

“RDF Vocabularies” Report 2012/13

Submitted by “Joachim Wackerow”, Chair of Sub-group on DDI-RDF Discovery Vocabulary

The charge of the group is to develop RDF vocabularies for efficient use of DDI metadata in the context of the Semantic Web / Linked Data.

Work is going on regarding two RDF vocabularies, DDI-RDF Discovery vocabulary, for publishing metadata about data sets into the Web of Linked Data, and XKOS, an RDF vocabulary for describing statistical classifications, which is an extension of the popular SKOS vocabulary.

There are two sub-groups, one focusing on the DDI-RDF Discovery vocabulary, one on XKOS.

Group Members of Sub-group on DDI-RDF Discovery Vocabulary

- Chair: Joachim Wackerow (GESIS - Leibniz Institute for the Social Sciences, Germany),
- Thomas Bosch (GESIS - Leibniz Institute for the Social Sciences, Germany)
- Sarven Capadislı (Bern University of Applied Sciences, Switzerland)
- Franck Cotton (INSEE - Institut National de la Statistique et des Études Économiques, France)
- Richard Cyganiak (DERI, Digital Enterprise Research Institute, Ireland)
- Arofan Gregory (ODaF - Open Data Foundation, USA)
- Larry Hoyle (University of Kansas, USA)
- Benedikt Kämpgen (Karlsruhe Institute of Technology, Germany)
- Olof Olsson (SND - Swedish National Data Service)
- Heiko Paulheim (University of Mannheim, Germany)
- Dan Smith (Algenta Technologies Inc., USA)
- Johanna Vompras (University Bielefeld Library, Germany)
- Benjamin Zapilko (GESIS - Leibniz Institute for the Social Sciences, Germany)

Group Members of Sub-group on XKOS

- Chair: Franck Cotton (INSEE - Institut National de la Statistique et des Études Économiques, France)
- Richard Cyganiak (DERI, Digital Enterprise Research Institute, Ireland)
- Daniel Gilman (BLS - Bureau of Labor Statistics, USA)
- Rob Grim (Tilburg University, Netherlands)
- Yves Jaques (FAO of the UN)
- Dan Smith (Algenta Technologies Inc., USA)
- Wendy Thomas (MPC - Minnesota Population Center, USA) Next member

This working group was built in the first quarter of 2013. The work on both vocabularies began in a workshop on "Semantic Statistics for Social, Behavioural, and Economic Sciences: Leveraging the DDI Model for the Linked Data Web" at Schloss Dagstuhl - Leibniz Center for Informatics, Germany, in

September 2011. This work has been continued at these three meetings: follow-up working meeting (Discovery vocabulary) at the 3rd Annual European DDI Users Group Meeting (EDDI11) in Gothenburg, Sweden, in December 2011; second workshop on "Semantic Statistics for Social, Behavioural, and Economic Sciences: Leveraging the DDI Model for the Linked Data Web" at Schloss Dagstuhl in October 2012; and a follow-up meeting (Discovery only) at GESIS-Leibniz Institute for the Social Sciences in Mannheim, Germany, in February 2013.

DDI-RDF Discovery Vocabulary

This specification is designed to support the discovery of microdata sets and related metadata using RDF technologies in the Web of Linked Data. The vocabulary leverages the DDI specification to create a simplified version of this model for the discovery of data files. It is based on a subset of the DDI XML formats of DDI Codebook and DDI Lifecycle. It supports identifying programmatically the relevant data sets for a specific research purpose. Existing DDI XML instances can be transformed into this RDF format and therefore exposed in the Web of Linked Data. The vice-versa process is not intended, as we have defined DDI-RDF components and reused components of other RDF vocabularies which make only sense in the Linked Data field.

XKOS

The use of the Simple Knowledge Organization System ([SKOS]) for managing statistical classifications and concept management systems is addressed, since SKOS is widely used. LOD is used to create Web artifacts that machines can interpret, so publishing machine readable statistical classifications and other concept management systems as SKOS instances is desired. We found that SKOS is insufficient for the problem. No aspect of SKOS was found to be wrong, just incomplete. Therefore, we propose an extension to SKOS, which we call XKOS.

XKOS extends SKOS for the needs of statistical classifications. It does so in two main directions. First, it defines a number of terms that allow the representation of statistical classifications with their structure and textual properties, as well as the relations between classifications. Second, it refines SKOS semantic properties to allow the use of more specific relations between concepts. Those specific relations can be used for the representation of classifications or for any other case where SKOS is employed. XKOS adds the extensions that are desirable to meet the requirements of the statistical community.

Current status

The work resulted in mature drafts of both specifications which are publicly available on a Github repository. The specifications were already presented on relevant conferences. A public review of the specifications is planned while 2013. It is intended to publish both RDF vocabularies as DDI Alliance specifications.

Work is currently in progress to integrate information on the RDF vocabularies on the DDI Alliance website. It will be available below a specific tab on RDF Vocabularies, at the working groups, and in future below Specification.