Ontology Development for the Masses: Creating ICD-11 in WebProtégé

Tania Tudorache,[1] Sean Falconer,[1] Natalya F. Noy,[1] Csongor Nyulas,[1]
T. Bedirhan Üstün,[2] Margaret-Anne Storey,[3] Mark A. Musen[1]

[1] Stanford Center for Biomedical Informatics Research, Stanford University, US
[2] World Health Organization, Geneva, Switzerland
[3] University of Victoria, Canada
{tudorache, sfalc, noy, nyulas}@stanford.edu, ustun@who.int,
mstorey@uvic.ca, musen@stanford.edu

Abstract. The World Health Organization is currently developing the 11th revision of the International Classification of Diseases (ICD-11). ICD is the standard diagnostic classification used in health care all over the world. In contrast to previous ICD revisions that did not have a formal representation and were mainly available as printed books, ICD-11 uses OWL for the formal representation of its content. In this paper, we report on our experiences in supporting the collaborative development of ICD-11 in WebProtégé—a web-based ontology browser and editor. WebProtégé integrates collaboration features directly into the editing process. We report on the results of the evaluation that we performed during a two-week meeting with the ICD editors in Geneva. We performed the evaluation in the context of the editors learning to use WebProtégé to start the ICD-11 development. Participants in the evaluation were optimistic that collaborative development will work in this context, but have raised a number of critical issues.

1 Creating a Formal Representation of ICD-11

Ontologies and terminologies are a critical component of many knowledge-intensive systems. In recent years, we have seen a considerable growth both in the tools that support the development of ontologies collaboratively and the projects that include contribution by a community of experts as a critical part of their workflow.

The development of large biomedical terminologies and ontologies is possible only in a collaborative setting. The Gene Ontology (GO) is one of the more prominent examples of an ontology that is a product of a collaborative process [3]. GO provides terminology for consistent description of gene products in different model-organism databases. Members of the GO community constantly suggest new terms for this ontology and several full-time curators review the suggestions and incorporate them into GO. The National Cancer Institute's Thesaurus (NCI Thesaurus) is another example of a large biomedical ontology that is being developed collaboratively [4]. The Biomed Grid Terminology (BiomedGT) restructures the NCI Thesaurus to facilitate terminology federation and open content development. NCI is using a wiki environment to solicit the feedback about the terminology from the community at large. The Ontology for Biomedical Investigations (OBI), a product of the OBI Consortium, is a federated ontology, which has more than 40 active curators, each responsible for a particular scientific community (e.g., cellular assay, clinical investigations, immunology, etc.). Developers of these ontologies use a variety of tools and a broad range of editorial workflows to achieve consensus and to ensure quality [11].

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. In the United States, use of the ICD is also a requirement for all medical billing. Thus, ICD is an essential resource for health care all over the world. The ICD traces its formal origins to the 19th Century, and the classification has undergone revisions at regular intervals since then. The current revision of ICD, ICD-10, contains more than 20,000 terms. 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; (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 also 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 will require a Web-based process with thousands of experts contributing to, evaluating, and reviewing the evolving content online.

We have developed iCAT—a custom tailored version of Protégé for authoring the alpha draft of ICD-11 (Section 2).[5] More specifically, iCAT is a customization of WebProtégé, a Protégé client that supports collaboration and enables distributed users to edit an ontology simultaneously, and to use their Web browsers for editing. The iCAT Web application presents users with simple forms that reflect the fields in the ICD-11 content model. The tool also incorporates many collaborative features, such as the ability to comment on ontology entities.

In September 2009, WHO gathered its ICD-11 managing editors for iCamp—a two week meeting with the goal of introducing the editors to the new development process and to the customized WebProtégé tool, developing requirements for further tool support, and evaluating the open development process. In this paper, we report on the results of the evaluation that we performed during the iCamp, where we focused on analyzing the feasibility of the open process for ontology development and the requirements to support such a process.

[4] `http://sites.google.com/site/icd11revision/home`
[5] A demo version is available at `http://icatdemo.stanford.edu`

To the best of our knowledge, the development of ICD-11 is the largest open collaborative ontology-development experiment of its kind. Thus, we believe that the insights that we gained from our evaluation will be informative to the organizers and developers of similar projects. Specifically, this paper makes the following contributions:

– We describe the customized WebProtégé system that is being used in the collaborative development of ICD-11.
– We use WebProtégé as the context for an evaluation of feasibility and requirements of a collaborative ontology-development process.

2 Description of WebProtégé

Our goal in developing a customized version of WebProtégé is to support the collaborative development of the ICD-11 content. In this section, we give an overview of the main artifact that we are building—the ICD Ontology (Section 2.1) and describe the WebProtégé architecture (Section 2.2). We highlight the key elements of the user interface in iCAT, the custom-tailored version of WebProtégé, in Section 2.3.

We joined the ICD revision project in its infancy, when many fundamental issues (content model, representation, workflow) and requirements for the tooling were undefined. Thus, we had to build tools that we can adapt on the fly when changes are made to the underlying model, user-interface requirements and the workflow. In Section 2.4, we describe our design of WebProtégé as a pluggable and extensible platform to enable each project to customize it according to its own requirements. iCAT is in fact a particular configuration of WebProtégé.

Finally, we present the support for collaboration among a large number of distributed users as an integral requirement of the ICD revision process. We discuss the collaboration features of WebProtégé in Section 2.5.

2.1 The ICD Ontology

The previous revisions of ICD stored only limited information about a disease, such as the code, title, synonyms, example terms, and simple conditions. The goal of the 11th revision process is to extend the description of diseases to include other attributes: a textual definition of the disease, clinical descriptions (body system, signs and symptoms, severity), causal mechanisms and risk factors, and the functional impact of a disease. To support the richer representation of diseases, the WHO has defined a formal representation of the model in OWL, the *ICD Content Model*. The content model describes both the attributes of a disease (e.g., Definition, Body System, Severity, Functional Impact, and so on) and the links to external terminologies, mainly to SNOMED CT [14].

The ICD Ontology[6] is the formal representation of the ICD content model in OWL (Figure 1). The class *ICDCategory* is the top level class of the ICD disease hierarchy. The ontology uses a meta-model layer to describe the attributes that a disease class may or should have. For example, the class representing *Acute Myocardial Infarction* disease has as a type (among others) the *ClinicalDescriptionSection* metaclass that prescribes that the range for the property *bodySystem* should be the class *BodySystemValueSet*. In this example, the class *Acute Myocardial Infarction* has the *CirculatorySystem* as a value for the property *bodySystem*.

[6] Accessible at `http://icatdemo.stanford.edu/icd_cm/`

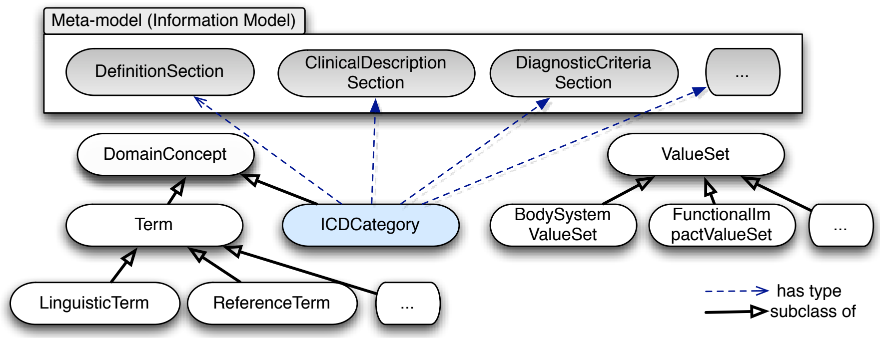

Fig. 1. A snippet of the ICD Ontology. The *ICDCategory* is the top-level class in the ICD disease hierarchy and has as types the metaclasses from the meta-model (gray background). The property values of a disease class are instances of the class *Term*. The *ValueSet* has as subclasses the different value set hierarchies used in the ontology.

All property values describing diseases are reified—they are instances of the class *Term*. For each value, we use this reification to record the source of the value (e.g., for a definition of a disease we need to record the supporting evidence in the form of citations or references) and other salient information. We use *LinguisticTerm*s to represent property values that have different labels in different languages. Because ICD aims to become a multi-language classification, providing support for multi-linguality is paramount.

Property values that are instances of the class *ReferenceTerm* represent links to other terms in external terminologies, such as SNOMED CT. For example, a disease has an associated body part. Rather than defining its own anatomy hierarchy to serve as values for the *bodyPart* property of a disease, ICD-11 references classes in SNOMED CT that represent anatomical parts. Because it is not practical to import the entire SNOMED CT into ICD-11, the *ReferenceTerm* class models all the information needed to identify uniquely an entity in an external terminology: the fully qualified name of the external entity, the name of the ontology, the label of the term, and other auxiliary information. This construct allows us to import references to terms in external terminologies and ontologies in a uniform and practical way.

2.2 Architecture of WebProtégé

Figure 2 shows a high level WebProtégé architecture diagram and the interaction of the different software components. The core functionality of the application is supported by the *Protégé server*, which provides access to the ontology content, such as retrieving and changing classes, properties and individuals in the ontology. The ontologies that the server accesses are stored in a database on the server side. To facilitate the management and reuse of the ICD ontology, we modularized it into several smaller ontologies that import each other. Both the Web-based Protégé client (WebProtégé) and the "traditional" Protégé desktop client access the Protégé server to present the ontologies to the users and to enable the users to edit. Any number of clients of either type can access and edit the same ontology on the server simultaneously. All changes that a user makes

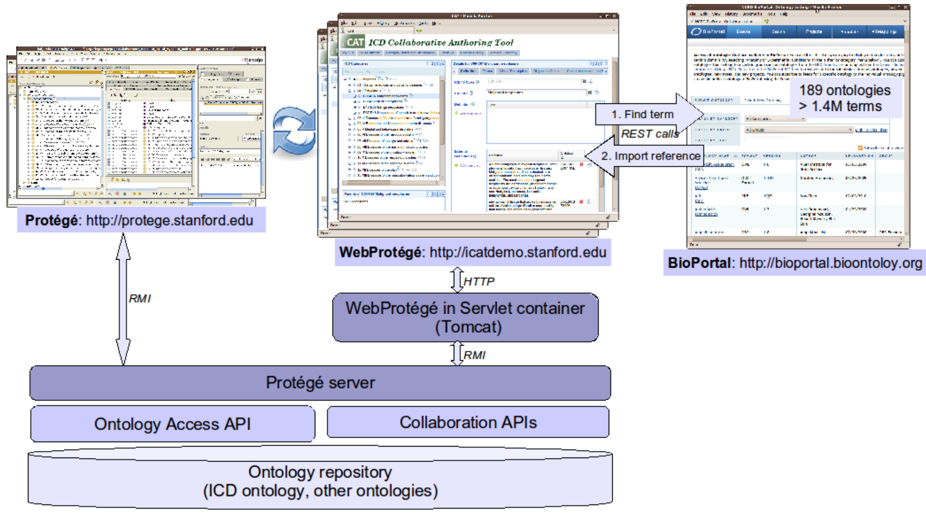

Fig. 2. An architecture diagram of the customized WebProtégé for ICD. The ICD ontology content is accessible through both a Protégé desktop client and in a Web browser. WebProtégé accesses BioPortal for searching terms to import as external references. Both WebProtégé and the Protégé desktop clients connect to a Protégé server to read and write the ontology content and information that supports the collaboration features.

in one of the clients are immediately visible in all other clients. The ICD editors use the WebProtégé client to browse and edit ICD-11. The technical-support team often uses the desktop client to make corrections or perform operations that are not supported in the Web interface.

In order to search external biomedical terminologies and to import terms from these terminologies, WebProtégé accesses *BioPortal*, a repository of about 200 biomedical ontologies and terminologies [9]. BioPortal provides REST service access that enables search across different ontologies and access to information about specific terms.

Support for collaboration among users is one of the key features of WebProtégé. We have developed a general-purpose collaboration framework in Protégé [15] and we use the same framework in WebProtégé. This framework provides Java APIs for tracking changes in an ontology, and for storing notes and discussion threads attached to ontology entities. We also reuse the generic access policy mechanism of the Protégé server that allows us to define customized access policies for an ontology (e.g., a user who has only read access will not be able to edit the ontology).

2.3 Features of the WebProtégé User Interface

WebProtégé is a web portal, inspired by other portals, such as myYahoo or iGoogle. Our vision is to enable users to build a custom user interface by combining existing components in a form that is appropriate for their project. The user interface is composed of *tabs*—either predefined ones or user-defined. A new tab is an empty container in which users can add and arrange by drag-n-drop several *portlets*. A portlet is a user interface component that provides some functionality. For example, the *Class tree port-*

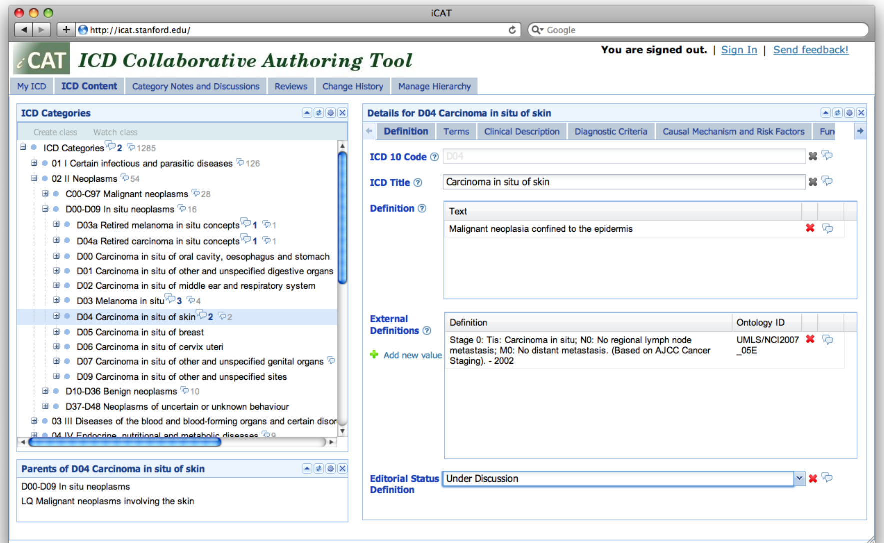

Fig. 3. The WebProtégé user interface customized for ICD. The interface is composed of tabs. Each tab contains one or more panels, called *portlets* that can be arranged by drag-n-drop in the interface. The left hand-side portlet shows the disease class hierarchy of the ICD ontology. The right portlet shows the fields of the selected disease in the tree, in this case *D04 Carcinoma in situ of skin*.

let displays the class hierarchy in an ontology and has support for class level operations (create and delete class, move class in hierarchy, etc.).

Figure 3 shows one of the tabs in the customized WebProtégé interface for ICD. The *ICD Content* tab contains two portlets: the class tree portlet—showing only a branch of the ICD ontology, and a details portlet—showing the property values of the class selected in the class tree in a simple form-based interface. The domain experts are familiar with this type of interface from many other application. For each property, we use a specific *widget* to acquire the property values. For example, we use a text-field widget to record the values of the *ICD title* property (Figure 3). As we have mentioned in Section 2.1, all values of properties describing a disease are reified as instances of the *Term* class. We use an *instance-table widget* to hide this extra reification layer from the user and to present all the details about the reified instance directly in the form for the disease. The widget presents a pre-configured set of property values for the term instance as columns in the table. You can see an example of this widget for the *External Definition* property in Figure 3.

A large part of attributes for diseases have values that are references to terms in external terminologies and ontologies. For example, the property *bodyPart* takes as values references to the Anatomy branch of SNOMED CT (see Section 1). We have developed a generic *Reference Portlet* that supports the simple import of an external reference with a single mouse click. The portlet uses RESTful Web services to search terms in BioPortal. For example, the *bodyPart* for *Acute Myocardial Infarction* should be a reference to "heart" from SNOMED CT. The search in BioPortal will return a list of matched terms. To decide which SNOMED CT term to import, the user may get more information about each search result either in textual form or as a graph visualization that are also retrieved via Web Service calls to BioPortal. The Reference Portlet is also configurable. We can specify in what ontology the search should be performed. We can also restrict the search to a particular ontology branch in the configuration of the portlet (e.g., Anatomy branch in SNOMED CT).

2.4 Configuring the User Interface

We noted earlier that one of our key goals in designing WebProtégé was to have a tool that can be configured easily for many different settings, workflows, and types of users. Indeed, users can configure almost everything in the WebProtégé portlets, by describing the configuration in an XML file with a a predefined schema[7]. Building a new tool based on WebProtégé can be as simple as defining a layout configuration for existing portlets.

To support this flexibility, each portlet has a property list attached to it in the XML layout file, which we can use to provide additional configuration information. For example, the class tree portlet in Figure 3 displays only the disease hierarchy of the ICD Ontology, with the *ICDCategory* class as the root. We defined one property *topClass* of the portlet that points to the *ICDCategory* class in the configuration file. Thus, we can reuse the class tree portlet to display different class-tree views by simply changing a property of the portlet.

The declarative user interface also allows us to define custom views for different users. In WebProtégé, layout configurations can be defined per user and per project. Therefore, different users can see the same ontology rendered in different ways. One can imagine a scenario in which a user works only on a branch of an ontology, or one in which users should see only a selection of portlets. We can support these scenarios by defining different configuration files for users.

We mentioned earlier that portlets provide independent pieces of functionality. Therefore, we tried to avoid creating hard-coded dependencies between portlets in order to be able to reuse them in different configurations. For example, selecting a class in the class tree portlet should trigger the display of property values in a different portlet. Rather than hard coding this dependency, we defined a generic selection-model mechanism. Each tab has a *controlling portlet*—the portlet that provides the selection for the other portlets in the tab. Each time the selection in the controlling portlet (e.g., the class tree portlet) changes, the other portlets are informed via a listener mechanism about the change and can update their content accordingly. XML layout configuration file specifies the controlling portlet for a tab that can be changed at runtime.

2.5 Support for Collaboration

We implemented the collaboration framework on the server side (Section 2.2) and we expose it in the user interface. Distributed users can edit the same ontology simultaneously and see immediately the results of one another's changes.

Users can add notes to classes, properties, and individuals in the ontology. They can also reply to notes that were posted by others. At the time of this writing, there

[7] XML layout configuration examples available at: `http://tinyurl.com/y35qazg`

are more than 1,300 notes in the production version of WebProtégé for ICD. Notes may have different types, such as *Comment* or *Explanation*. When a user browses the class hierarchy, he can see the number of notes that are attached to each class, and the number of notes in the subclasses of that class. In Figure 3, the icon next to the class name indicates, for example, that the class *D04 Carcinoma in situ of skin* has two notes attached to it. The shaded icon next to it indicates that there are also two notes in the subtree rooted at this class. Knowing the number of notes in a subtree, enables users to identify quickly the branches of ontologies that have most activity and discussions, and also to find the notes that are attached somewhere deeper in the class hierarchy. Users can also attach notes to specific triples. For example, a user may want to comment on a particular definition of a disease. The user may do so by clicking on the comment icon next to a particular property value (see Figure 3). The *Notes and Discussions Tab* is a dedicated interface for browsing and creating notes and discussions.

WHO plans to use peer review to ensure the quality of the ICD content. In the current implementation, WebProtégé supports a prototypical implementation of a reviewing mechanism in the *Reviews Tab*. A user with the appropriate priviledges can request a review for a particular disease class. The user may choose from a list of predefined reviewers who are specialized on the particular domain of the disease. Once the review is complete, the reviewer may log into the system and add a review to a class. Internally, we represent Reviews as a specific type of notes in WebProtégé.

The WHO is still working to define the workflow of the ICD-11 revision process. We envision that WebProtégé will support this workflow in a generic and flexible way. Currently, we support only parts of the workflow. WebProtégé already has a generic access-policy mechanism, which we use to define the different user roles (TAG member, managing editor, etc.) and their access rights. The user interface enforces the access rights and we can configure it for different user roles. However, much remains to be done. The main workflow defining how the operations should flow for different user roles is still under development. We currently plan to expose the WebProtégé platform to a larger audience, which will likely have a lower level of expertise than the current users. Members of this broader community should be able to make proposals for changes. We are currently working out the details on how such a proposal mechanism should work. Once we have a well-defined workflow, we will investigate how to develop the tool to support a flexible and generic workflow mechanism.

3 Evaluation

We evaluated the customized WebProtégé tool during *iCamp*—a two-week meeting of the members of the ICD-11 revision project. The meeting took place in September 2009 in Geneva. It brought together editors who will manage the revision process and classification experts. The goal of the iCamp was to discuss the plans for the ICD-11 revision and to gain experience using the WebProtégé software. The objective of the evaluation that we performed during that meeting was two-fold: (1) perform formative evaluation of the WebProtégé software and determine requirements for further development of the tool; and (2) to use WebProtégé and the users' experience with it during the iCamp as the context to evaluate the feasibility of and requirements for collaborative ontology-development process in general. In this paper, we report on the results of the second part of this evaluation—analysis of collaborative ontology development.

3.1 Research Questions for the Evaluation

Our objective for this evaluation was to determine how a collaborative authoring tool can support a diverse community of domain experts in developing a large terminology. Specifically, we wished to address the following research questions:

Q1: Do domain experts find a collaborative development process promising?
Q2: What are the features that users find useful? Which features are required?
Q3: What is the workflow for collaborative ontology-development that a tool must support?
Q4: What coordination and communication mechanisms do the users need?

By answering these questions, we hope to elicit further requirements for tool support for large scale collaborative ontology development and to understand better the process of collaborative development itself.

3.2 Participants

The participants involved in the tool evaluation consisted of eleven medical professionals that will be working as managing editors for ICD-11 and nine classification experts working to ensure the integrity of the terminology. The domain experts (i.e. managing editors) had varied areas of expertise: rare diseases, dermatology, external causes and injury, and so on. Each managing editor is responsible for the development of the part of ICD-11 in his area of expertise. The revision of ICD-11 is a large international project, thus, English is not the native language for many of the participants.

3.3 Materials and procedures

Developers of the WebProtégé software introduced the iCamp participants to the customized WebProtégé tool through a series of tool demonstrations and presentations. Then participants worked in pairs over several semi-structured ICD-11 editing sessions. Each of these sessions lasted for two to three hours, and took place over four of the iCamp days. The organizers instructed the pairs to explore the areas of the current ontology that they will be responsible for as managing editors and to begin filling out the different attributes of the content model. The organizers encouraged iCamp participants to use social media, such as Twitter and Facebook. There was also a team shooting daily videos—a few minutes each—that described the activities in the iCamp for that day and included interviews with participants, software developers, and iCamp organizers.[8]

Our evaluation consisted of two parts: a **survey** of the participants and a **focus group**. At the last day of the iCamp meeting, participants filled out a Web-based survey. Table 1 shows the survey questions. We also conducted a focus group with the managing editors on the last day of the first week of the iCamp. The focus group was moderated by a researcher and the four questions shown in Table 2 guided the discussion.

3.4 Results

In this section, we discuss results from the survey and the focus-group discussion. We use these results to derive findings that help address our research questions (Section 4).

[8] `http://www.youtube.com/user/whoicd11`

#	Question
1	Do you think the WebProtégé tool is developing into the right tool for authoring ICD-11?
2	Do you have faith that the open process of developing ICD-11 will succeed?
3	What are the two or three most important features you feel need to be incorporated into WebProtégé?
4	Did you find the use of social networking and media tools during iCamp like YouTube, Facebook, Twitter, and Blogger to be useful to the process?
5	Do you have any suggestions about how to evaluate the process of developing ICD-11?

Table 1. iCamp survey: Impressions about the collaborative development process

#	Question
1	Will the current WebProtégé approach/process work for you in terms of how you want or need to work as a managing editor?
2	Will the current annotation/commenting support fulfill your collaboration needs?
3	Based on what you have been able to produce so far during iCamp, do you feel that you will be able to develop an ICD-11 that is going to address your use cases or is the project moving in the wrong direction?
4	In terms of the review process, how do you see the process working?

Table 2. Questions to guide the iCamp focus group

Survey Results

The goal of the survey (Table 1) was to evaluate the participants' impressions of the proposed open process for developing ICD. Eleven participants responded to the survey.

We first asked participants whether they felt that the approach taken by WebProtégé was the right one for the development of ICD-11. The responses were primarily positive. Five of the eleven responses were a clear "yes," others were optimistic but had some concerns. For example, one respondent indicated that the "tool is focused on expert editing, not casual users."

The second question asked whether the participants felt that the open process of developing ICD-11 would succeed. Four out of eleven respondents answered "yes" to this question, while others had some level of concern. In particular, one participant was concerned that an open process will engage users with limited knowledge. Monitoring the contributions of these users will require extra time on behalf of the managing editors. Others were concerned that the editors will be overwhelmed with the feedback that they get or that a lack of ontological expertise among the participants will result in a poor terminology. However, one respondent felt that the "wisdom of the crowd is good."

Following this question, participants were allowed to describe their top three feature requests. Seven participants responded to this question with at least one feature request. The following features were listed by more than one participant: six respondents requested better hierarchy management support, two wanted to view the list of proposals in progress, two requested communication support, and two requested status reports.

We also asked participants about the use of social networking and media tools during iCamp. There was mixed feedback, only a few of the responses were positive. Many participants felt that face-to-face meetings are far more effective, and that the information needs to be centralized rather than spread across multiple social networks. One respondent suggested that the information should be integrated into the WebProtégé tool and one felt that the use of so many media was "media overkill."

Next we asked for suggestions on how to evaluate the development process of ICD-11. Respondents were keen on eliciting feedback about the development through future face-to-face meetings as well as periodic web-based surveys. One respondent also suggested monitoring participation by editors and the general public.

Focus group

The focus group consisted of a two-hour semi-structured moderated discussion. The participants felt that WebProtégé was a good initial step, but a lot of work needed to be done in terms of supporting an open collaborative process where anyone can propose a terminology change. Participants indicated that it was unclear to them at that moment what this process should entail in terms of peer review and conflict resolution.

Participants indicated that the current WebProtégé support for commenting was very granular, enabling users to comment on anything in the system. Participants proposed that another, higher-level type of discussion also needed to be supported. Such discussions would not be attached to any specific element in the terminology.

Awareness and tracking of changes to the model was also a key discussion issue. Participants felt that appropriate users need to be notified when certain changes take place. Moreover, participants felt that the WebProtégé model, where changes are immediately committed and visible to everyone, was "too permanent": There was no obvious way to experiment with changes to the terminology without having the changes immediately impact everyone else who was editing or navigating the terminology.

Finally, participants felt that the process of managing proposed changes needed to be defined better and eventually supported by WebProtégé. Participants raised the concerns about privacy for the reviewers of a change. Some participants felt that if a review was not anonymous, then reviewers would be reluctant to be completely honest about a proposal or that it would create more pressure for the reviewers.

4 Findings and Discussion

In this section we return to the research questions that we introduced in Section 3.1. Based on the results from our two evaluation procedures, we attempt to answer these questions in order to address our research objective.

4.1 Q1: Do domain experts find a collaborative development process promising?

Our survey results indicated that participants had mixed impressions about the collaborative development process. Their central concern about the open process was not only the resources that are required in order to develop the process and software for handling such an ambitious goal, but also the time it would take to manage change proposals by unqualified users of ICD-11. Such "crowdsourcing" approach has the advantage that potentially many participants will help with the development process and with maintaining the integrity of the terminology, yet it also opens the development to anyone who is interested, regardless of their expertise level.

We can compare the collaborative-development process of a terminology, such as ICD-11, to other "crowdsourcing" projects. We can draw parallels, for example, with open-source software development: the potential number of contributors is probably on a similar scale to the number of contributors to specialized terminologies, since making contributions requires expertise in a specific domain of knowledge (software development in one case, medicine in the other).

In open source software development, the openness of the source code can actually help increase the overall quality of the code. Because everything that a developer produces is freely available to the public, there is both social pressure to make the source code as readable, maintainable, and stable as possible and there are more people available to evaluate the code quality: "given enough eyeballs, all bugs are shallow" [10]. Mockus and colleagues [7] found that both open source projects Mozilla and Apache had very low defect density levels after release in comparison to commercial software, in part due to the larger community of users contributing to bug reports.

Few code contributions to large open-source projects come from unqualified developers. Instead, highly qualified developers often make up the core set of contributors, as their only incentive is to improve the software for their own use or for the intellectual challenge of contributing to the tool [6, 16]. Finally, since open-source developers often pick and choose what to contribute to, they choose the components that most interest them. This flexibility ensures that the developers are typically engaged in their work, versus a commercial project where they are generally assigned work, regardless of interest [7].

Researchers have observed similar results have with the articles contributed to Wikipedia. The online encyclopedia provides a means for those with specialized knowledge to share that information with the rest of the world [1].

However, there is also a significant difference between contributing to open-source software projects or contributing to Wikipedia and contributing to ICD-11. Contributing to source code requires certain level of technical expertise with the tools, if only to compile your contribution. Tools such as WebProtégé are designed to require little or no technical expertise to contribute. Thus, we must develop other mechanisms to control who can contribute content. For instance, we may require that contributors to ICD-11 have accounts that are validated by WHO or its subsidiaries in the respective countries.

At the same time, Wikipedia benefits from an extremely large user community, which helps ensure its quality. Naturally, the number of users who can understand and evaluate the ICD-11 content is far smaller than the number of Wikipedia users. Hence, WHO needs to develop a more structured revision and quality-assurance process.

4.2 Q2: What are the features that users find useful? Which features are required?

Among the features that participants felt were critical for a tool to support collaborative terminology development, multi-lingual support was mentioned by several participants both in the survey and during the focus group discussion. In a large, international development effort, this feature is critical. Participants also requested better hierarchical editing support. Although the users found the terminology navigation easy to use, they struggled with making modifications to the actual hierarchy. This is particularly interesting, as WebProtégé uses the same hierarchy navigation and editing paradigm as many other ontology-development tools. The participants struggled with understanding how to rename category entries or moving parts of the hierarchy using the drag and drop feature. They also struggled with tracking and revising changes that they made.

In collaborative terminology development, awareness of changes to the terminology are extremely important. Editors will need to be aware of change proposals as well as to have support for working together with other editors to incorporate or review proposals. Users will need to be able to see an overview of an entire change proposal. Also, users will need to know when major changes take place, changes that may influence their use of the terminology.

The software will need to support an audit trail for tracking how a category and its associated attributes change over time. There will need to be visibility about who made certain changes and how a proposal was reviewed. As mentioned, in the WebProtégé version that we used for the evaluation, it was difficult to track your own changes and navigation history, both of which will need to be present to help support the continued development of the terminology.

4.3 Q3: What is the workflow for collaborative ontology-development that a tool must support?

The workflow for incorporating change proposals to the terminology was a major concern of the participants in our evaluation. Currently, WebProtégé allows any authorized user to make direct changes to the terminology. However, once the wider community starts contributing to the content, they will be able to do so by creating change proposals. General users will make the proposals for changes, and there will be a series of steps that will be required to review and accept the proposal. For example, managing editors will receive the change-proposal and determine whether it should be reviewed. If review is necessary, the change-proposal will pass through a peer review stage and if approved, the change-proposal will then be reviewed by the Topic Advisory Group.

Any large scale collaborative development process will need tool support for incorporating, tracking, and reviewing changes. The exact process for performing this tracking and how it can be best supported is not yet clear, but it will most likely blend ideas from journal review processes, open source software development, and existing work on conflict resolution in ontology versions [5]. However, our analysis of workflows in several large-scale collaborative ontology-development projects [12, 8, 13] shows that the workflow for each project is different and involves different steps required by a changed approval. Indeed, there are several efforts to develop a declarative representation of customizable collaborative workflows [2, 11].

4.4 Q4: What coordination and communication mechanisms do the users need?

Our participants indicated that they needed at least two levels of communication support: low-level commenting on ontology elements and changes (already supported by WebProtégé) and higher-level discussion that could encompass multiple parts of the terminology (supported by the collaboration framework on the Protégé server, but not yet exposed in the user interface). Separate, perhaps private, communication support mechanisms will be necessary for editors working together to review proposals. These comments will possibly involve cross-cutting concerns that include categories within separate branches of the hierarchy or may be of a general nature, for example, comments relating to the review and editing process. The results or verdict of these reviews will also need to be communicated to the public.

Our analysis did not produce a clear picture of what role, if any, common social-networking sites, such as Twitter and Facebook, can play in the social process of terminology development. The participants did not enthusiastically embrace the use of this media in the context of iCamp. However, this may have been in part due to the participants all being co-located, thus, face-to-face communication was most effective.

Again, drawing parallels to open source software (OSS) development, successful projects like Apache and Mozilla rely heavily on somewhat archaic technologies like mailing lists and bugzilla as their main coordination and feedback tools [7]. However, these are older OSS projects. Newer companies like EclipseSource[9], make heavy use of tools like Twitter, blogs, YouTube, and Skype. There is great potential to re-use such tools in collaborative ontology development, but there needs to be buy in from the users as well as integration into the development process. As the ICD-11 project progresses, contributors will be working distributively. At this point, existing social-networking sites may play a larger role.

5 Conclusions

We have presented WebProtégé, a Web-based tool for distributed collaborative development of ontologies. WebProtégé is currently being used by the World Health Organization as the primary development environment for ICD-11, a key international medical terminology. Developers of other terminologies within the WHO Family of International Classifications (WHO-FIC) are beginning to use WebProtégé as well.

The open development of ICD-11 promises to be an exciting experiment in how far collaborative terminology development can go. At the same time, WHO members depend critically on the high quality and timely availability of the 11th revision of the ICD. Hence, the cost of failure is extremely high. However, the ICD experts believe that opening ICD development to a wider international audience is the only way to assure its comprehensive coverage, availability in multiple languages, and correspondence to the latest understanding of clinical practice.

Participants in our evaluation—none of whom are ontology experts or have previously participated in such an open development process—were optimistic that collaborative development will indeed work in this context. They have raised a number of critical issues, including the trade-off of having more experts contribute to the content and requiring more resources to assess the quality of these contributions. They also highlighted critical requirements for tool support such as change tracking, awareness of the state of the change proposals, discussions and notifications mechanisms.

We are currently working on the next version of WebProtégé that will address many of the requirements for the initial state of the development. Simultaneously with the tool development, ICD-11 contributors are already developing and extending content using the current version of the tool. Indeed, in the past month, 12 users have made over 4000 changes and more than 1500 comments.

Acknowledgments

We are very grateful to Jennifer Vendetti, Timothy Redmond and Martin O'Connor, and for their help with the design and implementation of WebProtégé and to Robert Jacob, Can Celik, and

[9] http://eclipsesource.com/

Sara Cottler for their work in developing the requirements for the project and in organizing and running the iCamp. Samson Tu has been instrumental in designing the ICD Content Model and his help during iCamp was invaluable. The work presented in this paper is supported by the NLM Grant 1R01GM086587-01. Protégé is a national resource supported by grant LM007885 from the U.S. National Library of Medicine.

References

1. N. Baker. The charms of wikipedia. *The New York Review of Books*, 55(4), 2008. Retrieved from: http://www.nybooks.com/articles/21131.
2. A. Gangemi, J. Lehmann, V. Presutti, M. Nissim, and C. Catenacci. C-ODO: an owl meta-model for collaborative ontology design. In *Workshop on Social and Collaborative Construction of Structured Knowledge at 16th Intl. WWW Conf. (WWW2007)*, Banff, Canada, 2007.
3. GOConsortium. Creating the Gene Ontology resource: design and implementation. *Genome Res*, 11(8):1425–33, 2001.
4. F. W. Hartel, S. d. Coronado, R. Dionne, G. Fragoso, and J. Golbeck. Modeling a description logic vocabulary for cancer research. *Journal of Biomed Informatics*, 38(2):114–129, 2005.
5. E. Jiménez-Ruiz, B. C. Grau, I. Horrocks, and R. B. Llavori. Building ontologies collaboratively using contentcvs. In *Description Logics*, volume 477 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2009.
6. K. R. Lakhani and R. G. Wolf. Why hackers do what they do: Understanding motivation and effort in free/open source software project. Techn. Report Working Paper 4425-03, MIT Sloan School of Management, September 2003.
7. A. Mockus, R. T. Fielding, and J. D. Herbsleb. Two case studies of open source software development: Apache and mozilla. *ACM Transactions Software Engineering Methodology*, 11(3):309–346, 2002.
8. O. Muñoz García, et al. A Workflow for the Networked Ontologies Lifecycle: A Case Study in FAO of the UN. In *The Conf. of the Spanish Association for AI (CAEPIA 2007)*, vol.LNAI 4788, p200–209. Springer-Verlag, 2007.
9. N. F. Noy, et al. Bioportal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research*, 10.1093/nar/gkp440, 2009.
10. E. S. Raymond. *The Cathedral & the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. O'Reilly Media, 2001.
11. A. Sebastian, N. F. Noy, T. Tudorache, and M. A. Musen. A generic ontology for collaborative ontology-development workflows. In *16th Intl. Conf. on Knowledge Engineering and Knowledge Management (EKAW 2008)*, Catania, Italy, 2008. Springer.
12. N. Sioutos, S. de Coronado, M. Haber, F. Hartel, W. Shaiu, and L. Wright. NCI Thesaurus: A semantic model integrating cancer-related clinical and molecular information. *Journal of Biomed. Informatics*, 40(1):30–43, 2007.
13. C. Tempich, E. Simperl, M. Luczak, R. Studer, and H. S. Pinto. Argumentation-based ontology engineering. *IEEE Intelligent Systems*, 22(6):52–59, 2007.
14. S. W. Tu, et al. A content model for the ICD-11 revision. Technical Report BMIR-2010-1405, Stanford Center for Biomed Informatics Research, 2010.
15. T. Tudorache, N. F. Noy, and M. A. Musen. Supporting collaborative ontology development in Protégé. In *7th Intl. Semantic Web Conference, ISWC 2008*, Karlsruhe, Germany, 2008.
16. G. von Krogh, S. Spaeth, and K. R. Lakhani. Community, joining, and specialization in open source software innovation: a case study. *Research Policy*, 32(7):1217 – 1241, 2003. Open Source Software Development.