The creation of the Flemish research discipline list, a huge step forward in harmonizing research information (systems)

(This extended abstract is submitted for presentation on the upcoming euroCRIS conference 2018. We plan to elaborate a bit more on the implementation issues incl. the creation of concordance tables to existing classifications in the final paper.)

Abstract

In 2011, a report was written by Peters et al. on the administrative simplification of research reporting in Flanders. Next to the description of several data flows that could be harmonized, 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, i.e. research discipline codes are used for denoting a researcher's discipline, while financial codes are used for reporting on the funding type, and technology and publication codes are used for depicting innovation sectors and publication types respectively,...

Next to the variety of classification schemes itself, each classification scheme has multiple classification lists depending on the authority to whom a researcher is obliged to report. This has led in the past to the existence of several research discipline lists for a region as small as Flanders. Therefore, one of the most important recommendations of the report by Peters et al. (2011) 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.

Introduction

Researchers and research institutions worldwide have to report on their research throughout the entire research lifecycle to governments, financers and third parties for granting and evaluation purposes. Next to the description of the research itself, these reports often contain figures on a wide variety of research output categories: publications, projects, financial data, patents, spin-offs, staff, etc. At first glance, these figures seem to provide an easily comprehensible and straightforward means to measure research performance. However, often the semantics underlying the data is overlooked by the research community, both when gathering the information as well as during research assessment possibly leading to erroneous conclusions.

This is partly due to the exponentially growing administrative burden, that is being put on the research community. As such, researchers and research administrations have to report on their research to many different stakeholders, each using different requirements with regards to the format, the actual content and the classifications used, that follow the stakeholders' goals, mind-set and thus specific semantics. In addition, the collection of the data is a daunting task, as the required information is hardly ever centralised in a single system but mostly resides in different databases. These systems are often designed or adopted based on the requirements of the research institution itself, which on their turn specify the manner and characteristics of the information being recorded and often reflect the information provider's mind-set. However the terms used in these databases are mostly not explicitly defined and the recording of the 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 such systems, without explicitly consulting the underlying semantics, different data can be combined erroneously leading to inconclusive data. The problem aggravates when information consumers 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.

The same applies for the classifications used in metrics. Hardly any classification is provided with an explicit semantic description, thereby opening the door for interpretative use by research administrators and evaluators.

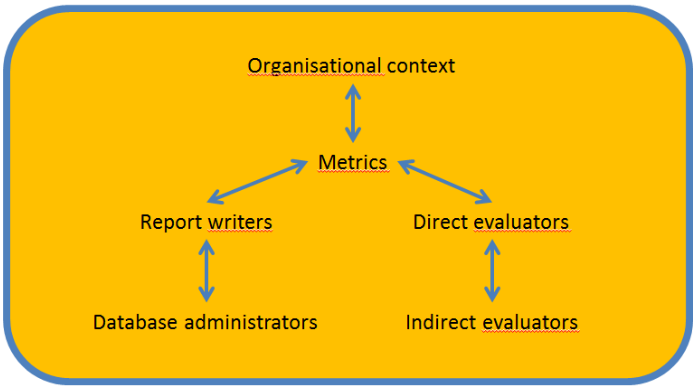

Figure 1: The origin of semantic misinterpretation. Database administrators record information mostly using their own informal definitions. This can lead to errors when information from different systems is combined by report writers, and sent to organizations that derive metrics from these data. These organisations generally use rather vague definitions, if present, thereby opening the door for interpretative use by information providers and consumers. The latter consist of direct evaluators, persons that are familiar with the indicators and classifications used, and

indirect evaluators, persons that have an interest in metrics but do not have an in-depth knowledge of the research environment.

Research classifications

Research classifications are mostly hierarchical systems that divide research information according to specific characteristics in different categories, each denoted with a unique code or number. Although a wide variety of research classifications exists, they basically fall into 3 categories describing either the identity, the input or output side of a researcher. 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. The input side of a researcher is often covered by classifications that describe the financial sources provided to research in terms of their origin (i.e. financer, geographical region, …). The largest variety of classifications, however, can be found on the output side of research 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 organizational levels, ranging from single institutions, to regional and (inter)- and (supra)national governments, developed according to the specific operational needs of each organization(al) entity (Alexander, J. and Lambe, P., 2015). 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 and their conclusions.

The Flemish landscape of science disciplines lists

In 2011, Peters et al. wrote a report on the reduction of research information reporting. Next to guidelines on the harmonisation of research reports, indicators and classifications, the report specifically states the importance of an unambiguous semantics underlying the indicators and accompanying classifications by all stakeholders. Although the importance of well-defined indicators is generally well recognized, not much attention has been drawn to the poor semantics of research-related classifications. However, these are fundamental for obtaining accurate and comparable data that allow a true understanding of research performance. This paper therefore focuses on the use of data and classification governance as a method to develop a semantically enriched classification of Flemish research disciplines.

Currently, Flemish research administrations use 4 different research discipline lists, i.e. 1) the revised OECD Field of Science and Technology Classification list (FOS) as published in the Frascati Manual, 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 revised 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 the early nineties, the IWETO/FRIS science discipline list 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 science 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 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 1029 codes. The third hierarchical level was sometimes even further categorized. 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 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.

Towards a draft of the Flemish research discipline list

Based on the report by Peters et al. (2011), 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.

Information collection

As a first step, we collected information on existing science discipline classifications used for research reporting both in Flanders and beyond. We specifically focussed on classifications that cover the full spectrum of research, thereby omitting lists used for bibliometric purposes as these only cover written scientific output. Furthermore, special attention was drawn 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 science disciplines regardless of their geographical origin (Thomson Reuters, 2016). In total, we considered 12 science 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 did use 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 even 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 science 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 science disciplines, as they contain various perspectives such as philosophical, anthropological, sociological, historical and managerial aspects that should be handled (Krishnan, A. *et al.*, 2009). On the other hand this can be explained by the implications that distinctions in science disciplines have on higher education research, policy and practice (Becher, T. *et al.*, 1994).

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	Australia & New Zealand	3
CASRAI	America	3
CSC	China	2
ERC	Europe	2
OECD	Europe	2
FAPESP	Brazil	2
RSA-RF	Africa	2
ANVUR	Italy	3
SSD	Italy	4
ÖFOS	Austria	3
RAE	UK	1
REF	UK	1
NWO	The Netherlands	2
FWO	Flanders	3
FRIS	Flanders	2
VLIR	Flanders	1

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 1-2 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 (Figure 2). In agreement with the report on reduction of research reporting, 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 organizations 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 organize 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.

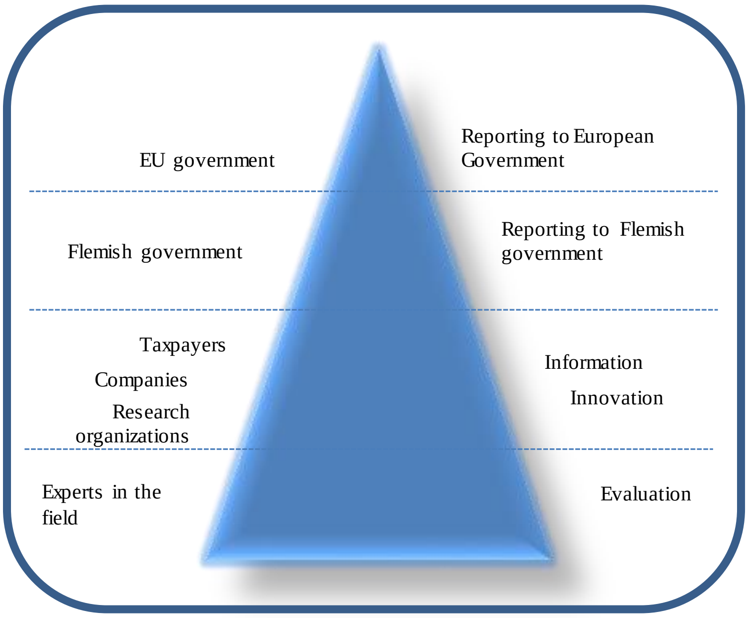

Figure 2: Use purposes of the new, harmonised research disciplines list

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

In an attempt 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 revised Field of Science and Technology list of OECD as this list should be used for reporting purposes to the European government. As not too many differences were observed, we proposed to adopt the first level of the revised Field of Science and Technology list of OECD 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 & technology, medical & health sciences, agricultural sciences, social sciences and humanities. A seventh code was added later on 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 OECD list, 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 that time. However, comparison with the other 15 research discipline lists resulted in the identification of 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.

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 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 and thus deserves mention. The ANZSRC-FOR research discipline list however, uses a semi-automated strategy to detect new emerging research disciplines. Moreover, the second level and the 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 system 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 system in the harmonised Flemish research discipline list and considered 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 organizations (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.

06	Humanities	0604	Arts	060403	Art studies and sciences	06040301	Architectural history and theory
06	Humanities	0604	Arts	060403	Art studies and sciences	06040302	Criticism and theory
06	Humanities	0604	Arts	060403	Art studies and sciences	06040303	Curatorship
06	Humanities	0604	Arts	060403	Art studies and sciences	06040304	Architectural design history and theory
his	Humanities	0604	Arts	060403	Art studies and sciences	06040305	Film studies
06	Humanities	0604	Arts	060403	Art studies and sciences	06040306	History of art
06	Humanities	0604	Arts	060403	Art studies and sciences	06040307	History of music
06	Humanities	0604	Arts	060403	Art studies and sciences	06040308	History of performing arts
06	Humanities	0604	Arts	060403	Art studies and sciences	06040309	History of stage craft
06	Humanities	0604	Arts	060403	Art studies and sciences	06040310	Iconology
06	Humanities	0604	Arts	060403	Art studies and sciences	06040311	Interior architecture history and theory
06	Humanities	0604	Arts	060403	Art studies and sciences	06040312	Landscape architecture history and theory
06	Humanities	0604	Arts	060403	Art studies and sciences	06040313	Musicology and ethnomusicology
06	Humanities	0604	Arts	060403	Art studies and sciences	06040314	Performance studies
06	Humanities	0604	Arts	060403	Art studies and sciences	06040315	Theatre science
06	Humanities	0604	Arts	060403	Art studies and sciences	06040316	Visual cultures
06	Humanities	0604	Arts	060403	Art studies and sciences	06040399	Art studies and sciences not elsewhere classified
06	Humanities	0604	Arts	060404	Artistic design		
06	Humanities	0604	Arts	060404	Artistic design	06040401	Autonomous design
06	Humanities	0604	Arts	060404	Artistic design	06040402	Costume design
06	Humanities	0604	Arts	060404	Artistic design	06040403	Craft design
06	Humanities	0604	Arts	060404	Artistic design	06040404	Design for performance
06	Humanities	0604	Arts	060404	Artistic design	06040405	Exhibition design
06	Humanities	0604	Arts	060404	Artistic design	06040406	Fashion design
06	Humanities	0604	Arts	060404	Artistic design	06040407	Furniture design
06	Humanities	0604	Arts	060404	Artistic design	06040408	Graphic design
06	Humanities	0604	Arts	060404	Artistic design	06040409	Jewelry design and gold/silversmith arts
06	Humanities	0604	Arts	060404	Artistic design	06040410	Product design
06	Humanities	0604	Arts	060404	Artistic design	06040411	Service design
06	Humanities	0604	Arts	060404	Artistic design	06040412	Social design
06	Humanities	0604	Arts	060404	Artistic design	06040413	Textile design
06	Humanities	0604	Arts	060404	Artistic design	06040414	Typography
06	Humanities	0604	Arts	060404	Artistic design	06040499	Artistic design not elsewhere classified
06	Humanities	0604	Arts	060405	Audiovisual art and digital media		
06	Humanities	0604	Arts	060405	Audiovisual art and digital media	06040501	Acting
06	Humanities	0604	Arts	060405	Audiovisual art and digital media	06040502	Animation film
06	Humanities	0604	Arts	060405	Audiovisual art and digital media	06040503	Cinematography
06	Humanities	0604	Arts	060405	Audiovisual art and digital media	06040504	Computer gaming and animation
06	Humanities	0604	Arts	060405	Audiovisual art and digital media	06040505	Digital media
06	Humanities	0604	Arts	060405	Audiovisual art and digital media	06040506	Documentary film
06	Humanities	0604	Arts	060405	Audiovisual art and digital media	06040507	Editing

Figure 3: The Flemish Research Discipline List_version 2018

Altogether, this resulted in 4-hierarchical draft list or respectively 7, 42, 382 and 2866 terms, reflecting research disciplines to a different level of granularity (Figure 3). This list later on served as a starting point on which the data and classification governance method was imposed, in order to create a semantically enriched, harmonised Flemish research discipline list.

Towards a semantically enriched classification for research disciplines

Classification governance as methodology

Classification governance (CG) basically stems from data and information governance (Logan, 2010). 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 (Vancauwenbergh *et al.*, 2016). These semantic definitions are extremely valuable, as the terminologies present in each classification list are often interpreted differently by each stakeholder according to their linguistic, thematic and situational context. 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, version 4.6.0 from Collibra has been used, as this comprises a data stewardship tool in combination with a business semantics glossary.

In order to apply the CG method on the draft research discipline list, we started with specifying the roles and responsibilities towards the research discipline classification in DGC using the RACI matrix. This matrix distinguishes roles as responsible, accountable, consulted and informed (groups of) individuals. In brief, the ECOOM-UHasselt team was denoted as accountable with the responsibility to guide 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 (seemingly) new research disciplines within the DGC tool and other related issues. The responsible role was taken by a so-called 'research discipline steering group', that had the responsibility to guard the overall process of creating the semantically enriched, harmonised classification list for research disciplines. Processes were designed in order to support the registration, review, discussion and possible approval of research disciplines. Furthermore, additional processes were designed that allowed the steering group to take decisions when no agreements on research disciplines were found. These were generally taken within expert groups, the so-called consulted, composed of established scientists that had a broad overview on their discipline and representing different research organizations types, i.e. universities, higher education institutions and strategic research centres. The consulted expert groups also provided their opinion on the semantic definitions proposed. The resulting information was then fed into the DGC tool, where all stakeholders could grasp the information kept, thereby representing the informed role. Obviously, the establishment of such a research classification governance board requires the involvement, responsibility and thus investments of all people involved. In the remainder of the paper, we will focus on the difficulties we encountered during the creation of the semantically enriched, harmonised research discipline list.

Granularity-related issues

A first difficulty concerned the creation of a single list that could serve multiple purposes, i.e. reporting to European and Flemish governments, providing an overview of the Flemish R&D potential as well as the possibility to obtain an in-depth knowledge on the research performed for evaluation purposes, in consultation with all stakeholders. Obviously, each hierarchical level required a varying degree of granularity corresponding to its use purpose. For the first level we adopted the highest hierarchical level of the OECD Field of Science and Technology indicators list, as this would simplify the creation of a concordance table afterwards. The second hierarchical level was largely based on the second level of the OECD FOS list. Additions were allowed by the expert panel if benchmarking analysis demonstrated their need or when profound, substantive arguments were found. As such 'Translational sciences' was added to the 'Medical and Health sciences', as this research discipline bridges the gap between 'basic sciences' and 'clinical sciences', and concerns research on model systems mimicking pathophysiological conditions in humans as exemplified by its definition, that cannot be allocated to either basic, nor clinical sciences. With each following hierarchical level, more terms were proposed by the stakeholders as possible new research disciplines. These were all analysed using the methodology mentioned above in order to omit possible conflicts of interest. In addition, care was taken to ensure that the granularity from one hierarchical level to the next, and even within a single hierarchical level was distributed evenly. When no agreements were found within the expert groups or conflicting ideas were formulated in between the expert groups, the steering group was consulted for a final decision.

Terminology and semantics-related issues

Secondly, the terminologies used for denoting research disciplines were often the subject of discussions within the expert groups. Some terminologies in our draft list were considered as old jargon that was no longer in use, while other terminologies were considered to reflect influences of political or other kind, at least to some extent. Moreover, even the name of the 'research' discipline list was under discussion, as some research fields were not comfortable with using the term 'science' discipline list. Bearing the goal of a harmonized research discipline list in mind, we therefore consulted several experts from each stakeholder group and asked their opinion on the most appropriate term. The most prevalent term was taken up in the research discipline list. Importantly, we composed the list in English and translated it only afterwards to Dutch. This not only facilitates the creation of concordance tables 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.

Next, a semantic definition had to be defined for each recognized 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), professional literature and Wikipedia. Afterwards the definitions were sent to experts for review and their adaptations were taken over in the current version of the research discipline list (Addendum 1).

Implementation-related issues

Finally, the thus composed list was implemented in the Data Governance Center software, 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, the 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 new 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' organization and beyond. As such, reporting obligations needed to be adjusted towards the use of the new, harmonized research discipline list, as well as data registration and collection processes. Likewise, business rules had to be defined that point out how the research discipline list should be used. In brief, researchers and research projects can be characterized by maximally 7 research discipline codes, each specifying research disciplines to the lowest hierarchical level, and appearing in order of decreasing importance. The 2 highest levels of the first code are used for reporting purposes, while the combination of all other codes can be used for dissemination on the research as well as for evaluation purposes, as a means to demonstrate their multidisciplinary character. In addition, concordance tables had to be drafted to existing (inter)national research discipline list, 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.

Discussion & conclusion

In conclusion, we applied data and classification governance as a methodology to create a semantically enriched, harmonized 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 currently 7, 42, 382 and 2866 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 that further enhance accurate data registration and collection. Finally, concordance tables 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.

However, many updates of the research discipline list will remain essential to keep up with the dynamics of the research world. These alterations can encompass both the granularity of the research discipline list as well as the terminologies and semantics used. Shortcomings in the granularity can be detected in a semi-automated manner (see above) or can be reported directly to the Centre for Research & Development Monitoring, who can initiate the discussion within the expert groups as mentioned above. In addition, ECOOM will also yearly evaluate the use of the discipline codes residing within the FRIS database. In addition, we plan a more in-depth review of the semantic definitions, as these currently contain a varying degree of detail, largely due to the lack of a controlled, uniform reviewing process. To this end, we will implement the principles of terminological theory in the data and classification governance methodology. Terminological theory is used by linguists to record how terminologies are interpreted according to their linguistic, thematic and situational context by different stakeholders (Kockaert and Steurs, 2014). Based on an analysis on the similarities and differences of the used connotations, harmonized semantic descriptions are proposed. These are reviewed and tested in iterative cycles until harmonized definitions are found that are disseminated to all stakeholders together with possible discrepancies. However, unlike the waterfall model of Bell and Thayer (1976), the process of classification governance does not end here. On the contrary, based on the dynamics of the research world, new cycli can be required (Boehm et al. 2000). In addition, the stakeholders' needs and wishes can change, that demand further adaptations to the research discipline list (Takeuchi and Nonaka, 1986; Henry and Henry, 1993). As such data and classification governance can be considered as a continuous, dynamic process essential for maintaining the semantically-enriched, research discipline list up-to-date, ensuring its appropriate use and thus allowing for a true understanding of research performance by all stakeholders.

Acknowledgments

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.

List of abbreviations

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
FOS: Field 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

References

Alexander, J., Lambe, P. (2015) *Classifying R&D: Why and how organizations develop taxonomies for research fields*. In: CASRAI ReConnect15 Conference Proceedings

Becher, T. (1994) *The Significance of Disciplinary Differences*. In: Studies in Higher Education 19:2, p. 159.

Bell, T.E., Thayer, T.A. (1976) *Software requirements: Are they really a problem?* Proceedings of the 2nd international conference on Software engineering. IEEE Computer Society Press.

Boehm B. (1986) *A Spiral Model of Software Development and Enhancement.* ACM SIGSOFT Software Engineering Notes, ACM, 11(4):14-24.

Debruyne, C., & De Leenheer, P. (2011) *Business semantics as an interface between enterprise information management and the web of data: A case study in the Flemish public administration.* In eBISS 2012,208–233.

Debruyne, C., De Leenheer, P., Spyns, P., et al. (2011) *Publishing open data and services for the Flemish Research Information Space*. In Advances in conceptual modelling. Recent developments and new directions. Lecture notes in computer science, 6999, 389–394.

De Leenheer, P., Christiaens, S., & Meersman, R. (2010) *Business semantics management: A case study for competency-centric HRM*. Computers in Industry, 61(8), 760–775.

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

Henry, J., Henry, S. (1993) *Quantitative assessment of the software maintenance process and requirements volatility*. In Proc. of the ACM Conference on Computer Science, pages 346–351.

Kockaert, H.J.; Steurs, F. (2014) *Handbook of Terminology*. Volume 1, Amsterdam, Philadelphia, John Benjamins.

Takeuchi, H. Nonaka, I. (2010) *New New Product Development Game*. Harvard Business Review. Retrieved June 9.

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

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).

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.