	06040399	Art studies and sciences not elsewhere classified

Fig 3. Extract of the Flemish Research Discipline List version 2018

Altogether, this resulted in 4-hierarchical draft list or respectively 7, 42, 382 and 2.493 terms, reflecting research disciplines to a different level of granularity (Fig. 3). This list served as the starting point on which the data and classification governance method was imposed, in order to create a semantically enriched, harmonised Flemish research discipline list.

4. Towards a semantically enriched Flemish classification standard for research disciplines

4.1 Classification governance as methodology

Classification governance (CG) basically stems from data and information governance [7]. Likewise, it comprises the specification of decision rights and an accountability framework that encourages desirable behaviour in the creation, storage, use, archival and disposal of classification systems. In addition, it includes processes, roles and standards that ensure the correct use of these classification systems by facilitating the incorporation of explicit semantic definitions and concordance tables to existing classifications [8]. In the last decade, several tools have been developed that facilitate the application of data and classification governance. In this project, the Data Governance Center® (DGC) software[18] (version 4.6.0) from Collibra was chosen because this tool features data stewardship and a business semantics glossary within the same tool.

In order to apply the principles of classification governance to the research discipline list, we started with specifying the roles and responsibilities towards the research discipline classification in DGC according to the RACI matrix [9]. This matrix distinguishes roles as *responsible*, *accountable*, *consulted* and *informed* (groups of) individuals. As shown in Figure 4, the ECOOM-Hasselt team was denoted as *accountable* with the responsibility to lead the process of creating the semantically enriched, harmonised classification for research disciplines and the subsequent management of the list, i.e. the registration and processing of detected and reported new research disciplines within the DGC tool. The *responsible* role was taken by a so-called 'research discipline steering committee', a committee composed of representative members of all stakeholders implementing or using the Flemish research discipline list. The steering committee has the responsibility to guard the overall process of creating and updating the Flemish research discipline list. The decisions of the steering committee were based on the advice of expert groups, the so-called *consulted*, composed of established scientists with a broad overview on their discipline and representing different research organisation types, such as universities, higher education institutions and strategic research centres. Subsequently, all stakeholders, the so-called *informed*, should be able the grasp the most up-to-date version of the Flemish research discipline list via the DGC tool.

After specifying the roles and responsibilities of the concerned parties, ECOOM-Hasselt and the research discipline steering committee determined the framework and procedures to be followed to transform the proposed draft of the research discipline list into a final operable research discipline list (Fig. 4, phase 0). This framework included the determination of the phases to review the draft research discipline list (Fig. 4, phase 1 to phase 3) as well as the procedures to follow to register, evaluate and finally approve or disapprove the reported feedback.

[18] https://www.collibra.com/data-governance-solutions/data-governance-center

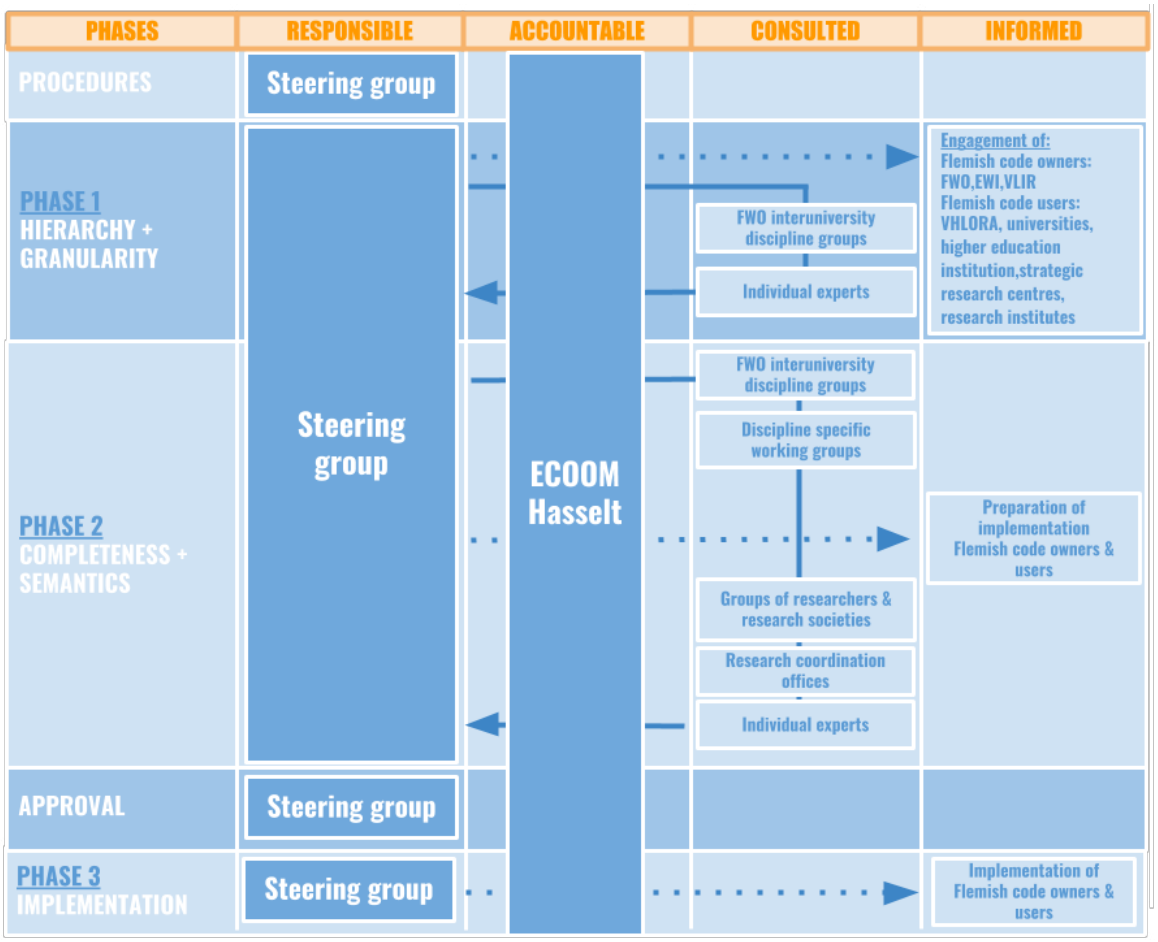

Fig 4. Schematic representation of the review process of the Flemish research discipline list according to the RACI matrix.

4.2 Phase 1: Hierarchy and granularity

One of the main incentives of this project was to create a single research discipline list, implemented by all stakeholders in the Flemish research landscape. This implied that the Flemish research discipline list was required to serve multiple purposes, i.e. reporting to European and Flemish governments, providing an overview of the Flemish R&D potential, obtaining in-depth knowledge on research performance, etc. The hierarchical structure of the Flemish research discipline list as well as its granularity had to reflect all of these purposes. Phase 1 of the evaluation of the draft research discipline list (Fig. 4) therefore focussed on evaluating the correctness of the drafted hierarchical levels and the granularity of the disciplines within each level.

In phase 1, the draft research discipline list was presented to 5 interuniversity expert groups, covering 5 disciplines ‘science and technology’, ‘cultural sciences’, ‘behaviour and society’, ‘medical sciences’, ‘biological sciences’ and consisting of researchers of all 5 Flemish universities that were involved in the past in the creation and/or update of the FWO scientific discipline code list. Additionally, 453 individual experts were consulted to give their feedback on hierarchy and granularity of the draft research discipline list out of which 294 provided feedback (response rate 64.9%). Recommendations of these experts and expert groups were incorporated in the Flemish research discipline list if benchmarking analyses demonstrated their relevance or when substantive arguments could be given that adjustments were necessary to cover the research landscape in Flanders. As an example, the discipline ‘Medical and Health sciences’ was expanded at level 2 by the discipline ‘Translational sciences’ in order to bridge the gap between ‘Basic sciences’ and ‘Clinical sciences’. In the definition of ‘Translational sciences’ it was clearly stated that this discipline concerns research on model systems mimicking pathophysiological conditions in humans, in other

words research that cannot be allocated to either basic, nor clinical sciences. Within this evaluation process, care was taken to ensure that the granularity from one hierarchical level to the next, and even between disciplines within a single hierarchical level was balanced. The expert recommendations as well as the conflicting ideas formulated by the experts were reported to the steering committee, who formulated a final decision on these matters and gave their approval to continue to the next phase.

4.3 Phase 2: Completeness, terminology and semantics

In the next phase, a series of experts and expert groups were consulted to evaluate whether level 4 of the draft research discipline list was complete and whether the predefined terminology for denoting the disciplines was correct. Additionally, these experts were asked to review the formulated definitions of the disciplines.

First of all, the same 5 interuniversity expert groups as in phase 1 were consulted. Subsequently, interuniversity and discipline specific working groups were constructed to evaluate those disciplines where substantial deviations were observed, compared to other (inter)national discipline lists in terms of incorporation and granularity. Disciplines covering the arts, for example, were evaluated by an interuniversity working group because 'arts' is a discipline that is generally only covered to a limited extent in other (inter)national discipline list and was elaborated to a more granular level in the Flemish research discipline list. For similar reasons, discipline specific working groups evaluating disciplines related to 'engineering and technology', 'architecture' and 'agricultural, veterinary and food sciences' were constructed. Finally, feedback was requested to individual researchers, groups of researchers, and research societies, directly via ECOOM-Hasselt or via the research coordination offices of the Flemish universities. During this phase, feedback of 127 out of the 469 consulted experts (response rate: 27,1%) was registered. Similar to the process in phase 1, all recommendations to add or remove certain level 4 disciplines were analysed to determine their relevance in function of other research discipline lists as well as in the context of the Flemish research landscape.

With respect to terminology, the most commonly reported feedback referred to the use of old jargon, i.e. terms that are no longer in use, and to the use of terminologies that are influenced, at least to some extent, by stakeholder policies (e.g. reflecting the name of a faculty or a teaching programme). Moreover, even the name 'research' discipline list was under discussion, as some research fields were not comfortable with using the term 'science' discipline list. Feedback on definitions on the other hand mainly focussed on demarcating neighbouring disciplines. Bearing in mind that this research discipline list is designed to cover the needs of all stakeholders in the Flemish research landscape, feedback on terminology and definitions was also presented to representatives of these stakeholder groups and their opinion on the most appropriate term or definition was also registered. All conflicting feedback that was gathered during this phase was consequently reported to the steering committee, who formulated a final decision on these matters and gave their final approval of the resulting Flemish research discipline list (available on request via ECOOM@uhasselt.be).

Important to note here is that the Flemish research discipline list was composed in English and translated after its final approval by the steering committee to Dutch. This not only facilitates the creation of concordance maps with existing international classification lists, but is also intrinsic to the nature of both languages themselves. In the past, Flemish research discipline lists were created in Dutch and translated afterwards to English, thereby causing multiple Dutch terms to be linked to a single English term and thus resulting in misinterpretations.

4.4 Phase 3: Implementation

After the Flemish research discipline list was approved by the steering committee, it was implemented in the Data Governance Center software[18], which allows for the design and inclusion of workflows that support the governance of the research classification list. These workflows are not merely focused on the maintenance of the research discipline list, but also include processes like the creation and deprecation of terms, commenting on definitions, etc. All information can be retrieved from DGC in an automated, efficient manner via an API, which ensures that stakeholders can constantly use the most recent version of the research discipline list. However, not every stakeholder is already integrating with DGC. This is largely due to the fact that the implementation of the Flemish research discipline list not merely requires its adoption by the data systems of the stakeholders, but also affects the policies and processes within the stakeholders' organisation and beyond. As such, reporting obligations needed to be adjusted towards the use of the new, harmonised research discipline list, as well as data registration and collection processes. Likewise, business and validation rules had to be defined pointing out how the research discipline list should be used by the research community, and how the proper use could be validated. In addition, concordance maps

were drafted to existing (inter)national research discipline lists, as this allows for continuity in data registration and follow-up over the years. To this end, the reference manager tool of DGC was used, as this allows crosswalk between different classification list that can be supplemented with clarifications.

Finally, the resulting Flemish research discipline list is currently in the process of being incorporated in the database systems of all Flemish stakeholders i.e. the Flemish universities, higher education institutions, research organisations and institutes, including the Flemish funders (i.e. FWO) and governmental organisations (Dept. EWI, Flemish Government) that collect and display research information, i.e. on the FRIS-portal. Obviously, the inclusion of the Flemish research discipline list in the FRIS-portal includes the active management of the research discipline list in the research information systems of all information providers, including the incorporation of the Flemish research discipline list in the FRIS-validator. This validator accepts research information sent by the information providers when in agreement with a strict set of business and validation rules. In addition, the FRIS-web portal has to be adapted towards the use of the new code list in order to display relevant information using this classification. Furthermore, a search module is being developed that will be added to the portal to allow a hierarchical as well as a text search, using the terms and semantic definitions as many stakeholders are not yet acquainted with the code list. Not surprisingly, the implementation roadmap of every stakeholder varies to a large degree depending on the characteristics of the database, the time and staff available, and the extent to which the Flemish research discipline list will be embedded in organisational processes.

4.5 The Flemish research discipline list compared to other classification systems

The Flemish research discipline list was designed to improve the efficiency and quality of research reporting in Flanders by creating a single, harmonised and semantically enriched classification system for registering and reporting on research in specific disciplines. This discipline list is, in comparison to the current (inter)national research discipline lists, unique worldwide because of two main aspects: its granularity and its semantically enriched content. That is, with its 4 hierarchical levels, the Flemish research discipline list is one of the most granular research discipline lists worldwide. Additionally, the Flemish research discipline list is the first research discipline list worldwide that includes definitions to semantically define what is included in every discipline.

Furthermore, the Flemish research discipline list was created by harmonising the current (inter)national research discipline lists that are used in Flanders and enriched by the recommendations of experts. To quantify how the content of the Flemish research discipline list differs from the current (inter)national research discipline lists used in Flanders, the number of disciplines underlying level 1 of each of these discipline code list was calculated and subsequently mapped onto the different level 1 disciplines of the Flemish research discipline list[19]. By expressing these numbers as proportions relative to the total number of disciplines within each research discipline list, we corrected for differences in the total number of disciplines in the different discipline code lists. Figure 5 shows that the Flemish research discipline list expands all currently used research discipline lists with the a code representing ‘General and logistic services’, that was added to ensure a continuity in research reporting from the previous VLIR science discipline codes to the Flemish research discipline list.

Compared to the FORD classification list [4], the Flemish research discipline list attributes a smaller proportion of disciplines to ‘Engineering and technology’ and ‘Agricultural, veterinary and food sciences’ but attributes a larger proportion of disciplines to ‘Medical and health sciences’. A similar pattern can be found in the comparison of the Flemish research discipline list and the VLIR scientific disciplines. That is, a smaller proportion of disciplines is attributed to ‘Engineering and technology’ whereas a larger proportion of disciplines is attributed to ‘Medical and health sciences’.

Compared to the FWO scientific disciplines, the proportion of disciplines attributed to ‘Medical and health sciences’ and ‘Agricultural, veterinary and food sciences’ is larger whereas the proportion of disciplines attributed to ‘Engineering and technology’ and ‘Social sciences’ is smaller in the Flemish research discipline list.

The IWETO/FRIS science discipline list and the Flemish research discipline list differ mainly in 3 disciplines. That is, the proportion of disciplines attributed to ‘Humanities’ is smaller whereas the proportion of disciplines attributed to ‘Medical and health sciences’ and ‘Agricultural, veterinary and food sciences’ is larger in the Flemish research discipline. The difference between the FRIS/IWETO scientific disciplines and the Flemish research discipline

[19] Because level 4 of the FWO scientific disciplines is not populated for all disciplines, the analysis of the FWO scientific disciplines only took the disciplines of level 2 and level 3 into account.

list in the field of 'Medical and health sciences' becomes even larger when correcting for the fact that the discipline 'Agricultural, veterinary and food sciences' was not recognised as a separate discipline in the FWO scientific discipline but was incorporate in the discipline 'Biomedical sciences' in this analysis classified in 'Medical and health sciences'.

This analysis clearly shows the emphasis of each of the analysed research discipline lists, which is perfectly in line with the use purpose for which these lists were created. In general, we can conclude that the Flemish research discipline list has a smaller proportion of disciplines attributed to 'Engineering and technology' compared to most of the current (inter)national research discipline lists used in Flanders. This is largely due to the integration of 'Engineering' on the one hand and 'Technology' on the other hand into a single 'Engineering and Technology' category. Additionally, the proportion of disciplines attributed to 'Medical and health sciences' in the Flemish research discipline list is larger compared to most of the current (inter)national research discipline lists used in Flanders. This increase in 'Medical and health sciences' disciplines is the result of the addition of the discipline 'Translational sciences' to the level 1 discipline 'Medical and health sciences'. The inclusion of this discipline in the Flemish research discipline list was requested by experts that reported a gap between medical research in the context of 'Basic sciences' and in the context of 'Clinical sciences'. Benchmark analysis could not identify a precedent to include the discipline 'Translational sciences' in the Flemish research discipline list. However, in the ANZSRC a distinction between 'Basic sciences', 'Translational sciences' and 'Clinical sciences' can be made based on the classification of activity types[5]. Therefore the discipline 'Translational sciences' was allowed to be included at level 2 of the Flemish research discipline list, with the agreement that for reporting purposes 'Translational sciences' would be merged with 'Basic sciences'. Finally, as shown by Figure 5, the disciplines attributed to 'Natural sciences', 'Humanities and the arts' and 'Social sciences' are largely comparable to the current (inter)national research discipline lists used in Flanders.

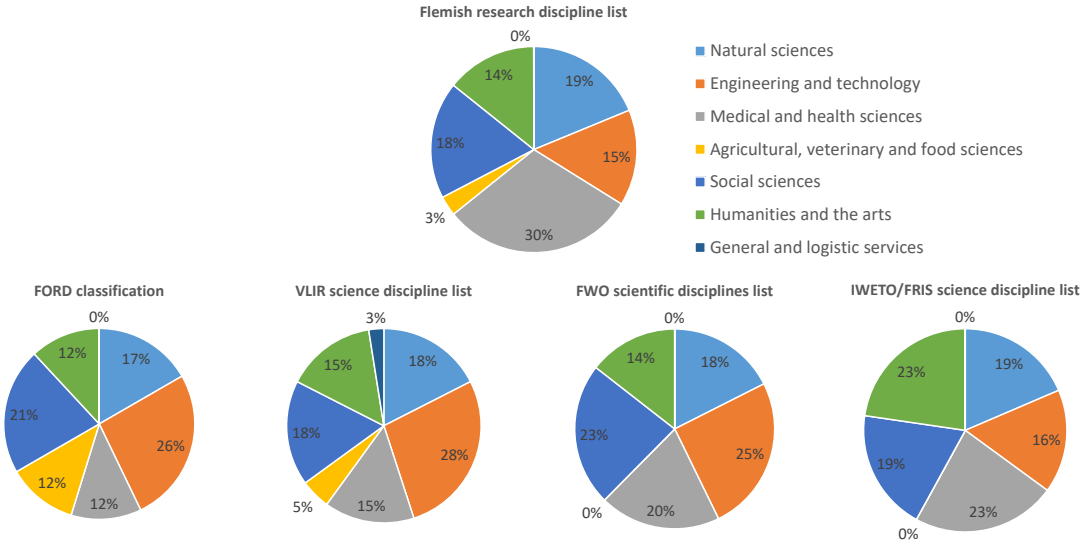

Fig 5. The proportion of disciplines that can be attributed to the 7 different level 1 disciplines of the Flemish research discipline list for the FORD classification list, the IWETO/FRIS science discipline list, the FWO scientific disciplines, the VLIR scientific disciplines and the Flemish research discipline list.

5. Discussion & conclusion

In conclusion, we applied data and classification governance as a methodology to create a semantically enriched, harmonised research discipline list, that can be used to classify researchers and research projects for reporting, dissemination and evaluation purposes. Using this methodology, we created a research discipline list consisting of 4 hierarchical levels, with 7, 42, 382 and 2.493 codes respectively that can be updated in a semi-automated manner.

Each code is accompanied with a semantic definition that allows for an accurate use of the research discipline list by all stakeholders. In addition, business rules were drafted to further enhance accurate data registration and collection. Finally, concordance maps to existing research disciplines were created in order to facilitate the follow-up on previously recorded data. Altogether the application of the DCG methodology assisted the creation of a semantically enriched research discipline list, that can be used for an accurate data registration, collection and interpretation, thereby allowing for a true understanding of research performance by all stakeholders. Obviously, the Flemish research discipline list will need to be updated in the future according to the dynamics of the research world. These alterations can encompass both the granularity of the research disciplines as well as the terminologies and semantics used for which semi-automated mechanisms were included in the list. In addition, the Flemish research community can address issues directly with ECOOM-Hasselt, which will continue to govern the Flemish research discipline code list in the future. Furthermore, ECOOM-Hasselt will actively seek to promote the use of semantically enriched research classification systems and ultimately the harmonisation thereof into an international research classification standard with concordance schemes to existing (inter)national research classifications.

Nomenclature

ANVUR	Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca
ANZSRC	Australian and New Zealand Standard Research Classification
CASRAI	Consortia Advancing Standards in Research Administration Information
CG	Classification Governance
CSC	China Subject Categories
DGC	Data Governance Center
ERC	European Research Council
FAPESP	Fundaçáo de Amparo Á Pesquisa do Estado de Sáo Paulo
FOR	Fields of Research
FORD	Fields of Research and Development
FOS	Fields of Sciences and Technology
FRIS	Flanders Research Information Space
FWO	Fonds voor Wetenschappelijk Onderzoek
NWO	Nederlandse Organisatie voor Wetenschappelijk Onderzoek
OECD	Organisation for Economic Co-operation and Development
ÖFOS	Österreichische Systematik der Wissenschaftszweige
RAE	Research Assessment Exercise
REF	Research Excellence Framework
RF	Research Fields
RSA	Republic of South Africa
SSD	Settore scientifico disciplinare
VLIR	Vlaamse Interuniversitaire Raad
VLHORA	Vlaamse Hogescholen Raad

Acknowledgements

This work is part of the Classification Governance project carried out for the Expertise Centre for Research & Development Monitoring (ECOOM) in Flanders, which is supported by the Department of Economy, Science and Innovation, Flanders.

References

[1] Peters, A., Lambrechts, L. (2011). De vereenvoudiging van onderzoeksverslaggeving, een analysetraject uitgevoerd door de Vlaamse universiteiten en hogescholen en de VLIR, in opdracht van de Vlaamse Overheid (EWI).
[2] Alexander, J., Lambe, P. (2015). Classifying R&D: Why and how organizations develop taxonomies for research fields. In CASRAI ReConnect15 Conference Proceedings.
[3] OECD (2002). Frascati Manual 2002: Proposed standard practice for surveys on research and experimental development, The measurement of scientific and technological activities, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264199040-en.
[4] OECD (2015). Frascati Manual 2015: Guidelines for collecting and reporting data on research and experimental development, The measurement of scientific, technological and innovation activities, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/9789264239012-en
[5] Thomson Reuters (2016). Global institutional profiles project: academic reputation survey, 2016 Report of findings, http://ip-science.thomsonreuters.com/m/pdfs/GIPP_AcamRep_report.pdf
[6] Becher, T. (1994). The significance of disciplinary differences. In Studies in higher education, 19(2), 159.
[7] Logan, D. (2010). What is information governance? And why is it so hard? In http://blogs.gartner.com/debra_logan/2010/01/11/what-is-information-governance-and-why-is-it-so-hard/
[8] Vancauwenbergh, S., De Leenheer, P., Van Grootel, G. (2016). On research information and classification governance in an inter-organizational context: the Flanders Research Information Space. Scientometrics, 108, 425-439.
[9] Project Management Institute (2013). 9.1.2.1 Organization charts and position descriptions. In A Guide to the project management body of knowledge (PMBOK Guide) (5th ed.), 592.