Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 106 (2017) 329 – 334

Procedia Computer Science

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

Say, "S" (as) semantics – and mean it! Path to semantically interoperable digital research services

Suvi Remes[a]*, Miika Alonen[a], Patrik Maltusch[b], Mikael af Hällström[c], Stina Westman[a]

[a]*CSC – IT Center for Science, Keilaranta 14, Espoo FIN-02101, Finland*
[b]*Aalto University, PO BOX 11000, Aalto FIN-00076, Finland*
[c]*Finnish Tax Administration, PO BOX 325, Vero FIN-00052, Finland*

Abstract

The more we invest in open science and research, the more we need to ensure that metadata enabling discovering and digital preservation of research material is of high-quality and semantically coherent. Still, interoperability of information systems and the lack of shared semantics, both between humans and machines, is an internationally recognised issue.

In Finland we are in the process of implementing information systems and harmonising the legacy data models in the way that it makes use of the shared semantics, standards and other best practices according to the common architectural vision. This basic infrastructure for information management is built by combining terminological theory, linked data and adaptable data modelling practices. The idea of the Semantic Interoperability Model and new tools, IOW – Interoperability Workbench, supporting it are presented in the context of research and science in Finland, but the vision of the linked information components is generic.

Peer-review under responsibility of the Organizing Committee of CRIS2016

Keywords: semantic interoperability; terminological theory; linked data; metadata modelling; information architecture

1. Introduction

Many efforts have been made to define best practices and guidelines for interoperability in the field of research. In Finland, a national data model has been defined to support data storage, data collection and reporting activities in the

* Corresponding author. Tel.: +358-50-320-4792; fax: +358-9-457-2302.
E-mail address: suvi.remes@csc.fi

Peer-review under responsibility of the Organizing Committee of CRIS2016
doi:10.1016/j.procs.2017.03.032

field of higher education and research administration. The XDW model[1] is designed to support institutions in implementing a data warehouse. Data in the data warehouse can be any functional data, ranging from students' grades to financial indicators or research publications. Mapping from the XDW model to the CERIF model[2], developed by the euroCRIS community, was studied on the conceptual level, but formal mapping is not supported in the current XDW implementation.

The loosely CERIF compatible XDW model itself has been used as a starting point in data warehouse implementations. This kind of a database approach to information management is typical, but architecture based on traditional databases which have their own data models require lot of harmonisation and high cost expert labor work to move the data from one information system to another. Currently more focus is needed in support for more adaptable data modelling and invocation of linked data practices, and on development of interfaces and data transfers between information systems — all including semantics.

The new approach (Fig. 1) to data management and interoperability we describe in this paper is a mix of existing practices, such as utilising core vocabularies as proposed by Charalabidis et al. (2010)[3] and documenting the use of such data vocabularies by defining application profiles as first defined by Heery and Patel (2000)[4]. Semantic technologies and linked data are used to describe machine-readable terminology, core vocabularies and application profiles. Expectations for the machine-readable application profiles as argued by Diane I. Hillmann and Jon Phipps (2007)[5] are realized by using shape expressions as described by Prud'hommeaux et al. (2014)[6]. Framework forms a structured, common architecture for connecting conceptual modelling of business, services and processes to defining and maintaining controlled terminology and further to constructing data models for information systems. It also offers Finnish agents operating in the field of research and science a new way to create and maintain linkages to relevant existing international work and resources.

2. Framework for information architecture

2.1. Concept modelling using terminological method

The modelling process of any information system should start from conceptual modelling. This is generally accepted statement, but according to our experience the actual meaning of it seems to vary. Moreover, what is usually missed at this stage is a systematic and formalised method for concept defining. It is not that we do not have them, it is that we have not fully recognised the value of terminological theory[7]. This originally humanistic approach, typically used in cross-human communication, argues that concepts within a subject field are interrelated and form concept systems. And we go on arguing that this shared understanding, the concepts we use in business and operations of agencies, should form the solid foundation also for semantics of data models used in information systems. The point is that these terminologies, resulting from terminological concept modelling, are domain specific, aimed at concept clarification, mutual understanding of concepts and consistent use of terms. They do not specify database or information system specific information or their mutual relationships, resulting from concept modelling[8]. And further, these terminological concept systems are adapted to the language and cultural realm they were created in. In multilingual terminologies and communication situations careful attention should be paid to glosses.

The controlled, methodologically intact terminology should be openly available both in human and machine readable formats. In Finland, terminologies are currently published mainly in printed or digital booklets or in term banks. These medias do not support machine readable formats, such as SKOS[9], required to apply terminologies as the cornerstone of semantics in information system development according to the architectural vision described in this paper. In Finland, however, situation is good as we already have a suitable service available: SKOS vocabularies and