

Michael Panzer, M.A. (OCLC Online Computer Library Center, Inc., Dublin, OH)  
**Towards the “webification” of controlled subject vocabulary: A case study involving the Dewey Decimal Classification**

Controlled vocabularies and other KOS terminologies are as ubiquitous in the library world as they are elusive in the world of the Web at large. Right now, with the maturing of the standards accompanying the Semantic Web, a strong case can be made to reconnect those vocabularies and taxonomies to the current of semantic technologies and services. To seize this opportunity, it will be necessary to take these resources to the network level, making them truly Web-scale.

Many initiatives inside the library community have in the past tried to tackle problems of semantic interoperability by using controlled vocabularies in novel ways, although the data that is created usually stays locked in, either due to unsettled intellectual property issues or proprietary formats and systems. A culture of sharing results and prototypes at early stages is often not actively promoted or enforced; neither is the exploration of use cases outside of classic library applications. Rarely are these vocabularies presented as building blocks ready to support a variety of applications.

The presentation will briefly introduce a series of major principles for bringing subject terminology to the network level. A closer look at one KOS in particular, the Dewey Decimal Classification, should help to gain more insight into the perceived difficulties and potential benefits of building taxonomy services out and on top of classic large-scale vocabularies or taxonomies.

The discussed elements of taxonomy services include:

- URI design: Establishing persistent concept and representation URIs for both machine and user interfaces
- Caption design: Creating self-contained, semantically profound concept labels aimed at end user display out of the original application context
- Data representation: Embracing of new standards (RDF, SKOS) and implementing common encoding standards (Unicode) in predictable ways
- Enhancement of the scheme itself: Mappings, typing of relationships and entities
- User contribution: Providing frameworks for user-generated content and metadata
- Versioning: Using methods for tracking and exposing the ontogenesis of concepts, preferably derived from the original data
- Vocabulary registries: Making the scheme findable by providing access points and vocabulary-level metadata

The discussion of the DDC will focus on the first three of these elements. The migration from a proprietary data format to the MARC Classification and Authorities formats lays the groundwork for new data representations that can be derived directly from MARC. The proposed changes to the representation of Dewey numbers in the MARC Bibliographic format facilitate new uses in information retrieval and content description. A cool, hackable URI scheme, derived directly from the new data format, would allow

referring to Dewey concepts from a variety of web applications. Further changes will have to be made not only to the structural, but also to the semantic representation (“the verbal plane”) of the scheme to allow concepts to be comprehensible outside of the complete hierarchies.

The goal should be to retain the baroque richness and often conspicuous complexity of large controlled vocabularies through the struggles of the Web revolution, leading these seemingly hopeless nobles into the world of the sobering *Code Napoléon*, and enabling them to become good *Citoyens* without cutting their heads off.