

iCAT: A Collaborative Authoring Tool for ICD-11

Tania Tudorache, Csongor Nyulas, Natalya F. Noy,
Timothy Redmond, Mark A. Musen

Stanford Center for Biomedical Informatics Research, Stanford University, US
{tudorache, nyulas, noy, tredmond, musen}@stanford.edu

Abstract. We present iCAT—the Collaborative Authoring Tool for the 11th revision of the International Classification of Diseases (ICD-11). ICD is a fundamental health-care resource developed by the World Health Organization (WHO) and applied in all United Nations countries for a variety of uses. A landmark change in this ICD-11 revision compared to previous ones is the use of OWL as a representation language and of Semantic Web technologies for the collaborative authoring of the ICD content. A community of international medical experts develops ICD-11 in a collaborative setting using the Web-based iCAT platform. Besides its extensive collaboration support, iCAT also fosters the interconnectedness of ICD-11 with other biomedical ontologies and terminologies by providing interlinking and reusing capabilities as a fundamental feature of the tool, while also storing the provenance metadata. The generic and extensible infrastructure as well as its declarative user interface allowed us to easily deploy iCAT in a production setting for the development of ICD-11 and two other WHO classifications. The declarative user interface allowed us to custom tailor the platform for domain experts use. iCAT is in production use since October 2009; it was used to author over 105,000 changes in the ICD-11 ontology, to create more than 40,000 cross-links to other biomedical ontologies, and to produce over 19,000 notes and discussions. The software is open source and a demo version of the platform is available at <http://icatdemo.stanford.edu>.¹

1 ICD-11 – Using Semantic Web Technologies

The International Classification of Diseases (ICD) is the standard diagnostic classification developed by the World Health Organization (WHO) to encode information relevant for epidemiology, health management, and clinical use. Over the years, ICD has become an essential resource in all United Nations countries, who are applying ICD for a variety of uses, ranging from compiling basic health statistics, to billing, and to informing policy making.

To keep up to date with the scientific progress, WHO publishes new revisions of ICD every decade. ICD-10 is the latest revision that is currently in use. Our group has been collaborating with WHO since 2007 to support the development of ICD-11.² A large community of medical experts around the world are involved in the authoring of ICD-11 in a collaborative Web-based platform, called iCAT [2]. The most important change

¹ For the OCAS challenge we created a user account: ocas and password: ocas.

² <http://www.who.int/classifications/icd/revision/en/index.html>

in ICD-11 compared to previous revisions is the decision of WHO **to adopt a solid formalization of ICD-11 and to use Semantic Web technologies for its development.** As a result, the underlying formalization of ICD-11 is OWL, and the platform used for the collaborative authoring of ICD-11, iCAT, a customization of a generic Web-based ontology editor, WebProtégé.³

2 The ICD-11 Semantic Web Platform—iCAT

With the radical change in the ICD-11 revision with respect to the underlying representation, we needed to develop tools and methods that, on the one hand, are suited for domain experts and on the other hand support the much richer Semantic Web content of ICD-11. Some of the main requirements for the platform are: a) support the **Web-based collaboration** of experts all around the world, b) provide features to **inter-connect and inter-link** ICD-11 with other biomedical ontologies and terminologies, and c) provide a **user interface tailored to the needs of the domain experts** that would hide or sugar-coat complex OWL formalization details.

As an overall strategy, we decided to implement **generic features** that can be used for the development of other Semantic Web applications in the more generic WebProtégé platform, and have only a very small number of plugins that are specific for ICD in iCAT (which is a customization of WebProtégé).

The development of the ICD-11 core ontology and of the Web platform took place in parallel, with the former informing the latter. A WHO assigned committee of ontology experts developed the core ontology in OWL. We defined **forms and templates** based on this core ontology that domain experts use to enter the actual ICD-11 content in iCAT. During the last 4 years, the core ontology has evolved significantly, and we had to make sure that our tools were adaptable enough to support seamlessly such frequent changes. Our solution was to implement a **declarative user interface**⁴ as a generic feature in WebProtégé, and iCAT became mainly a specific UI configuration of WebProtégé. This feature allowed us to re-configure iCAT on-the-fly while the Web application is running, and also to define **user-specific views** of the ICD ontology.

Collaboration support: One of the main features of iCAT is its extensive support for collaboration in the development of Semantic Web content, including change tracking, contextualized threaded discussions, watches and notifications, an extensible access policy mechanism, and generation of statistics of the ontology-development process⁵ [2]. We reused the functionalities of Collaborative Protégé [3], which are themselves implemented using Semantic Web technologies.

Ontology interconnectedness and reuse support: ICD-11, as many other ontologies, reuses terms from external ontologies and terminologies. We implemented a generic plugin for WebProtégé that enables simple import of terms from external ontologies stored in the BioPortal⁶ ontology repository [1]. This plugin allows users to search for terms in BioPortal ontologies, to browse their details, and then to import them into the ontology with a single click. All these operations are supported through

³ <http://protegewiki.stanford.edu/wiki/WebProtege>

⁴ <http://protegewiki.stanford.edu/wiki/WebProtegeLayoutConfig>

⁵ <http://protegewiki.stanford.edu/wiki/ChangeAnalysisTab>

⁶ <http://bioportal.bioontology.org>

REST calls to BioPortal. Several properties in ICD-11 have values coming from external ontologies, such as *body part*, *morphology*, or *genomic linkages*. Based on the property range definitions in the ICD ontology, we configured the fields in the UI for these properties to search specific ontologies, or their subsets, in BioPortal. Some of the external ontologies and terminologies we have linked to ICD-11 include SNOMED-CT, The Gene Ontology (GO), the Online Mendelian Inheritance in Men (OMIM), the International Classification of Functioning, Disability and Health (ICF) and the International Classification of External Causes of Injuries (ICECI). During the last year, users have created more than 40,000 links between ICD-11 and terms from external biomedical ontologies. More than 14,000 of these links are for *body part* associated to a disease and take values from the SNOMED CT Anatomy branch. Another use case for inter-linking ICD-11 with its previous revision, ICD-10 (also stored in BioPortal) is to keep track of the ICD-10 to ICD-11 mappings that are essential for transitioning existing medical software to the new ICD-11 coding system.

Infrastructure reuse: We implemented WebProtégé as a pluggable and extensible architecture that can be customized to the needs of a particular project. As a proof, iCAT is one particular customization of WebProtégé. We reused this generic infrastructure to deploy similar platforms to support the production development of two other WHO classifications: the International Classification of Traditional Medicine (ICTM)⁷ and the International Classification of Patient Safety (ICPS).⁸ These two platforms required only a new user interface configuration file and no code changes. Other customizations of WebProtégé are available on the WebProtégé demo server.⁹

3 Conclusions

We presented iCAT, a customization of the WebProtégé platform for the development of ICD-11. We discussed the advantages of using Semantic Web technologies for ICD-11 elsewhere [2]. iCAT has been in production use since 2009 and other customization of WebProtégé for real-world Semantic Web applications are available. All software is open source, pluggable, reusable and under active development.

Acknowledgments

We thank our WHO collaborators for developing the project requirements and for the fruitful collaboration. The work presented in this paper is supported by the NIGMS Grant 1R01GM086587.

References

1. N. Noy, T. Tudorache, C. Nyulas, and M. Musen. The Ontology Life Cycle: Integrated Tools for Editing, Publishing, Peer review, and Evolution of ontologies. In *AMIA Annual Symposium Proceedings*, volume 2010, page 552. American Medical Informatics Association, 2010.
2. T. Tudorache, S. Falconer, C. Nyulas, N. Noy, and M. Musen. Will Semantic Web Technologies Work for the Development of ICD-11? In *The 9th Intl. Semantic Web Conference (ISWC 2010)*, pages 257–272. Springer, 2010.
3. T. Tudorache, N. F. Noy, and M. A. Musen. Supporting Collaborative Ontology Development in Protege. In *Seventh International Semantic Web Conference, ISWC 2008*, Karlsruhe, Germany, 2008.

⁷ <http://icatdemo.stanford.edu/ictm/>

⁸ <http://icat-ps.stanford.edu>

⁹ <http://webprotege.stanford.edu>