

ISO-Cats: The Revised and Future TC 37 Data Category Registry

Sue Ellen Wright, Marc Kemps-Snijders, Menzo Windhouwer

Abstract

The current Syntax Metadata Registry solution for registering, retrieving, and manipulating data categories used across the spectrum of ISO TC 37's four sub-committees was developed in 2004 as a kind of proof of concept. Several major sets of data categories, e.g., SC 4 metadata categories, morpho-syntax data categories, and several others have been successfully added, but problems remain. Indeed, enough issues have arisen involving the use of the online interface that users have agitated for a major revision of the system. In fall of 2006, the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands, assumed responsibility for the resource and initiated a system analysis with the intention of creating a new set of requirements specifications for a fully new design. At a specialist meeting held in Tilburg in January, 2007, the framework for this new development was established.

Criteria for the new design comprise the following:

Data integrity and flexibility

Robustness and data integrity will be ensured. Entry status will be clarified, and entry "owners" will be able to edit existing entries. Critical features, such as the coordinated creation of closed and simple data categories (parent/child relationships) will be resolved not only to improve user-friendliness, but also to ensure stable relations between parent data categories and value domains (picklist items).

Program features

New features will include automatic responsibility and data encoding, doublette control, clear save features, ability to view lists and profiles in a meaningful way, ability to document ontological relations between data categories in specific Data Category Selections. Search features will be improved. Export formats will be clearly defined.

Interface design

The English interface will be carefully reviewed and vetted by native speakers in order to ensure clarity as well as standard usage and practice.

Icon selection and functionalities will be adapted to standard MAC and PC practice in order to facilitate intuitive operation.

The input template will be designed to be more user-friendly, less time-consuming, and more stable, and users will be able to customize their views.

Difficulties involving some browsers will be resolved.

System organization

Clear role specifications will be defined, together with administrative control functions on the part of the web master or system coordinator.

Administrative issues involving the establishment of a Registration Authority on the part of Max Planck will be resolved.

Functionality for facilitating thematic-domain-related balloting will be resolved. The interaction of human organization in the thematic domain committees must be integrated into the actual software design.

From the beginning it is intended to design the system in a way that the authorized ISO DCR will be hosted at least at three sites with dynamic synchronization capability and that other institutes and initiatives can run their own fully compliant service.

Progress to Date

The new metadata registry has been nicknamed *ISO-Cats*, providing a more meaningful system name than the original *Syntax* name, which was confusing because the system has everything to do with data category (data element) names and with semantics, but very little to do with syntax.

NetKernel™ 1060® has been selected as the back-end application server, and the scaffolding for a set of layered modules has been created to handle major functions like data access, access control and session management. The web front-end uses a Rich Internet Application (RIA) library, i.e. Tibco General Interface™, which provides a state-of-the-art user interface. The data already entered in the Syntax Metadata Registry has been ported to the ProgreSQL database management system.

The first functionality ISO-Cats will implement is the creation of a Data Category Selection by specifying a query on the registry, and the inspection of individual Data Categories. In the next phase this functionality will be extended to adding new and modifying existing Data Categories in a user's own workspace, followed by adding support for the balloting process. In order to enable these core functionalities, ISO-Cats will also have to be extended with supporting functionality, e.g. user management.