

Will Semantic Web Technologies Work for the Development of ICD-11?

Tania Tudorache, Sean Falconer, Csongor Nyulas, Natalya F. Noy, and Mark A. Musen

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

Abstract. The World Health Organization is beginning to use Semantic Web technologies in the development of the 11th revision of the International Classification of Diseases (ICD-11). Health officials use ICD in all United Nations member countries to compile basic health statistics, to monitor health-related spending, and to inform policy makers. While previous revisions of ICD encoded minimal information about a disease, and were mainly published as books and tabulation lists, the creators of ICD-11 envision that it will become a multi-purpose and coherent classification ready for electronic health records. Most important, they plan to have ICD-11 applied for a much broader variety of uses than previous revisions. The new requirements entail significant changes in the way we represent disease information, as well as in the technologies and processes that we use to acquire the new content. In this paper, we describe the previous processes and technologies used for developing ICD. We then describe the requirements for the new development process and present the Semantic Web technologies that we use for ICD-11. We outline the experiences of the domain experts using the software system that we implemented using Semantic Web technologies. We then discuss the benefits and challenges in following this approach and conclude with lessons learned from this experience.

1 The International Classification of Diseases—A New Beginning

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. Health officials use ICD in all United Nations member countries to compile basic health statistics, to monitor health-related spending, and to inform policy makers. ICD is one of the most important classifications used for health care all over the world. ICD is created by a large collaborative effort among international medical experts. To keep up to date with scientific findings about diseases and to address new uses of the classification, the WHO publishes revisions of the classification approximately every decade. In 2007, the WHO started work on the 11th revision of ICD (ICD-11).

Our group is working closely with the WHO to support the collaborative development of ICD-11. The new requirements for ICD-11, which we describe in Section 3, call for a complete revamping of the classification representation in order to build a more solid and flexible formal foundation. ICD-11 will use OWL as the underlying representation language. The workflow for the new development process is also going

to change fundamentally. The process will become a Web-based open process that is powered by collaboration and social features.

This paper makes the following contributions:

- We analyzed the representational and functional requirements for supporting the new collaborative workflow for the development of ICD-11 (Section 3).
- We developed a customization of WebProtégé, a Web-based version of Protégé, to support distributed collaborative development of ICD-11 (Section 4).
- We performed a formative evaluation of the tool (Section 5).
- We analyzed lessons learned and the challenges and advantages of using Semantic Web technologies for the development of large medical terminologies (Section 6).

2 ICD History, Use, and Development

ICD traces its origins to the 19th century. The initial work on disease statistics actually began in the 16th century with the London Bills of Mortality that listed the number of burials as a warning against the onset of the bubonic plague. The London Bill of Mortality enumerated 81 causes of death and it is the predecessor of international mortality classifications.¹ Several governments and health organizations recognized the importance of this classification and became interested in it. In 1948, the World Health Organization (WHO) took over the responsibility for ICD and its creation and included for the first time the causes of morbidity, in addition to classifying causes of mortality.² Since then, ICD underwent revisions approximately every decade. The current revision of ICD, ICD-10, contains more than 20,000 terms and is used in over 100 countries around the world. ICD-10 is available in the six official languages of WHO (Arabic, Chinese, English, French, Russian, and Spanish) as well as in 36 other languages [11].

2.1 Uses of ICD

ICD is an essential resource for health care all over the world. Its strength comes from enabling researchers to undertake studies of temporal and spatial distributions of certain diseases and to make estimates of the effects of diseases on populations [3]. ICD also enables the study of numerous other epidemiological aspects of diseases in human populations. More recent uses include indexing and retrieving of medical records, or use in reimbursement, audit systems, and public policy. At its core, the most important contribution of ICD is the ability to exchange comparable data from different regions and allowing the comparison of different populations over long periods of time.

2.2 The Previous ICD Development Process

WHO publishes three types of ICD revisions and updates: Every decade, a *revision process* takes place and results in a new ICD revision, such as ICD-9 or ICD-10. To

¹ Mortality is the proportion of deaths to population, or the rate of death.

² Morbidity is defined as the incidence of disease, or the rate of disease.

keep up to date with new scientific findings that occurred between 2 subsequent ICD revisions, WHO makes *yearly updates* and *3-year major updates* of the classification [4]. The yearly update usually contains hundreds of changes, while the 3-year update involves more significant changes that impact the mortality and morbidity statistics.

The revision from ICD-9 to ICD-10 was done mainly via regular mail. Non-governmental organizations, statistical offices, and scientific societies proposed changes to ICD. WHO sent the proposed updated chapters to the involved parties for review and comments. Then, experts discussed controversial topics and agreed on the chapters in face to face meetings. In the last step of the process, WHO experts reviewed the changes for consistency across different chapters and for structural integrity of the overall classification.

Starting with ICD-10, the proposals for updates came from national stakeholders through the WHO collaborating centers [10]. WHO circulated the proposal for updates in a formal way via email. Experts then met in one or two teleconferences to seek agreement on edits or overall acceptance of the proposals, in addition to emailing of the proposals with lists of comments and originators. The centers usually went back to their national scientific societies to get analysis of the proposals from a scientific point of view. WHO made final decisions at the annual face to face meeting of the WHO Family of International Classifications Network, strongly believing that this was the best way to solve open issues with the proposals.

In an effort to streamline the ICD-10 update process, WHO developed a Web-based application, *ICD-10 Plus*,³ to serve as the common platform for incremental updates to the ICD-10 revision. The goal of the platform was to make the development process transparent and to encourage the participation of external experts. ICD-10 Plus functions as a workflow engine that starts when an expert creates a change proposal and ends when the proposal is either removed from the system or implemented in ICD. Users also have different levels of authorization ranging from *standard users*, who can only submit proposals and participate in discussions, to *moderators* and *administrators*, who have more access permissions.

ICD-10 Plus is implemented as a database-backed system with a fixed scheme for storing the classification and the information attached to a disease. The types of structured proposals (e.g., a proposal to introduce a new category) are modeled as database tables with predefined fields. The workflow implementation is very specific to the ICD revision process and it is hard-coded in the tool.

3 Requirements for the New ICD Development

In 2007, WHO initiated the work on the 11th revision of ICD (ICD-11) with the mission “to produce an international disease classification that is ready for electronic health records that will serve as a standard for scientific comparability and communication.”⁴ ICD-11 will introduce major changes to ICD, which the WHO characterizes as (1) evolving from a focus on mortality and morbidity to a *multi-purpose and coherent classification* that can capture other uses, such as primary care and public health;

³ <http://extranet.who.int/icdrevision>

⁴ <http://sites.google.com/site/icd11revision/home>

(2) creating a *multilingual international reference standard* for scientific comparability and communication purposes; (3) ensuring that ICD-11 can function in electronic health records (EHRs) by *linking ICD to other terminologies and ontologies* used in EHRs, such as SNOMED CT; (4) introducing *logical structure and definitions* in the description of entities, and representing ICD-11 in OWL and SKOS. In addition to these changes in structure and content, the WHO is radically changing the revision process itself. Whereas the previous revisions were performed by relatively small groups of experts in face-to-face meetings and published only in English and in large tomes, development of ICD-11 requires a Web-based process with thousands of experts contributing to, evaluating, and reviewing the evolving content online.

Thus, the requirements for the new ICD revision fall in two categories (1) developing a richer and formal representation for ICD-11 that will support the new goals of the classification, and (2) designing and implementing an open social development environment to support the richer content acquisition.

3.1 Representation Requirements

ICD-10 is a statistical classification and it lacks a formal representation. A classification is a set of categories (buckets) into which one can place all the objects in the universe, for which the classification was designed. In the ICD case, the universe is represented by all diseases and health related problems. As such, ICD has to comply to the classification principles [1,4]:

- ICD must have a category for each (possible) disease.
- Categories cannot overlap, which means that a disease cannot be placed into two or more categories.
- Each category must have at least one disease; thus, a category cannot be empty.

To maintain its usefulness for statistical purposes, ICD must follow these principles. However, with its extension of goals, *to become a multi-purpose classification* for a much larger number of usages, the current ICD already faces a number of issues. The use of **different classification axes** by different branches of ICD allows the classification of a disease in two or more categories. For example, bacterial pneumonia is both an infectious and a respiratory disease. Such classification of a category in more than one branch violates one of the principles of classification.

Different uses will also require **different level of details** in the classification. For instance, in primary care it will be enough to have appendicitis as a category, but in a clinical-care setting we will need a much higher level of detail, and even more details in a research setting. The ICD-11 representation will have to encompass all its uses with their possibly different properties. From this representation, we will have to be able to extract valid classifications for the different uses at the appropriate level of detail, while maintaining the coherence among them.

If previous ICD revisions contained only minimal information about a disease, usually just a code, WHO will significantly extend the ICD-11 **representation of a disease** to cover different aspects of diseases, such as clinical description, causal mechanisms,

risk factors, treatment, functional impact, and so on. These aspects can serve as different classification axes. We will need to devise representation patterns for the new attributes of a disease and find ways of **linking them to predefined value sets**.

ICD is in use in many countries around the world. Some of the countries, including the USA, Canada, Germany, and Australia, found ICD to be insufficient for the level of detail that they needed for clinical and administrative uses, and created extensions of the classification, known as *Clinical Modifications* [4]. For example, the ICD-10-CM in use by USA has more than 60,000 categories. As a consequence, there are now multiple extensions of ICD with no formal linkages among them, restricting the compilation of international statistics only to certain cases. ICD-11 will try to **integrate the clinical modifications** into one consistent representation. In the initial step, ICD-11 will merge the clinical modifications into one representation, which medical and classification experts will curate in a second step. From the all-encompassing representation, we should be able to generate the country-specific classifications that will represent a subset of the original. To support this requirement, we need to **maintain the metadata about the provenance** of the country specific categories. Our representation will also need to be able to **model the relevant usages for a category** (for example, a disease is relevant for morbidity use, but not for mortality).

The content of ICD-11 will also be evidence-based: for each piece of information stored in ICD (e.g., the risk factors of a disease), the experts will have to provide scientific evidence in form of links to publications or official documents. The ICD representation will have to **store the evidence** in form of metadata attached to each assertion.

Furthermore, ICD-11 has to be **language-agnostic** and provide translations of the labels used in the classification in several languages. WHO also intends to maintain a mapping between the ICD-10 code and ICD-11 code of a disease to support the **migration of existing medical software** to the new classification.

One important WHO desiderata is to make ICD useful in electronic health records by linking it to other standard terminologies and biomedical ontologies, such as SNOMED CT. We must develop a representation and methodology for creating **references to terms in external resources**. For example, the description of a disease will include a body part. Rather than creating its own anatomy taxonomy, ICD will reference a term in the *Anatomy* branch of SNOMED CT. The reference will have to store metadata, such as the source of the term, terminology version, identifier, and link to the term.

3.2 Development Process Requirements

While the previous ICD development happened mostly in face-to-face meetings behind closed doors, WHO envisions to use an open social process for ICD-11 that will involve a large international community of experts. The process will be similar to Wikipedia, where a large number of people contribute to the content. In the ICD case, WHO hopes that a large number of medical experts will contribute to the content of ICD-11.

The development of ICD-11 will happen in several phases. The *alpha phase* is open only to WHO experts. The goal of this phase is to develop the new representation, to test it internally, and to fill the content of the *alpha draft*, the initial draft of ICD-11. The alpha phase will end in May 2011. We performed the work presented in this paper as

part of the alpha phase. In the *beta phase*, WHO will open ICD-11 to a large community of experts for feedback, and will also use it for field trials. In the last phases, ICD will be open to the entire public for viewing (2014). Following the approval by the World Health Assembly (WHA), which is planned for 2014, the implementation of ICD-11 in health care systems across the world will start.

The workflows that each of the development phases needs differ significantly. In the alpha phase, the main focus is on finding an agreement on the ICD formal representation, and having WHO experts pre-fill a large part of the content. In this phase, it is critical to enable many experts to fill in effectively as much content as possible. In the beta phase, the process will change completely. The number of users will increase from around a hundred to thousands. In such a situation, enforcing access policies becomes a priority. The focus changes from having an effective editing platform to having a platform where users can make change proposals and discuss issues in the classification. Collaboration and workflow support becomes one of the most important features. WHO envisions a reviewing process of the ICD-11 content by external domain experts similar to the scientific peer review process that will ensure a high quality of the classification. WHO will also define a quality assurance process for ICD that should become an integral part of the development cycle.

The main challenge in our experience so far was the lack of a well defined collaboration workflow. The different groups of WHO experts could not agree on a concrete workflow for the alpha phase, the roles of users, their access policies and the sequence of steps and responsibilities. However, proceeding to the next phases of ICD will be impossible without a well defined workflow.

4 The Semantic Web Approach

As the representation requirements on ICD-11 have become more complex (Section 3.1), the WHO decided to use OWL as the underlying formal representation language. WHO created a committee—the Health Informatics Modeling Topic Advisory Group (HIM-TAG), to design an appropriate OWL representation for ICD-11, the *content model*. One of the paper authors (Prof. Mark A. Musen) serves as a chair of the HIM-TAG. The Revision Steering Committee (RSG) serves as the planning and steering authority in the update and revision process. The RSG in collaboration with the HIM-TAG are in charge of defining workflows for the different phases of the ICD development.

Our group has been involved in both committees from the beginning of the ICD-11 revision process, and has contributed to the development of the content model. We have also designed and implemented a Semantic Web tool that the WHO domain experts use to edit the ICD-11 content in the alpha phase. The tool is based on WebProtégé—a lightweight ontology editor for the Web [9], which extends the popular Protégé platform. Besides browsing and editing support, WebProtégé supports collaboration processes and has a highly customizable user interface for knowledge acquisition.

We describe the design of the ICD-11 OWL ontology in Section 4.1, and WebProtégé in Section 4.2.

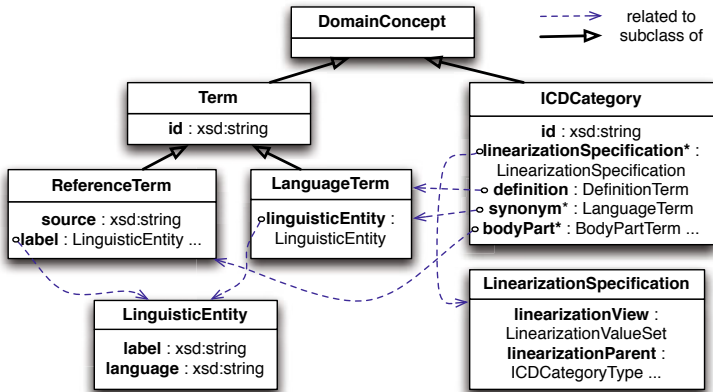


Fig. 1. A snippet of the ICD Ontology. The *ICDCategory* is the top-level class in the ICD disease hierarchy. The usage of a class is modeled using the *linearizationSpecification* property. The property values of a disease class are instances of the class *Term*.

4.1 The ICD-11 OWL Ontology

The ICD ontology⁵ should serve as the underlying representation for all information related to diseases, including the definition of disease characteristics, linkages to external terminologies, as well as linguistic information for translation in multiple languages. One of the major challenges in designing the ontology was supporting the different usages of ICD that have to conform to valid classification principles (see Section 3.1). The HIM-TAG proposed a layered model of the ICD Ontology. The *Foundation Layer* will contain an all-encompassing model of all the usages of ICD and it will allow multiple parents of a category. The *Linearization Layer* will provide a view on the Foundation Layer, called a *Linearization*, for each specific usage. The main characteristic of a Linearization is that it is *linear*—each category will have exactly one parent, which satisfies the most important principle of a classification. There will be linearization corresponding to mortality, morbidity, primary care, and so on. Tu and colleagues [7] provide a full description of the ICD Ontology and the content model. The Foundation Layer ensures that all linearizations provide a consistent representation of diseases, and that they are coherent with one another.

Figure 1 shows the main components of the ICD Ontology that correspond to the Foundation Layer. A class in the ontology represents an ICD category. The *ICDCategory* is the root of the ICD disease hierarchy. We used a metaclass level to describe the different classification axes corresponding to the properties in the ontology. For example, a disease has one or more associated body parts. The representation in the ontology is as follows: A disease class has a property *bodyPart* that is prescribed by the *ClinicalDescription* metaclass⁶. In other words, we associate a property *bodyPart*

⁵ Available online at: <http://tinyurl.com/icd-ontology>

⁶ The disease classes have a number of other metaclasses corresponding to different classification axes.

to each disease class, rather than to its instances. Other properties include textual definition, synonyms, clinical descriptions (body part, body system), manifestation properties (signs and symptoms, investigations), causal properties, temporal and functional properties, treatment, and so on.

The values for these properties are reified instances of the *Term* class and its subclasses. Reification enables us to encode additional information for a property value, such as scientific evidence, translations in multiple languages, or metadata about the linkages to external terminologies. For example, a synonym for a disease name is an instance of *LinguisticTerm* that has a unique identifier, but also provides labels for different languages. Similarly, the value for the *causalMechanism* property is an instance that must contain links to scientific evidence.

The values for most of the properties should come from predefined value sets. These property values are represented as instances of the *ReferenceTerm* class and its subclasses. The ICD ontology uses two types of reference terms: terms in external terminologies and terms in value sets defined locally in the ICD ontology. A reference term instance has metadata associated with it, such as a unique identifier, a preferred label, the source terminology, a direct URL link to the term (if available). For example, the *Myocardial Infarction* class has as a value for the *bodyPart* property a reference to the SNOMED CT term for *Entire Heart*. The value of *bodyPart* is therefore an instance that has as property values: *id*=302509004 (identifier in SNOMED CT), *source*=SNOMED CT, *preferredLabel*=Entire heart, a direct URL⁷ to the term in BioPortal, and other properties to store additional metadata.

The ICD ontology is built in a modular way. The upper ontology (i.e, the content model) provides the basic structures and is imported directly or indirectly by all other modules. The right hand side of Figure 2 shows the import hierarchy of the ontology modules that make up the ICD-11 ontology.

4.2 WebProtégé—The Knowledge Authoring Platform Used for ICD-11

Our group developed WebProtégé—a highly customizable Web interface for browsing and editing ontologies, which provides support for collaboration. We have created a specific customization of WebProtégé for the content acquisition in ICD-11.⁸ In the remainder of the section, we describe the architecture and the features of the tool.

Architecture

Figure 2 shows an overview of the WebProtégé architecture. The core of the system is the **Collaboration Framework** [8], which provides all the collaboration and ontology access services that the client applications need through the *Ontology Access API*, *Notes and Discussions API* and the *Change Tracking API*. The *Workflow API* provides the workflow support, including access policies and user management. The clients of the Collaboration Framework are the Protégé desktop application, WebProtégé, and any other application that access its services either directly through the Java APIs or through the remote RMI interfaces.

⁷ Direct URL in BioPortal to *Entire Heart*: <http://tinyurl.com/bp-heart>

⁸ The ICD demo platform is available at: <http://icatdemo.stanford.edu>

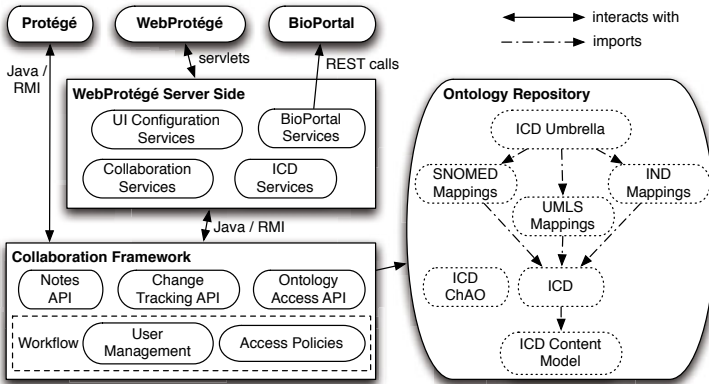


Fig. 2. An architecture diagram of WebProtégé used in the ICD context. WebProtégé UI uses the services on the server side to display information. The WebProtégé server side and the desktop client Protégé connect to the *Collaboration Framework* to access the ontology and the collaboration services. The *Ontology Repository* stores the ontologies available to the clients. The modules and the import hierarchy of the ICD ontologies are also shown.

The WebProtégé server connects to the Collaboration Framework to access the ontology content and the collaboration services. The WebProtégé server also provides services for the user-interface configuration and for creating links to external terminologies stored in BioPortal—a web-based repository for biomedical ontologies and terminologies [6]. In the WebProtégé deployment for the ICD-11 authoring, we have used the Collaboration Framework and implemented additional services for the ICD-specific functionality.

Features

The main feature of WebProtégé is the support for Web-based **browsing and editing** of ontologies. Medical experts from around the world are using the system to edit the ICD-11 content simultaneously. Every change that one editor makes is immediately committed to a shared copy of the ontology. The change is then propagated to all other clients in the collaboration framework and other editors see the changes in real time.

Another key feature of WebProtégé is the support for **collaboration**. The support for notes and discussion ensures that domain experts can raise questions and discuss different issues that arise during the authoring. Several note types are available in the Web interface. Users can attach notes to any entity in the ICD ontology (class, property, individual), or even to a specific property-value assertion (for instance, to one of the synonyms of a disease class). The notes and discussions are threaded and can contain any HTML formatted text. **Change tracking** is one of the most important features that WHO uses in the quality assurance. Indeed, for each action that the user performs, there may be several granular changes that happen in the ontology (e.g., creating a reference to another term involves creating a new instance, adding property values to it, and setting a property value in the container class to refer to that instance). We store granular changes for analysis and debugging purposes, but present the user only with the user-friendly high-level descriptions of changes that correspond to the user actions.

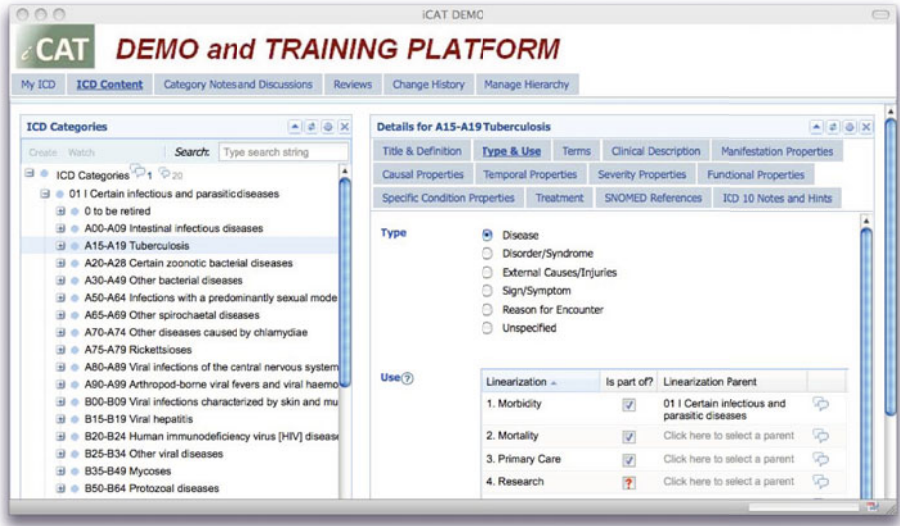


Fig. 3. The ICD authoring tool using WebProtégé. Each tab contains one or more panels, called *portlets* that can be arranged by drag-n-drop. The left hand-side portlet shows the disease class hierarchy of the ICD ontology. The right panel shows the uses (linearizations) of the selected disease in the tree, in this case *Tuberculosis*.

Protégé stores both notes and individual changes that the authors make as instances in the *Changes and Annotation Ontology (ChAO)* [5]. The changes are hierarchical, meaning that a change may be composed of other sub-changes, similar to nested database transactions. We have implemented a tool that uses the structured changes log to analyze the activity in the ICD ontology and to generate statistics.⁹

The new development process for ICD also envisions that the content of ICD-11 will be reviewed in a manner similar to scientific paper reviews (see Section 3.2). We have implemented a rudimentary **reviewing feature** that allows a user with the appropriate privileges to request the review of a disease description. WebProtégé sends a review request to the reviewer, who can later log into the system and enter her review. As any other note types, reviews can be attached to a class representing a disease, or to individual property values.

Figure 3 shows the user interface of WebProtégé as deployed for the ICD-11 authoring. We have implemented a **declarative user interface** that allows us to define the user interface components and layout in an XML configuration file. The configuration file declares the binding of the user interface elements to the underlying ontology entities. For example, a text field can be used to edit the values of a string property, or a radio button for functional properties. By simply changing the XML configuration file, we can quickly change the user interface without the need to compile or re-deploy the application. This feature provides great flexibility for projects such as the ICD

⁹ The Change Analysis Plugin is available at: <http://tinyurl.com/ch-analysis>

development, in which the underlying ontology structure (even the upper ontology) is still under discussion and active development. We implemented the user interface as a portal, in which users can arrange *portlets*—components providing independent pieces of functionality, simply by drag-n-drop. WebProtégé is **extensible** and has a **plugin infrastructure**. We implement some ICD-specific portlets for the ICD infrastructure. The high customizability of the user interface also allows us to define different layouts for different users based on their interest and domain of expertise. One of our main reasons for developing a highly configurable user interface was to be able to hide from domain experts ontology details that are not relevant to them. We have spent a lot of effort in ensuring that the user interface does not look like an ontology editing environment, but is **customized for the domain experts**.

One of the important requirements in the ICD-11 development is to support the **linkages to terms in external terminologies**. We implemented a *Reference Portlet*, which allows users to import terms from terminologies stored in BioPortal with a single click. For instance, if the user wants to import a reference to a term *Heart* from the SNOMED CT Anatomy branch, we invoke a RESTful service call to BioPortal and fetch the results. Then, the user can simply click on an import link next to one of the search results. On the backend, we create an instance of a *ReferenceTerm*, and present it in a table format to the user.

5 Usage of the Semantic Web Platform

During the alpha phase of the ICD-11 revision project, editors have been updating and making changes to the ontology from November 2009 to present day. Figure 4a displays the changes that the editors have made during this time period. Different colors in the stacked area chart correspond to changes made by different editors. The x-axis is the months of the project and the y-axis is the number of changes made by a given author.

After six months of use, there has been a total of 15,025 changes made by 16 different editors out of 48 users who have logged into the system. The platform does not require a sign in for read access, and therefore we did not keep track of other users of the system. Contributions have ranged greatly from as little as a single change to as many as 7,709. The average number of changes per editor is 684, while the median is 85. The editors have created a total of 483 definitions for terms, added 2,464 completely new terms to the hierarchy, removed 149 ICD-10 terms and moved 1,415 ICD-10 terms to a new location in the ontology hierarchy. While previous versions of ICD did not support multiple inheritance, there is already 464 terms with multiple parents within ICD-11.

Editors have also been actively participating in discussions. Figure 4b displays a stacked area chart representing editor note contributions per month in the alpha phase. The editors created a total of 5,035 notes. Similar to change contributions, the amount of note contribution has varied greatly amongst the editors. Contributions range from as few as 1 to as many as 2,422. The average is 315, while the median is 105.

These statistics demonstrate that the tool is being used actively by the editors, but it does not provide details about their feelings on the usability of the tool. To gather this information, we carried out a web-based survey, which we describe in the next section.

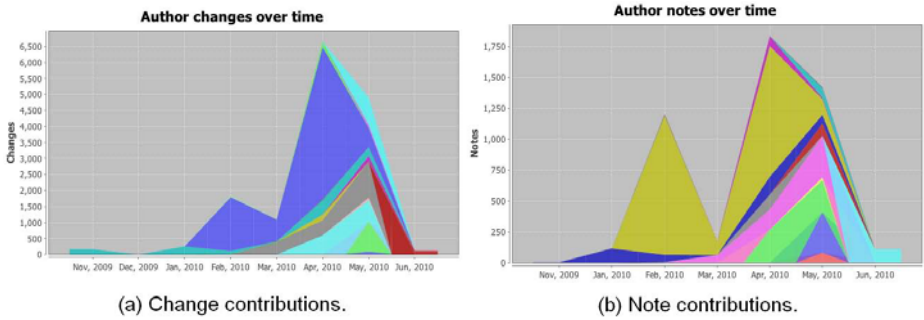


Fig. 4. Visual representation of editor changes (a) and notes (b) contributions per month during the alpha phase of the ICD-11 revision process

Table 1. Survey questions. iCAT stands for “The ICD Collaborative Authoring Tool” and is the customized WebProtégé platform used for editing ICD.

#	Question
1	Rate your experience with iCAT?
	I thought iCAT was easy to use.
	I found iCAT to be unnecessarily complex.
	I think I would need technical support to be able to use iCAT.
	I found the various features of iCAT to be well integrated.
	I thought there was too much inconsistency in iCAT.
	I think most people would learn iCAT quickly.
	I found iCAT cumbersome to use.
	I felt very confident using iCAT.
	I needed to learn a lot about iCAT before I could effectively use it.
2	What did you like about iCAT?
3	What did you dislike about iCAT?
4	Did you find the discussion and commenting features helpful?
5	Do you have any additional comments not already covered?

5.1 Survey Feedback

To solicit feedback from our users, we conducted a web-based survey consisting of five questions (see Table 1). The first question consisted of completing nine sub-questions, each evaluated on a 5-point Likert scale ranging from “Strongly disagree” to “Strongly agree”. These sub-questions were adapted from the System Usability Scale (SUS) [2]. The other five questions were all open-ended.

Survey results

Thirteen experts responded to the survey¹⁰. The respondents were content experts (medical doctors) and classification experts. None of the respondents had ontology expertise. Some of the experts took part in a training session for WebProtégé in September 2009, but others had to start using the tool with no prior training. The results of the first

¹⁰ The results of the survey are available at: <http://tinyurl.com/icat-survey>

question shows a range of answers. Five of the editors found the tool easy to use, while 4 editors had difficulties with it, and 4 others were neutral. Half of the respondents found the tool too complex and some felt that they needed training or support to use it. On the positive side, the respondents found the tool to be well integrated and consistent.

The second and third questions provide more detail about these ratings. The second question asked about the features that the editors liked. The answers indicate that it is “logical,” has “easy structure and clear layout”, “easy navigation” and the good “integration of features.” Other features mentioned were the easiness in performing hierarchical changes and the integration with BioPortal.

On the negative side, 7 of the 10 responses highlighted the complexity and time needed to enter information. Some of the respondents found that the tool was too related to the informatics side, forcing users to enter too much detail. One respondent also indicated that they wanted to be able to display more information, in report form, on a single screen and be able to export this information for discussion.

The fourth question asked about the usefulness of the notes feature. Five of the 9 respondents indicated that they found the feature useful. The other respondents either had not used the feature, or indicated they had trouble finding new comments.

The final question asked whether the respondents had any comments not already covered by the survey. Only four respondents provided further feedback. Two indicated great concern about having an open editing process. They felt that too many users contributing without classification expertise could lead to inconsistencies in the hierarchy and continual changes to certain areas. The other two respondents re-iterated the time needed to enter information and felt that there needed to be support for submitting a less structured form of the content model.

6 Discussions and Future Work

The results of the survey provide important feedback both to the developers of the collaboration tools and to the team developing the content model itself (the HIM-TAG). Indeed, WebProtégé only exposes the underlying ontology in a form-based interface. Most users, however, do not distinguish the tool and the model that the tool exposes. The editors in our survey felt the representation was very complex and that they did not understand many of the fields. Indeed, ICD-11 is enormously more descriptive than previous ICD revisions. If in previous revisions, a category contained mainly a code and a handful of properties, in ICD-11 editors can fill in over 40 fields attached to a disease. The description of the fields is also not well defined in the content model yet and there are fierce discussions in the HIM-TAG related to representation of disease characteristics. For example, including a severity in a disease description can be problematic because the severity may vary in different stages of disease or based on other factors. Given this newly added complexity of the model and the constant changes that it is still undergoing, we believe that some of the issues that the editors raised in our survey, were actually related to the content model itself and not to the tool.

Of course, there were also issues related to the tool itself. We implemented the initial version of the tool in a very short period of time with no formal requirements. The experts and HIM-TAG were able to formulate the requirements only after they saw the

first prototype of the tool in September 2009. Ever since then, the tool, the ICD ontology, and requirements evolve together. Supporting the ICD-11 process as it is today was possible only by having a flexible system that can be easily adapted to changes in the underlying ontology. Any application that would hard-code the binding between the user interface and the ontology would not have worked in this setting.

We are aware of several user interface issues in the tool that make the work of the editors more difficult. In several discussions with the editors and also from the survey, we know that the lack of synchronization between the class trees in the different tabs causes a change of context and editors can make mistakes much easier. We are also aware of some performance problems due to the concurrency support in the ontology APIs on the backend side that use a coarse-grain lock at the level of an ontology. Meanwhile, we have developed an implementation of the ontology APIs with a finer-grained locking mechanism that addresses this issue.

Some problems are related to the backend side and the fact that the Changes and Annotation ontology grows at a much more rapid rate than the domain ontology. Therefore, we must address the question of whether to store the change tracking information as instances in an ontology makes sense. The crucial question is whether the structure and type information of the changes provide any advantage. We believe that it does. The Change Analysis plugin generates different kind of statistics and makes use of the type structure information for some of the statistics (for example, how many classes have been moved from one branch of the ontology to the other). We need to find better technical solutions to address this scalability problem.

Our experience with ICD-11 made us realize that developing and providing content to formal models is just cognitively difficult. This process is even harder for domain experts who have no prior knowledge of ontologies or knowledge representation. In developing the web platform, we tried to “hide” as much of the underlying formal representation as possible. We implemented simplified editing widgets, such as instance tables, to present and support the acquisition of reified relationships. Most of the relations in the ontology are reified, but domain experts are not aware of their underlying representation. We also tried to model the entire user interface as a form-based interface using common editing widgets (textfields, radio buttons, check boxes, etc.) that were already familiar to the users, and did the “heavy-lifting” on the server side. However, we think that there may be a limit to how much the user interface will be able to simplify a task that is similar to ours.

Initially, we developed the ICD ontology using a frame-based formalism and soon after we migrated the representation to OWL. Although in the current model we use very few DL-specific constructs, there are several benefits of using OWL. First, it is a W3C recommendation and has a well defined formal semantics, which will enable the reuse of ICD-11 in other ontologies and by other tools. Second, we make use of two of the most common modeling patterns in OWL—defining value sets and reified relationships—which form the basis for the entire representation. Third, we make use of inverse properties, cardinalities, domain and ranges, which are important in the model and are enforced in the user interface. Fourth, we plan to use OWL annotations for storing the metadata and linguistic information, which play a central role in the model. Fifth, in order to manage manage poly-hierarchies and multiple inheritance, we plan to

convert the current representation into a DL form using OWL definitions and restrictions. We will then be able to check the consistency of the manually created polyhierarchies using a DL reasoner.

One concern that the editors raised is whether the development of the ICD-11 can be open to the community at large, as planned by WHO. It is not yet clear whether a development *à la* Wikipedia would work for ICD. The main issue is whether a domain expert with no understanding and training in the new content model will be able to contribute in a significant way to the development of ICD. We do not have answers to these questions yet. For the beta phase, we plan to implement a much stricter access policy mechanism (access rights at the level of branches and depending on many workflow variables). The collaboration workflow will also be significantly different. It will switch its focus from editing to a proposal-based process. External experts will be able to submit structured change proposals. However, the exact workflow and quality assurance for the beta phase is still undefined. As in many other cases, the technology and implementation is only secondary, and the main challenge is social: the various teams of experts that are developing the workflow are yet to agree on one.

The general evaluation of WebProtégé as a platform for the collaborative authoring of ICD is positive. We are currently working with WHO to build two other WebProtégé deployments to be used for the collaborative development of two other WHO classifications: the International Classification of Patient Safety (ICPS) and the International Classification of Traditional Medicine (ICTM). The flexibility and versatility of WebProtégé allows very quick customization of the tool for different ontologies. In fact, we have built a first prototype¹¹ for the Traditional Medicine classification in less than 2 weeks. We spent most of this time adapting the upper level ontology to the ICTM content model, and then we were able to configure the new user interface showing Chinese characters in a very short time. For the near future, we will add collaboration features needed for the beta phase, and will also work on providing internationalization support. We also plan to perform further evaluations of the tool, and we are particularly excited to witness the use of the platform in a much larger setting.

7 Conclusions

Developing ontologies is a cognitively hard process and we do not yet have a good grasp of simple interfaces for this type of development. Even though defining a new class in ICD-11 is as “simple” as filling out a number of pre-defined terms, with value sets for many of the fields also pre-defined, users still found the process difficult and cumbersome. And while some of this difficulty was in fact the difficulty in understanding the meaning of the fields themselves, some of it was from the amount of information that we must present on the screen. We need to consider ways to custom-tailor interfaces dynamically, based on the role that a particular user is playing in the workflow, based on the parts of the class tree that he is interested in, and parts of the content model that he is either qualified or interested in filling out.

Our experience working with WHO on the ICD-11 alpha draft also helped both teams understand better what the advantages of the semantic technologies were. The

¹¹ The ICTM prototype is available online at: <http://icatdemo.stanford.edu/ictm/>

need to reference other ontologies, the need for multiple inheritance along with single-inheritance linearizations, the ability to integrate labels in multiple languages are exactly the strong points of semantic-web technologies such as RDF and OWL.

Acknowledgments. We are very grateful to Jennifer Vendetti, Timothy Redmond, Jack Elliott, and Martin O'Connor for their help with the design and implementation of WebProtégé, and to Bedirhan Üstün, Robert Jakob, Can Çelik, and Sara Cottler from WHO for their work in developing the requirements for the project and for providing documentation on the previous ICD development process. We would like to thank Alan Rector, Samson Tu and all the members of the HIM-TAG and RSG for their invaluable work in designing the ICD content model. The work presented in this paper is supported by the NIGMS Grant 1R01GM086587-01. Protégé is a national resource supported by grant LM007885 from NLM.

References

1. Bailey, K.D.: *Typologies and taxonomies: An introduction to classification techniques*. Sage Publications, Inc., Thousand Oaks (1994)
2. Brooke, J.: Usability evaluation in industry. In: Jordan, P.W., et al. (eds.) *SUS: a 'quick and dirty' usability scale*, pp. 184–194. Taylor & Francis, Abington (1996)
3. Israel, R.A.: The International Classification of Disease. *Two Hundred Years of Development*. *Public Health Rep.* 93(2), 150–152 (1978)
4. Jakob, R.: Disease Classification. *International Encyclopedia of Public Health* 2, 215–221 (2008)
5. Noy, N.F., Chugh, A., Liu, W., Musen, M.A.: A framework for ontology evolution in collaborative environments. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L.M. (eds.) *ISWC 2006*. LNCS, vol. 4273, pp. 544–558. Springer, Heidelberg (2006)
6. Noy, N.F., Shah, N.H., Whetzel, P.L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D.L., Storey, M.-A., Chute, C.G., Musen, M.A.: BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research*, 10.1093/nar/gkp440 (2009)
7. Tu, S.W., et al.: A content model for the ICD-11 revision. Technical Report BMIR-2010-1405, Stanford Center for Biomedical Informatics Research (2010)
8. Tudorache, T., Noy, N.F., Tu, S., Musen, M.A.: Supporting Collaborative Ontology Development in Protégé. In: Sheth, A.P., Staab, S., Dean, M., Paolucci, M., Maynard, D., Finin, T., Thirunarayan, K. (eds.) *ISWC 2008*. LNCS, vol. 5318, pp. 17–32. Springer, Heidelberg (2008)
9. Tudorache, T., Vendetti, J., Noy, N.: WebProtege: A lightweight OWL ontology editor for the Web. In: *OWL: Experiences and Directions*, 5th Intl. Workshop, OWLED 2008, Karlsruhe, Germany (2008)
10. World Health Organization: Collaborating Centres for the WHO Family of International Classifications (WHO-FIC), <http://www.who.int/classifications/network/collaborating/> (last accessed: June 2010)
11. World Health Organization: International Classification of Diseases (ICD), <http://www.who.int/classifications/icd/> (last accessed: June 2010)