XKOS: An SKOS Extension for Statistical Classifications

Franck Cotton,
Institut National de la Statistique et des Études Économiques, France franck.cotton@insee.fr

Richard Cyganiak,
DERI, Ireland richard.cyganiak@deri.org

R.T.A.M. Grim,
Tilburg University, Netherlands rob.grim@uvt.nl

Daniel W. Gillman,
Bureau of Labor Statistics, USA gillman.daniel@bls.gov

Yves Jaques
Food and Agriculture Organization of the United Nations, Italy Yves.Jaques@fao.org

Wendy Thomas[*]
Minnesota Population Center, USA wlt@umn.edu

Resource Description Framework (RDF) based approaches to data management are growing in use and acceptance as governments and international organizations increasingly embrace the Semantic Web and Linked Open Data (LOD). This paper describes the efforts to marry LOD techniques to the needs of the international statistical community during two workshops on Semantic Statistics held at Schloß Dagstuhl, Leibniz Center for Informatics with the support of the GESIS Leibniz Institute for Social Sciences and the Data Documentation Initiative (DDI). Specifically, we address the extension of the Simple Knowledge Organization System (SKOS), an RDF vocabulary, to satisfy the requirements of classification systems and concept management for the statistical community.

The SKOS vocabulary was heavily influenced by the needs of thesauri, which rely on loosely defined notions of 'broader than', 'narrower than, and 'related to' relationships. Statistical classifications on the other hand use more formally defined hierarchical relations, which are referred to as generic (generic-specific) and partitive (whole-part). Further, statistical classifications, through their hierarchies, are structured according to levels that correspond to increasingly detailed views of the field covered. Finally, statistical concept management requires the use of associations that are more specific than the generic 'related to'. Causal, sequential, and temporal relations therefore need to be defined. XKOS addresses the greater detail required to manage statistical classifications by extending the existing SKOS definitions of object classes and relationships.

The proposed extensions to SKOS are guided by the needs of the statistical community as well as the requirements laid out in ISO standards on terminology, as reflected in ISO 704:2000 ([ISO704]) and ISO 1087-1:2000 ([ISO1087]). These standards describe and define the constructs and relations necessary for concept management and for a more complete description of statistical classifications. This paper describes concepts, their designating terms and codes,

their associative relations, their structures and finally the relationship between concepts and the real-world objects they classify. The goal is to provide a common means for making statistical classifications available through emerging web technologies, while also building on the RDF good practices of reuse.