13th International Conference on Current Research Information Systems, CRIS2016,
9-11 June 2016, Scotland, UK

Modelling National Research Information Contexts based on CERIF

Christoph Quix[a,b,*], Mathias Riechert[c,d]

[a]Fraunhofer Institute for Applied Information Technology FIT, 53754 St. Augustin, Germany
[b]RWTH Aachen University, 52056 Aachen, Germany
[c]German Centre for Higher Education Research and Science Studies (DZHW), 10117 Berlin, Germany
[d]Technical University of Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany

Abstract

We present the results of the data modelling in the project 'Research Core Dataset' (2013-2015) that provided the context for developing a set of core definitions for research information for the German science system. In this paper, we focus on the data modelling aspects of the project, whereas another submission focuses on the management of the discussion phase and visualization of the argumentation process in the project. We present how the data model has been developed and synchronised with the argumentation process. As compatibility with CERIF was a major requirement for the data model, we present our approach to link the data model to the CERIF standard.

Keywords: research information; CERIF; data modelling

1. Purpose

Modelling a standardised data model for a national context of research information is a complex challenge, especially if the interests of hundreds of stakeholders in research information management are to be taken into account. Within the German project for defining a 'Research Core Dataset' for the German science system this national context has been specified for the German Science System[1]. In this paper, we report the results with respect to the data modelling part and its relationship to CERIF. A related paper presents the framework for documenting the standardisation process[2].

Harmonising the reporting of research information on a national level requires the agreement of the participating stakeholders and 'clearly defined requirements' for a sustainable conceptual model[3]. To achieve this goal, we characterised the standardisation process for research information as a wicked problem[4,5], i.e., a problem that does not have a clear definition and its solution cannot be true-or-false, but only good-or-bad. For wicked problems, the focus is

E-mail address: christoph.quix@fit.fraunhofer.de

* Corresponding author

rather on the problem-solving process instead of the final solution[6]. The solution of wicked problems involves various stakeholders who exchange arguments pro and con certain issues.

Thus, the visualisation of the argumentation and discussion process is an important issue for the acceptance of a data model for research information. In[7], we presented a framework in which techniques from CSAV (Computer Supported Argumentation Visualization,[8]) and IBIS (Issue-based Information Systems,[9]) are combined to produce a comprehensive but still comprehensible documentation of the argumentation process.

The result of the framework is a set of definitions for elements of a data model for research information in a semi-structured form[a]. The arguments are linked to the corresponding definitions. Although the elements and their definitions have been captured in a tabular form, the definitions are still insufficient for a formal exchange of research information. Therefore, the project included a working group that aimed at the translation of the definitions into a formal data model, which should guarantee at the same time that the new standard can be implemented with reasonable efforts in current research information systems.

2. Approach

From the perspective of a data modeller, the definitions are the requirements for the data model and, thus, form an important input for the data modelling process. The data modelling process was done in a classical way: we created first a conceptual data model and then mapped it to a logical data model[10]. But why is it necessary to create new data models for research information although there is already the CERIF standard?

2.1. Limitations of existing standards

CERIF provides a wide-ranging model for research information. It provides definitions for most elements which are relevant for research information management. However, as a European standard, CERIF cannot provide all the details which might be relevant only in one country. This requirements-driven approach has also been pointed out in a recent study that emphasised the need to standardise research contexts in contrast to only standardise research information[3].

Furthermore, the definitions in CERIF are rather 'abstract' as the standard should be applicable to several use cases. A specific contextual semantics that defines the elements with respect to a given application context is not part of the standard[3]. For example, it has been not specified which kind of projects should be considered as an instance of 'cfProject' or which prizes are relevant for 'cfPrize'.

A similar question as for CERIF could be also raised for CASRAI. CASRAI maintains and develops profiles for research administration information[3]. The process of defining research information based on business needs is similar to our approach to model research information based on requirements. At the time of our project, the existing definitions in CASRAI and their underlying business needs did not match the requirements for the German Science System which expected more detailed definitions and data models.

The definition of these contextual boundaries was the main goal of the discussion and argumentation phase. This included, for example, the discussion about which staff categories should represented in the German core dataset, or which types of projects should be considered. In addition, classification, categorisation, and attributes of the research information entities have been discussed.

2.2. Representation of the Definitions in the German Research Core Dataset

As mentioned above, this information has been captured in a semi-structured form, which could be visualised in tables[b], argumentation graphs, and other more aggregated views[7,2].

---

[a] The resulting table is available at `http://kerndatensatz-forschung.de/version1/Spezifikationstabelle_KDSF_v1.html`.

[b] In addition to the table with the final definitions (`http://kerndatensatz-forschung.de/version1/Spezifikationstabelle_KDSF_v1.html`), there is also a table with all discussed elements (`http://kerndatensatz-forschung.de/version1/Spezifikationstabelle_KDSF_v1_komplett.html`) which includes the elements that have been rejected or postponed.