

On research information and classification governance in an
inter-organizational context: the Flanders Research Information Space

Peer-reviewed author version

VANCAUWENBERGH, Sadia; De Leenheer, Pieter & Van Grootel, Geert (2016) On research information and classification governance in an inter-organizational context: the Flanders Research Information Space. In: *Scientometrics*, 108 (1), p. 425-439.

DOI: 10.1007/s11192-016-1912-7

Handle: <http://hdl.handle.net/1942/20437>

On research information and classification governance in an inter-organizational context: the Flanders Research Information Space

Sadia Vancauwenbergh^{a, b}, Pieter De Leenheer^c, Geert Van Grootel^d

^a *Centre for R&D Monitoring, Division ECOOM-UHasselt, Hasselt (Belgium)*

^b *Hasselt University, Research Coordination Office, Hasselt (Belgium)*

^c *Collibra, Inc., New York (USA)*

^d *Flemish Government – Department Economy, Science and Innovation, Brussels, Belgium*

Abstract:

In a knowledge-based economy, a good overview of the scientific and technological portfolio is essential for policy formation and driving knowledge transfer to the industry and the broad public. In order to enhance open innovation, the Flemish public administration has created a Flanders research information portal that integrates information available from its data providers (research institutions, funding organizations...) using the Common European Research Information Format standard. Although this standard allows for almost unlimited flexibility for modelling the research information, it has limitations when it comes down to communication to end-users, in terms of semantics. However, interoperability of research information is only meaningful when a well-defined semantics is used and hence acts as a leverage for comparability of the information provided on the portal, including its derived services offered to the research community, research-driven organizations and policy makers. This paper describes the implementation of a business semantics tool that governs the meaning of the data concepts and classifications used for research information, in particular research funding, as a means to unambiguously exchange and interpret research data.

Keywords: research information systems, semantics, interoperability, data and classification governance, concordance tables, comparability, research funding, inter-organizational context

Address for correspondence:

SADIA VANCAUWENBERGH

ECOOM-UHasselt & Hasselt University - Research Coordination Office,
Martelarenlaan 42

BE-3500 Hasselt (Belgium)

Email: Sadia.Vancauwenbergh@uhasselt.be

Introduction

A couple of decades ago, the demands on the research community to report on research information were rather low. Scientific findings were published in preferably highly-rated journals and rather limited research reports were written. Other types of information related to a researcher's innovativeness (i.e. patents, products), infrastructure (i.e. equipment, facilities) or project portfolio were difficult to grasp or even non-existing. At best, the pieces of information were stored in various databases, predominantly developed per organization and hence diverging in content, format and storage system (**Figure 1**). Moreover, the semantics accompanying the collected data and classifications used, were not explicitly defined which was largely due to the rather low volume of digital data and amount of people involved in designing and maintaining these databases. In addition, tools to define and manage semantics in inter-organizational contexts were lacking.

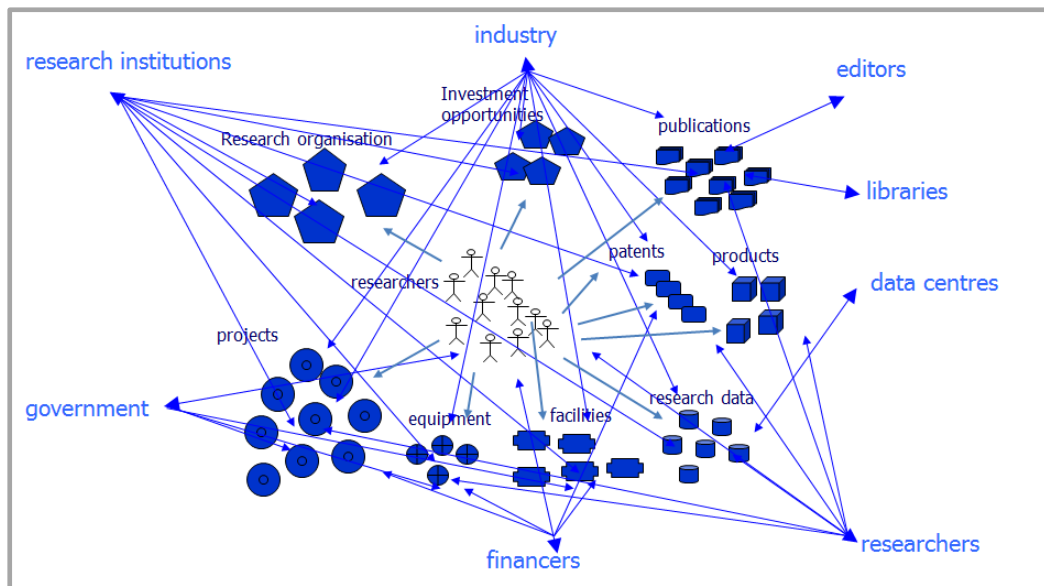


Figure 1: Overview of the research information landscape

Researchers are associated to research organizations where they carry out research projects using equipment and facilities. This results in a large amount of research data, described in publications, potentially leading to patents and innovative products. All these entities are in fact different types of research information, reported to or stored by specific stakeholders, i.e. research institutions, financiers, industry, government, ... This results in a large amount of databases, diverging in content, format and storage system.

Over the years, the research system has vastly expanded, resulting in a massive increase in the amount and nature of the information stored as well as its information consumers. These changes are not only due to the advancements made in the research field itself, but are also explained by the global efforts undertaken to transfer the obtained knowledge to industry and the broad public. Obviously, in a knowledge-based economy, a good overview of the scientific and technological portfolio is essential for driving open innovation. In many countries, this resulted in similar initiatives leading to the creation of *current research information systems* (CRIS). In general, a CRIS can be defined as a database or other information system used to store and manage data about research conducted at a (group of) research institution(s) (euroCRIS, 2015). As such, CRIS systems not only unveil research information in a better and more complete manner, but also allow for the development of webservices and applications that serve the needs of all involved stakeholders. For instance, CRIS can provide information services that assist policy makers in monitoring and analyzing research performance more efficiently and effectively. This increases a nation's strategic intelligence and leads to more informed research policies and better investments. As research output can easily be retrieved via the CRIS, entrepreneurs and technology transfer organizations have an early access to novel ideas and technologies, which stimulates innovation and eventually leads to a more prosperous society. Other benefits of such knowledge-assisted environments, include the possibilities that CRIS offer to support researchers and research administrators by helping to reduce the administrative burden. As information is kept in a central system, innovative systems can be build that simplify administrative processes by maximally reusing the information kept. In addition, funding agents can use the CRIS and its derived webservices to optimize the entire funding process, i.e. from making informed decisions on delineating funding opportunities, facilitating the grant application and evaluation process as well as automating the mapping of the resulting research output both on the short and long term. This obviously is also beneficial for the researchers who can invest the time saved on repeatedly providing similar information in grant applications

and reports, into their core business, i.e. research. Furthermore, CRIS allows researchers to discover new knowledge depending on the amount and type of interconnected (open) data and research managers can easily measure and analyze research activities. As CRIS systems offer a multitude of benefits, many countries, institutions, projects and vendors have created their own CRIS. An inventory of the European CRIS systems can be found at the Directory of Current Research Information Systems (DRIS, 2015).

The Flanders Research Information Space

In Flanders, these needs and drivers for a CRIS resulted in a program of the Flemish Department of Economy, Science and Innovation (EWI) to create the Flanders Research Information Space (FRIS) portal, which is an online research information space making Flemish research information publicly available to all stakeholders in science, economy and innovation (Van Grootel, G., *et al.*, 2009; Spyns, P., *et al.* 2011).

In its first realization, the FRIS-portal primarily acts as an information retrieval platform. Currently, the portal displays information on 29.267 researchers, their affiliated organizations and 29.735 research projects (**Figure 2**). Furthermore, the platform can easily be browsed for individual researchers, research groups, projects and collaborations. This information is retrieved directly from the information providers, i.e. the Flemish university database systems via exchange through the Common European Research Information Format (CERIF)-XML, after which it undergoes validation checks and finally is arranged at the back-end via the CERIF data model (euroCRIS, 2013). The explicit implementation of CERIF equips the FRIS environment with a powerful instrument for setting up scalable and quality-oriented information systems. Since its release in 2000, CERIF has become the European standard for research information management. As such, CERIF comprises a data model with declared semantics for most but not yet all concepts on research information and offers a high level of flexibility for modelling research information. Importantly, CERIF uses a standard formal syntax which allows data exchange at the international level. Recent findings by Jörg *et al.* (2014) have indeed shown that the CERIF format can be transformed to CASRAI, another well-known standard in the domain of research information management.



Figure 2: The Flanders Research Information Space 1.0 (*researchportal.be*)

Although both standards offer many advantages towards the creation of research information systems, both have limitations when it comes down to communication to end-users, in terms of semantics, especially in an inter-organizational context. We have analyzed these problems in-depth in an earlier publication, i.e. Debruyne and De Leenheer (2011). In brief, the problem arises with the requirement that all users coming from different organizations have to use exactly the same terminology and corresponding meaning as used in the CERIF scheme. However, information providers mostly have their own data systems with accompanying classifications and vocabularies already in place, even long before the FRIS initiative was launched. As a result, the information providers often use the CERIF terms for (slightly) different concepts, or alternatively, use a different terminology for a similar CERIF concept. Obviously, a true qualitative research information system can only be guaranteed when the semantics used by each individual information provider is mapped formally to a common exchange format, in order that research information delivered to FRIS is uniform and comparable.

As the FRIS-portal is currently being expanded both in the number of information providers as well as in terms of content (publications, research output), the focus on semantics is even more important in FRIS 2.0. Moreover, by extending the landscape of data providers from the universities to university colleges, strategic research centers, research institutes, funding organizations and other independent research organizations, the demands on FRIS 2.0 are even bigger as it is faced with a multitude of data environments. These are either self-designed or commercially available systems, ranging from CRIS to an (integrated) system of databases and

repositories, each with different data storage formats, different data input and upload techniques as well as classifications used. Altogether, these challenges urged the need for the integration of data and classification governance as a prerequisite for a long-term, qualitative and scalable FRIS system.

Towards a semantic interoperable FRIS 2.0

In order to achieve such a vigorous semantic interoperable data platform that allows for impeccable data interoperability and information exchange, the overall FRIS design had to be re-evaluated. In its first realization, the FRIS design was largely built on the CERIF model, which is a relational database model available as SQL scripts based on common entity-relationship model constructs (Jörg, B. *et al.*, 2009). In the CERIF model, research information is represented by means of entities, i.e. Person, Project, OrganisationUnit being the most important ones, and the relationships between these. Importantly, the relationships can be contextualized by applying a formal semantics in a timely context. Although very powerful, the CERIF format does not take into account the many contextual differences that exist at the side of the individual data providers. Based on the CERIF model, FRIS 1.0 largely consisted of an application layer combined with a technological data layer to which research information was sent by IT professionals. However, these professionals were mostly not aware of the exact meaning of the data nor the classifications used, which is kept with the domain experts, i.e. research coordinators, librarians, technology transfer officers and database administrators who are mostly not familiar with specific database modelling methods. In order to implement the semantic information into the overall FRIS design, these actual domain experts had to be involved in an easily accessible method. This is realized by the addition of a semantic layer on top of the FRIS application and technological layer, resulting in a three-tier architecture (*Figure 3*). The topmost tier of this architecture comprises the business layer in which the semantics of all concepts on research information is being governed. The application layer involves the services and interfaces of the FRIS portal and the technological layer involves the actual data, the software and the application programming interface (API). This three-tier architecture offers many advantages to FRIS 2.0 because of the clear separation between the different layers, yet the capability to maintain formal relationships with the concepts within as well as across the layers. Most importantly, this ensures that due to traceability across the layers, the impact of changes in one place can be easily be evaluated which supports a scalable, better maintainable and future-proof FRIS system.

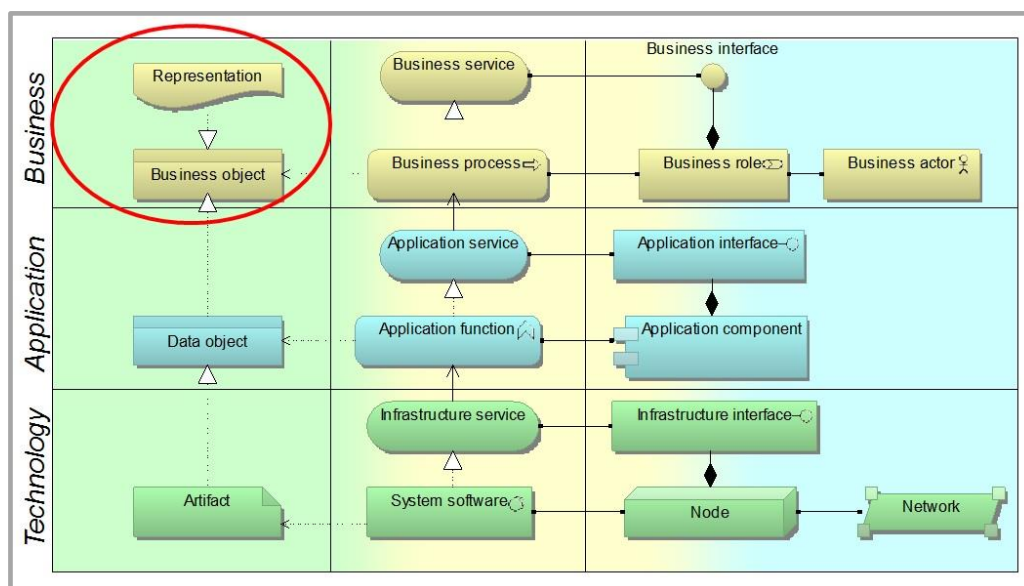


Figure 3: Three-tier architecture of FRIS 2.0

Although the three-tier model provides domain experts with an easier access to describe and manage their own research information without in-depth knowledge of the underlying information model, it is not yet sufficient for creating a true semantic interoperable research information system. This can only be reached when all information providers involved, generally agree upon a common ontology that renders the data meaningful provided by these, mostly autonomously developed information systems. Therefore, the actual domain experts have to be involved as they have the knowledge of the specific meaning of a piece of information and classifications used. However, the domain experts consist of a large and heterogeneous group, ranging both in

terms of the content owned as well as their physical location because of their affiliated institutions. In order to create semantic communities of topic-related business experts, a *data governance tool* named Collibra® Data Governance Center (DGC) was added to the overall FRIS architecture (**Figure 4**). DGC is an online software platform with a suite of data stewardship applications. The latter allows domain experts (in practice referred to as ‘data stewards’ (Plotkin, D., 2014) to create meaning agreements on concepts in an easy accessible, collaborative and machine-readable manner.

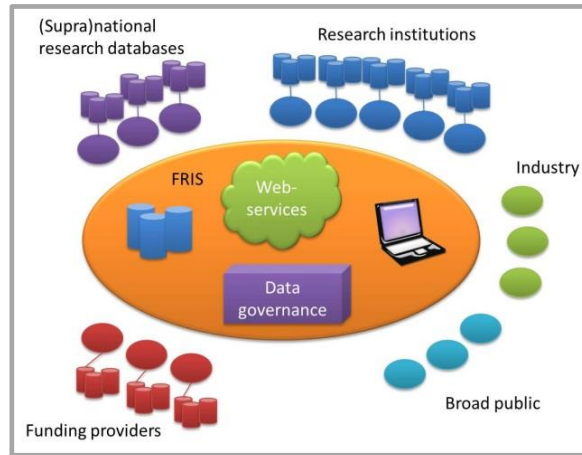


Figure 4: Implementation of a data governance tool in the FRIS 2.0 design

Foundational to DGC is a fact-oriented operating framework¹ and modelling methodology, based on Semantics of Business Vocabulary and Business Rules (SBVR) to capture concepts and their relationships in facts (Verheijen, G. *et al.*, 1982; De Leenheer, P. *et al.*, 2010; OMG SBVR). As such, all information providers can feed their own business models and classifications for describing research information together with the institution-specific semantics. Following the business semantics management method (De Leenheer, P. *et al.*, 2010), the information suppliers are then virtually brought together in the data governance tool to discuss and agree on the conceptual FRIS metamodel together with the meaning of the concepts used. By running the process in iterative cycles, the method achieves to align business concepts, their semantics and corresponding representations between organizations and FRIS in a collaborative and dynamic manner (**Figure 5**). Importantly, this model can be converted into a CERIF-based E-R model, that is implemented as optimized MySQL database tables and the accompanying ontology can be exported into an implementation in RDFS/OWL (Van Grootel, G. *et al.*, 2009; Spyns, P. *et al.*, 2011; Debruyne, C. *et al.*, 2011).

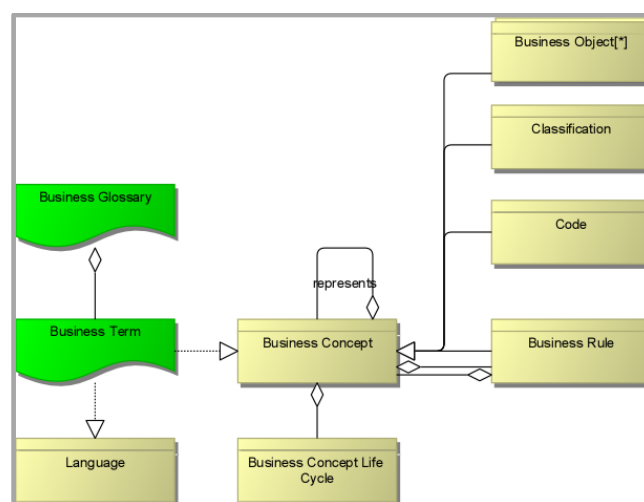


Figure 5: Alignment of the institution-specific models with the FRIS metamodel

¹ <https://compass.collibra.com/display/COOK/Data+Governance+Operating+Model>

By introducing model driven formal semantic management at the business level, the FRIS 2.0 methodology tries to create a complete and managed modeling stack from business to the technical level (**Figure 6**). As the FRIS data governance platform is used to formally represent and relate all necessary representations of business concepts and the technical representation, this approach allows for a traceability across the complete FRIS 2.0 model stack, from the different business objects to their CERIF ER and CERIF XML representations and back. The FRIS 2.0 environment is even further strengthened as the information can be extracted in a machine-readable manner leading to a plethora of webservices that can be created around the data. These services can assist in reducing the administrative burden accompanying research reporting and open avenues for developing master data webservices on concepts like organizations, journals, persons, projects,... that are being standardized within FRIS 2.0. Other possibilities include improved knowledge discovery via the FRIS research portal which could act on the long term even as a linked open data store (Dimou, A. *et al.*, 2014).

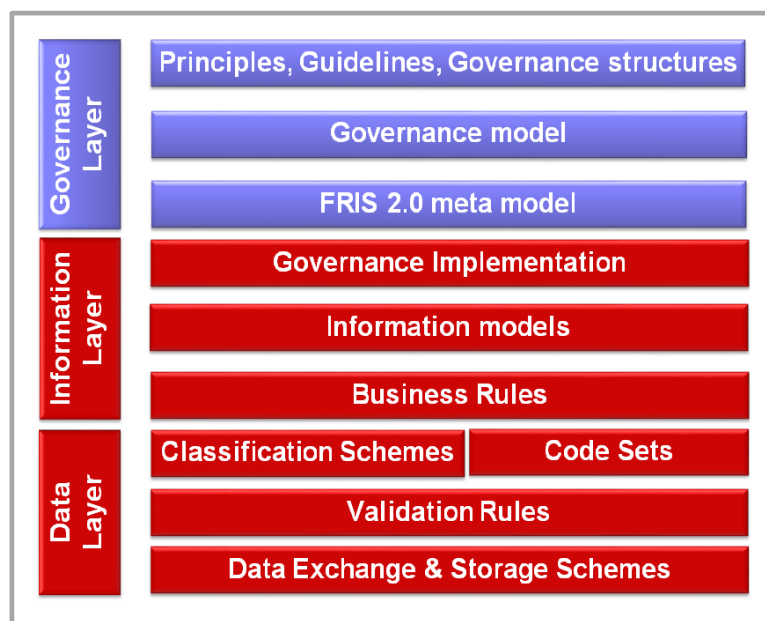


Figure 6: High level view on the FRIS 2.0 modelling stack

From design to implementation

Using this design, FRIS 2.0 has adopted an architecture that incorporates data and classification governance as a prerequisite for a semantic interoperable CRIS. Although this offers many benefits, several practical challenges still need to be met when implementing such a data and classification governance framework. First, a data governance board had to be installed where all stakeholders are aware of the roles and responsibilities of every person. While rather informal rules on responsibilities can be sufficient in smaller groups, the involvement of a many, inter-organizational information providers demands a more formal process to track responsibilities and reduce confusion. To this end, the data governance tool uses the RACI responsibility assignment matrix which assigns specific roles as responsible, accountable, consulted and informed to (groups of) individuals. In brief, the persons consulted for delineating definitions for research information, which sometimes even still need to be formalized, are termed subject matter experts. The definitions itself are fed into the governance tool by so-called stewards, comparable with the responsible role of the RACI matrix. The tool itself is maintained by a data governance administrator also denoted as accountable and all stakeholders can grasp the information kept, representing the informed role. Obviously, the establishment of such a data governance board that tackles each piece of research information requests the involvement, responsibility and thus investments of all people involved.

Another issue that had to be tackled, involved the drafting of data governance processes and workflows that support the entry, approval and automated provisioning of business and data concepts to the FRIS metamodel. These are extremely valuable for realizing agreements on the meaning of concepts between geographically separated communities in a robust, fast and more cost-efficient manner. As such, the use of these workflows guarantees that all concepts go through a similar automated and qualitative approval process, before being implemented. Furthermore, additional workflows had to be designed that allowed adaptations to the resulting business glossaries in a traceable and dynamic manner.

Next, business rules had to be created in a collaborative manner that basically determine the content and format of the information being exchanged by the information suppliers. Obviously, this imposed adaptations to the information export methods as used by the IT professionals. Likewise, validation rules had to be redesigned or newly created according to the business rules and incorporated into the information exchange process. Altogether, the realization of FRIS 2.0 requested a serious investment of the information providers, which prompted the Flemish government to establish an impulse financing program that actively supports the information providers to achieve these goals.

Classification governance

Another important issue that had to be dealt with, included the incorporation of semantically agreed upon classification systems which are crucial for retrieving comparable information from the FRIS portal. To this end, the Centre for Research and Development Monitoring (ECOOM) was contracted, which is a interuniversity consortium with participation of all Flemish universities that develops indicator systems for R&D and innovation monitoring. These ECOOM activities support the Flemish government in its ambition to perpetuate Flanders as a leading, innovative intensive region. In line with these objectives, ECOOM is also responsible for the development and dynamic management of classification systems for the monitoring of the Flanders' research project portfolio in terms of funding, publication and innovation output, and its assignment to scientific disciplines. Obviously, the addition of a semantic layer to these classification systems in a collaborative and governed manner, renders these classification systems into powerful tools for overseeing the Flanders' R&D and innovation potential in a comparable, meaningful manner. Moreover, the integration of such classifications into the FRIS 2.0 environment also supports other webservices, i.e. automated research reporting, the development of master data services, ... which further boosts the FRIS 2.0 environment.

Although many research classification systems already existed in FRIS 1.0, none of them included *semantic classification governance*. This discipline, like information governance, comprises the specification of decision rights and an accountability framework that encourages desirable behaviour in the creation, storage, use, archival and deletion of research classification systems (Logan *et al.* 2010). 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. In this paper, the creation of such a semantically governed classification system on research funding is exemplified.

The Flanders research funding landscape

In Flanders, scientific research is carried out by a wide variety of organizations ranging from the 5 Flemish universities, to university colleges, strategic research centers (VIB, VITO, IMEC, iMINDS, Flanders' Make²), research institutes (INBO, ILVO, ITG, KMSKA, ...³) and other independent research organizations (VLIZ,...⁴). These organizations differ in many aspects ranging in terms of focus – fundamental or applied research - and topics being investigated, their mission as well as the degree of (in)dependency. The research funding allocated to each of these institutions even comes with a bigger variety. As such, research funding is primarily provided by the Flemish government either directly, via ministerial assignments or governmental agencies, of which the Flanders Innovation and Entrepreneurship Agency (AIO) is the most known one. Alternatively, the research funding can be granted indirectly to researchers via independent agencies, like the Research Fund Flanders (FWO), that organizes the entire funding cycle from call to research report and divides the obtained funding sources based on a stringent evaluation process. Furthermore, a substantial amount of research funding comes from the European level via framework programs like Horizon 2020, the national level, i.e. via the Belgian Science Policy Office (BELSPO) and a large number of international organizations. Additional resources come from the private sector, although more focused on close-to-market research rather than blue-sky ideas or technologies. Finally, donations from charity organizations and even legacies, especially in relation to non-curable diseases, further augment the total amount of research funding that is available for the support of research projects, strategic research programs, scholarships and research infrastructure.

² VIB: Flemish Institute for Biotechnology, VITO: Flemish Institute for Technological Research, IMEC: Interuniversity Micro-Electronic Centre

³ INBO: Research Institute for Nature and Forest

⁴ VLIZ: Flanders Marine Institute, ILVO: Institutes for Agricultural and Fisheries Research, ITG: Institute of Tropical Medicine Antwerp, KMSKA: Royal Museum for Fine Arts Antwerp,

The FRIS 1.0 research funding classification scheme

In order to get an overview of the funding sources and their respective contribution to research in Flanders, a classification scheme for research funding was designed by the steering group of the Inventory of Scientific and Technological Research Flanders (IWETO) which was implemented in the FRIS 1.0 environment. This research funding classification used a hierarchical structure consisting of 3 levels deep, where each level aggregated the funding origin to a different amount of detail. This hierarchy was also reflected in the corresponding code scheme, which consisted of 4-digit code sets (*Table 1*).

Funding Code Term	Funding Code	Funding Program	Start Date	End Date
1. Own means				
2. Institutional subsidies (of the governmental authorities)				
3. International means				
3.7 EU – FP7 Program for O&O 2007-2013				
	3701	Health	01/01/2007	31/12/9999
3.75 EU - H2020 Program for O&O 2014-2020				
	3750	H2020 - Frontier Research (ERC)	01/01/2014	31/12/9999
	3752	H2020 - Skills and Career Development (MC) actions	01/01/2014	31/12/9999
3.8 EU - Programs and initiatives outside FP				
	3800	EU - out of framework - indefinite	01/01/1980	31/12/9999
3.9 Multilateral International Collaboration outside EU				
3.91 International European Institutions (outside EU)				
	3912	COST	01/01/1980	31/12/9999
4. Public Funding by the Federal Government				
5. Public funding by the regions/communities				
6. Financing by decentralized governments				
7. Research Foundation Flanders				
8. Private national funding				

Table 1: Extract of the FRIS 1.0 Funding Classification scheme

Although this code scheme had practical advantages in terms of ease of use by research administrators, it simultaneously imposed several constraints on the system. For instance, a given funding source could only be allocated to a single hierarchical level, which is not reflecting the true nature of research funding. For instance, European research programs occasionally integrate their funds with governmental ones. The representation of these funding programs by a distinct hierarchical funding code not only limits the viewpoint to a single funding origin, but also creates flaws when these code sets are used for reporting purposes. Furthermore, the use of such a hierarchical 4-digit code set, limits the amount of funding sources that can be positioned at each level. Consequently, the deepest level can only contain 9 funding programs, causing problems when a higher number of sources is present. Moreover, even the higher hierarchical levels are confronted with these problems, which is exemplified by the funding code “3.75”. The latter code represents the funding provided by the EU – Horizon2020 framework program, which should have been denoted as 3.8 according to the hierarchical structure. However, as this code already existed when the framework was installed, the steering group was forced to find a creative solution, in contrast with the conceptual guidelines. Other issues that were raised on the funding classification scheme included the decentral maintenance in spreadsheets, which resulted over time in a small number of similar, though slightly different code sets that were used by the information providers. Moreover, the use of spreadsheets did not allow to track the time when changes occurred to the code scheme nor whom was responsible. Finally, the FRIS 1.0 research funding classification scheme lacked a well-defined accompanying semantic vocabulary, which is an absolute prerequisite if one wants to ensure the identical use of the code scheme in an inter-organizational context.

Classification governance of research funding code schemes

In order to overcome the limitations and issues accompanying the FRIS 1.0 research funding classification, ECOOM-UHasselt was contracted for designing a semantic interoperable classification scheme suitable for information retrieval and reporting purposes. Altogether these requirements prompted a complete

revision of the classification design and its corresponding research funding model, as the classification system would be implemented in the FRIS 2.0 environment, which similarly also includes the principles of data and classification governance.

In brief, the new classification list comprises a non-hierarchical list of funding programs encoded by 5-digit code sets. These code sets are defined by means of auto-numbering, thereby creating endless possibilities for further growth. Besides, dimensions are added to the classification scheme, which are basically characteristics of each funding program that are in direct relation with the requirements put forward by regional, national and international reporting obligations. As such, these dimensions not only facilitate information and reporting services in the FRIS 2.0 environment but also satisfy the needs of research administrators to keep a good overview of the codes. In addition, the funding programs are encoded with a higher degree of granularity in order to allow for more detailed and thus accurate information.

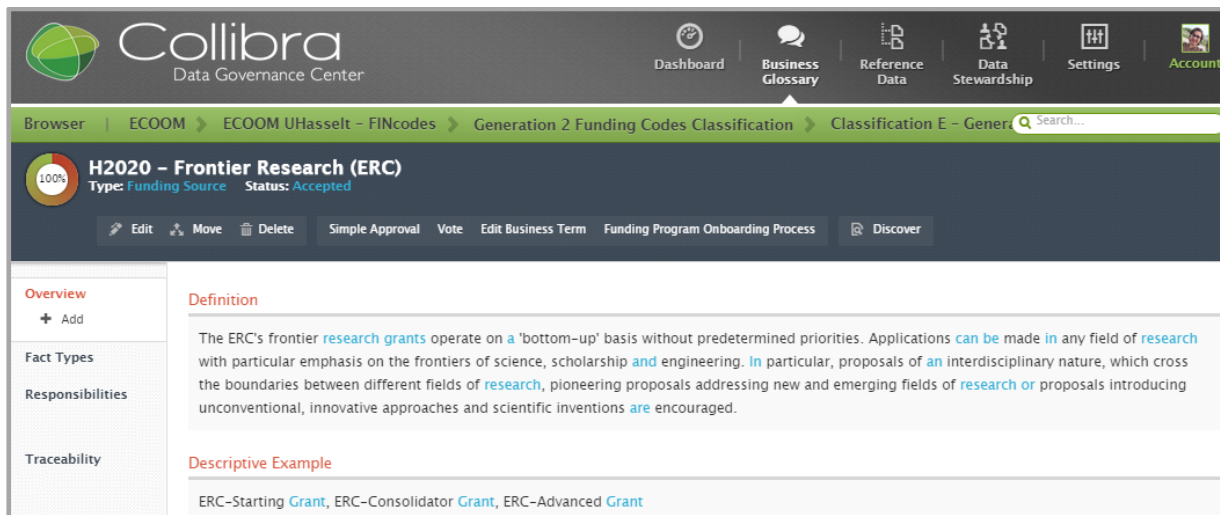


Figure 7: The Collibra® Data Governance Center as a governance tool for research funding classifications.

Furthermore, a data governance tool is used in order to realize the unambiguous definition of funding concepts and codes in a collaborative, traceable and machine-readable manner (**Figure 7**). To this end, the Data Governance Center of Collibra® is implemented as it comprises both a business semantic glossary in combination with a data stewardship tool. This allows for the organization of a data governance board carrying out the different roles of the RACI matrix when defining funding concepts, codes and the classification scheme, and assists in keeping track of the changes that occur together with the persons responsible. Additionally, DGC allows for the design and inclusion of workflows that support the governance and information retrieval process in an automated, efficient manner. Finally, the tools also comprises a reference manager tool which can be used for the creation of concordance tables to existing (inter)national classifications.

Conclusion

Over the past decades, the research ecosystem has vastly expanded, resulting in a massive increase in the amount and nature of information produced as well as the systems used for information storage, ranging from sophisticated research information systems, to databases and even spreadsheets. Simultaneously, the demands on the research community to report on their findings to funding providers, authorities, the research community and the public at large have increased tremendously, resulting in a multitude of largely similar reports. Despite of this abundant reporting, a global overview of the complete scientific and technological portfolio was lacking in Flanders, which is an absolute prerequisite if one wants to transfer the obtained knowledge to industry and the broad public. This prompted the Flanders Government, to create the FRIS-portal which makes Flemish research information publicly available.

In its first realization, the FRIS portal primarily acted as an information retrieval platform, displaying information on researchers, their affiliated organizations and research projects, which is obtained directly from universities via exchange through CERIF-xml. Although CERIF, the European standard for modelling research information with declared semantics on all concepts used, offers many advantages in terms of large flexibility for modelling, it requires end-users to use exactly the same terminology and corresponding meaning as in the CERIF scheme. Information providers however, often use a different terminology for a similar concept or alternatively,

use a similar terminology for a different concept. In order to ensure data communication in the same language, a semantic layer needs to be added on top of the data exchanged by the information providers, which is realized by implementing data governance. This specific focus on semantic alignment adds further to existing initiatives like VIVO⁵ and CERIF based CMS (Guéret *et al.*, 2013). CERIF and VIVO are in essence technical formats geared towards data representation in a specific implementation context. In concrete, this includes the Entity Relation methodology for CERIF and Web Ontology in the case of VIVO. Both VIVO and CERIF suffer from the lack of coherent logical and conceptual model layers. In the case of CERIF, the logical level is a part of the naming conventions identical to the physical level and the CERIF semantics is a list of possible values for the different classification vocabularies. This leads to the fact that both standards have little to no control on the accurate implementation of the business meanings. Nevertheless, both technical standards are specifically aimed at interchange of research information across different organizations and systems. The weak semantic control has led in Flanders in the past to CERIF data that is technically coherent but with large semantic discrepancies making querying and analysis of data across different CERIF sources at least very difficult. The FRIS 2.0 semantic alignment reduces these risks considerably as it allows for far more semantic alignment than the actual practice of CERIF and VIVO. The implementation of the FRIS 2.0 methodology in the FRIS environment, a CERIF-based system, will undoubtedly augment data quality and allows for much more precise querying across CERIF systems.

As such, the implementation of data governance opens the door to applications, like information retrieval in a transparent and automated manner for research reporting purposes. Research reports often come in a wide range of flavours, each with their own requirements with regards to the formats and classifications used, thereby creating a multitude of largely similar research reports. Obviously, this places a large burden on the research community and thus reduces the value for money in terms of productivity of a researcher. In line with the growing concern on the administrative burden put on the research community, a report was published by Peters *et al.* (2011) providing guidelines for the reduction of redundant research information reporting. One of these advices, included the development of classification governance on the classifications used for reporting purposes, i.e. publications, funding sources, science disciplines and technological applications. Until recently, funding providers and authorities were using their own classification schemes which were semantically poorly defined and lacked concordance mappings to other (inter)national classifications. This resulted in dissimilar reporting by the data providers, thereby lowering the quality of the data and the interoperability of research information. Data unambiguity is becoming increasingly important, in an era where many initiatives have seen light to measure and benchmark research and where public research reporting obligations are vastly growing. Obviously, the lack or incomplete definition of semantics puts large constraints on the interoperability of research information, and in extension on the policies drawn out of these data.

In order to implement data and classification governance in the FRIS 2.0 environment, the Collibra[®] data governance tool has been used. This tool comprises a business semantics glossary, a data governance board manager and a reference data module which facilitates the creation of meaningful concordance tables. The business semantics glossary has been used for the description of each individual component of the FRIS research model. By explicitly defining all concepts, the governance tool assists in the swift identification of semantic inter-organizational misalignments when mapping corresponding concepts by stakeholders. The resulting ontologies can be exported and used to annotate data in relational databases, and hence render data meaningful. Similarly, the DGC tool has been used for defining the semantics of classifications and code sets on research information, which is essential when it comes down to consistent and unambiguous reporting to third parties. Obviously, the research community at large will benefit from this, as the information retrieved via FRIS will be much more reliable and accurate which improves its business intelligence capacity.

Altogether, the use of a data governance tool for data and classification governance, and its implementation into the FRIS 2.0 environment opens new avenues in terms of efficiency of the research ecosystem. Not only will governments be able to delineate better founded policies, also research administrations and researchers themselves can gain tremendously as research reporting could be automated from the FRIS-portal in a reliable and accurate manner, thereby reducing the administrative burden at the benefit of scientific discovery and innovation.

Acknowledgements

⁵ VIVO is a tool for representing information about research and researchers, their scholarly works, research interests and organizational relationships (Börner, K. *et al.* 2012). To this end, VIVO provides an expressive ontology, tools for managing the ontology, and a platform for using the ontology to create and manage linked open data for scholarship and discovery.

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

- Börner, K., Conlon, M., Corson-Rikert, J., Ding, Y. (2012) VIVO: a semantic approach to scholarly networking and discovery. In: *Synthesis lectures on the semantic web, theory and technology*, 1-160
- 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. In: *Computers in Industry*, 61(8):760-775.
- Dimou, A., De Vocht, L., Van Grootel, G., et al. (2014) Visualizing the information of a Linked Open Data enabled Research Information System. In: *Procedia Computer Science*, 33, 245-252
- DRIS. (2015) Retrieved, January 8, 2016. From: <http://www.eurocris.org/activities/dris>
- EuroCRIS. (2013). Statistic Information. CERIF 1.6. Retrieved August 13, 2013, From: <http://www.eurocris.org/cerif/feature-tour/cerif-16>.
- EuroCRIS. (2015) Retrieved, January 1, 2016. From: <http://www.eurocris.org/why-does-one-need-cris>
- 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/
- Guéret, C., Chambers, T., Reijndoudt, L., et al. (2013) Genericity versus expressivity – an exercise in semantic interoperable research information systems for Web Science. In: <http://arxiv.org/pdf/1304.5743.pdf>
- Jörg, B., Höllrigl, T., Baker, D. (2014) Harmonising and formalizing research administration profiles CASRAI/CERIF. In: *Procedia Computer Science* 33, 95-102
- Jörg, B. (2009) CERIF: Common European Research Information Format. From contextual relations to guide through the maze of research information. In: *Proceedings International Conference Research Information Systems in the EU*. ISBN 978-80-89354-02-03
- OMG SBVR, In: <http://www.omg.org/spec/SBVR/>
- 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)
- Plotkin, D. (2014) Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance. ISBN: 978-0-12-410389-4
- Spyns, P., Van Grootel, G. (2011) Realising the Flanders Research Information Space. In: *On the Move to Meaningful Internet Systems: OTM 2011 Workshops. Lecture Notes in Computer Science*; 7046, 138-141
- Van Grootel, G., Spyns, P., Christiaens, S., Jörg, B. (2009) Business semantics management supports government innovation information portal. *On the move to meaningful internet systems*. In: *On the Move to Meaningful Internet Systems: OTM 2009 Workshops. Lecture Notes in Computer Science*, 5872, 757-766
- Verheijen, G., Van Bekkum, J. (1982) NIAM, an information analysis method. In: *Proceedings of the IFIP TC-8 Conference on Comparative Review Of Information System Methodologies (CRIS 82)*