Submitted on: June 4, 2013

Library linked data and its relationship to knowledge organization systems

Aida Slavic
UDC Consortium, Den Haag, The Netherlands.
aida.slavic@udcc.org

Christophe Gueret
DANS-KNAW, Den Haag, The Netherlands.
christophe.gueret@dans.knaw.nl

Chris Overfield
ISKO, United Kingdom.
chris@acorweb.net

Andrea Scharnhorst
DANS-KNAW, Den Haag, The Netherlands.
andrea.scharnhorst@dans.knaw.nl

Abstract:

One of the most exciting potentials of linked data development in the library domain lays in cross-collection information discovery that is made possible by linking library subject metadata and knowledge organization systems (KOS) in an open linked data environment. The benefits are twofold: users gain better access to information resources provided by libraries; and subject metadata in library collections can be enhanced.
This paper highlights some opportunities that are now opened to subject metadata exploitation. At the same time we also point to some as yet unresolved issues that need to be discussed in the open linked data standards communities.

Our focus is on linking bibliographic metadata containing classmarks on the one hand and classification systems published as linked data on the other. The problem we would like to focus on is connected to the evolution and changes in classification schemes and their discordance and mismatch with the subject metadata in library catalogues which often contain obsolete data or data from prior versions.

As an example we use the Universal Decimal Classification (UDC). But we would like to stress that the problem we observe is relevant and of interest to all those publishing KOS as linked data and all bibliographic subject metadata containing KOS.

Bibliographic classifications are continuously developed to follow scientific developments and accommodate new and emerging knowledge. The implementation of these changes in library collections is usually delayed due to practical difficulties when reclassification involves changes in large physical collections of documents. As a consequence when library subject data is published as linked data it will most certainly contain classmarks that are cancelled and replaced by some other classes in the latest edition of the system. Such was the case with UDC data in e.g. National Library of Hungary catalogue (OSZK), TEKORD data (University of Oslo Library) and the AGRIS collection when these were published as linked data.

The evolution of knowledge in the UDC scheme can be traced through historical revision data for over a century. But more interestingly for the past twenty-five years records of classes and concepts that have been moved or changed are available as machine readable records. In the UDC Master Reference File database we can search and present all classes that have been deprecated and their relationship to the revised structure.

In 2011 UDC Summary, an abridged schedule containing 2,600 classes was made available as linked data. In 2013 the UDC Consortium plans to publish the entire system of 70,000 classes and 10,000 cancelled i.e. historical classes, as linked data. But, the decision to publish this information as linked data has been postponed primarily because of the fact that the current SKOS format is lacking support for historical data and for the relationship 'replaced by'.

Using UDC as an example we discuss some solutions in managing and exposing 'deprecated' (i.e. historical UDC) terms as linked data in order to support linking with library catalogues. Our proposal discusses a solution that does not depend on the availability of 'classification provenance data' that may or may not be preserved in a bibliographic format or in subject authorities.

Keywords: Linked Open Data, Knowledge organization, UDC, libraries, knowledge spaces