Semantics, a Key Concept in Interoperability of Research Information -The Flanders Research Funding Semantics Case

Sadia Vancauwenbergh

Sadia.Vancauwenbergh@uhasselt.be
ECOOM-UHasselt, Hasselt University, Research Coordination Office, Martelarenlaan 42, BE-3500 Hasselt (Belgium)

Introduction

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 (FRIS, http://www.researchportal.be) that integrates information available from its data providers (research institutions, funding organizations…) using the CERIF (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. This paper describes the implementation of a business semantics tool on data concepts and classifications for research funding as a means to unambiguously exchange and interpret these data.

The need of semantics

A couple of decades ago, the demands on the research community to report on research data were rather low. Results were published in preferably highly-rated journals and rather limited research reports were written. Over the years, more research data became available and the need for research databases grew. Unfortunately, these databases were predominantly developed per organization without consultation of other organizations. Moreover, because of the rather low data volume and people involved, there seemed no explicit need for defining an accompanying semantics.

However, as the research system expanded, there has been 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. In Flanders, this resulted in the creation of the FRIS-portal which makes Flemish research information publicly available. This information is provided via a multitude of data providers that often use a different terminology for a similar concept or alternatively, use a similar terminology for a different concept. The correct interpretation of the information at the FRIS portal is realized by the addition of a semantic layer on top of the data by the data providers, which later on is translated to a general FRIS semantics resulting in data communication in the same language. The focus on the explicit semantic alignment with the data providers, adds further to existing initiatives like VIVO and CERIF based CMS (Guéret et al., 2013). Data unambiguity is 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.

The Flanders research information landscape

In Flanders, research institutions receive funding from a broad range of research funding providers going from the regional to national and international level. Obviously, each funding provider has its own requirements with regards to the formats or classifications used for reporting on the resulting research output, thereby creating a multitude of largely similar research reports. Obviously, this places a large burden on the research community. Until now, the data providers tried to keep pace with this vast expansion of research reporting by improving or even creating databases, unfortunately without generally agreed upon semantics. At the same time, the data providers were feeding their information to the FRIS-portal in order to increase the visibility of the research in Flanders to third parties (i.e. companies, research institutions and individual researchers).

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. Following these advices, the Flemish Department of Economy, Science and Innovation (EWI) is currently improving the FRIS-portal in order to be used as a virtual research information space, for information retrieval in a transparent and automated manner that can be used for research reporting (Figure 1) (Debruyne et al., 2011). This implicates

the use of unambiguous data concepts and research funding classifications. Until recently, funding organizations were using their own funding classification schemes which were semantically poorly defined and lacked concordance mappings to other (inter)national classifications.

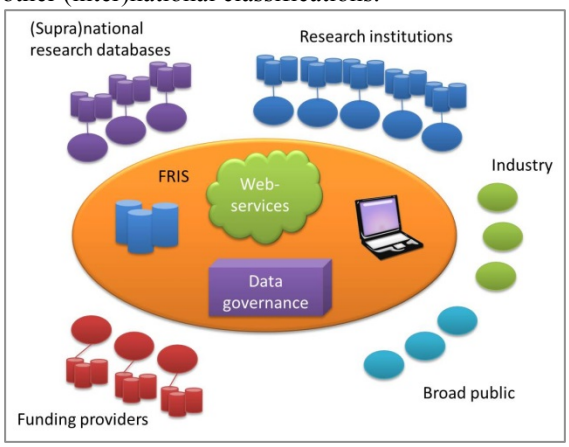

Figure 1: Representation of the FRIS design.

Funding data and classification governance

In order to add a semantic layer on top of the FRIS database layer, the Data Governance Centre® (DGC) platform of Collibra has been used. This platform allows data suppliers to manage their own data models used to describe, i.e. research funding together with the corresponding institution specific semantics (Figure 2).

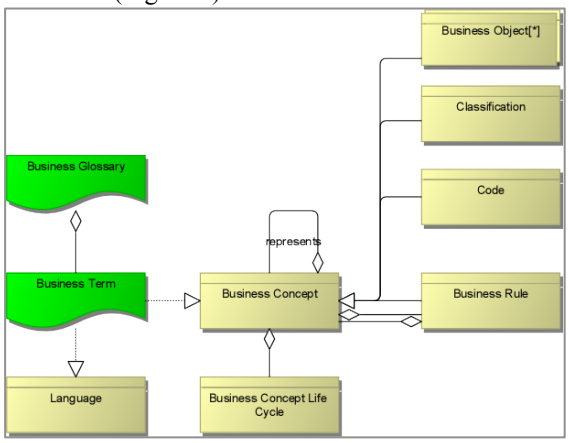

Figure 2. Incorporation of a business semantics glossary on the research funding model.

At the same time, the DGC platform has been used for the description of each individual component of the FRIS research funding model using definitions (Figure 3). By explicitly defining all concepts, the governance tool assists in the swift identification of semantic inter-organizational misalignments when mapping corresponding concepts by the stakeholders. The resulting ontologies can be exported and used to annotate data in relational databases, and hence render data meaningful. Furthermore, the DGC tool has been used for defining the semantics of classifications and code sets on research funding, which is essential when it comes down to consistent and unambiguous reporting on research funding 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.

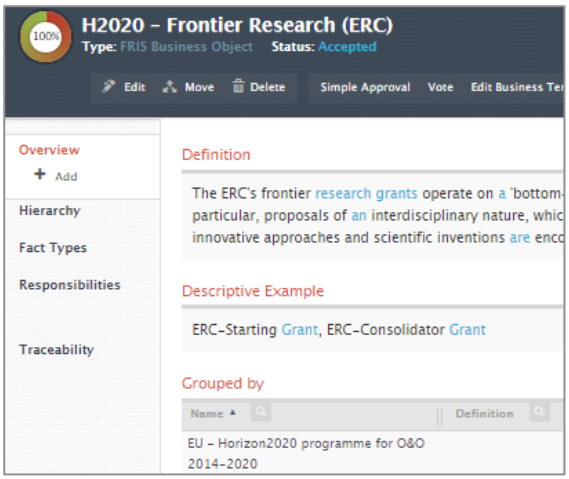

Figure 3: DGC as a governance tool for research funding classifications.

Altogether, the use of a data governance tool focused on semantics 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 manner, thereby reducing the administrative burden at the benefit of scientific discovery and innovation.

Acknowledgement

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

Debruyne, C., De Leenheer, P., Spyns, P., Van Grootel, G., & Christiaens, S. (2011). Publishing open data and services for the Flemish research information space. In *Advances in Conceptual Modeling. Recent Developments and New Directions,* Springer.

Guéret, C., Chambers, T., Reijnhoudt, 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*

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).