Collaborative Ontology Development

Tania Tudorache

Protégé Short Course

Stanford University
October 11, 2017

Collaborative Ontology Development

- **Collaboration**: a process in which a community of users contribute to the development of one or more ontologies
- Ontologies vary quite a lot: from simple taxonomies to fully fledged OWL
- Usually the community is made of (some) ontology experts and (several) domain experts with varying expertise
- Common concern: how to get the domain experts to contribute without overloading them with representation details
- Think how you author a Word document with your colleagues: same and more challenges apply to ontologies!

Lorem Ipsum

Introduction

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed vitae faucibus enim, non viverra turpis. Aenean semper hendrerit diam vel ultrices. Maecenas porttitor ante lectus, a porta justo pellentesque eu. Donec elementum justo suscipit viverra hendrerit. Donec at egestas purus. Interdum et malesuada fames ac ante ipsum primis in faucibus. Morbi hendrerit massa ligula, non egestas justo vestibulum sit amet. Suspendisse et rhoncus massa. Integer ultrices tempus erat. Quisque eu mauris mi. Donec eu consectetur sem. Morbi libero ipsum, aliquet vitae ex id, suscipit sollicitudin orci. Suspendisse egestas dolor tellus, in malesuada diam condimentum eget. Morbi imperdiet sapien non risus auctor, eget fringilla sem maximus. Cras commodo ultricies metus id tempus.

2. Methods

Definition: Duis ut lorem venenatis, interdum est sit amet, facilisis nunc. Curabitur dapibus, neque ut tempor semper, nunc ex hendrerit sapien, eu hendrerit ex dolor et est. Integer sit amet mauris aliquam leo vehicula tristique sit amet nec nunc. Nulla commodo est quis tellus dictum vehicula. Nunc et luctus libero. Phasellus finibus semper maximus. Fusce sit amet libero sem. Suspendisse potenti. Donec lorem tortor, venenatis et sem ac, volutpat lacinia dolor. Suspendisse tristique metus non semper posuere. Proin at suscipit elit. Nam vel augue turpis. Proin sed ex ultricies, tincidunt ligula ac, placerat tellus.

2.1. Lorem Method

Nam porttitor mollis enim, quis hendrerit ipsum elementum in. Integer odio dui, hendrerit eu consectetur at, ullamcorper eget leo. Nulla facilisi. Fusce nisl felis, lacinia vel dictum ut, rutrum eget neque. Nam pellentesque ipsum sed lacus sollicitudin placerat. Morbi in nunc id lacus maximus luctus. Nullam eget neque et nulla convallis accumsan. Pellentesque elit ligula, venenatis sed odio et, viverra venenatis tortor.

Google Docs

Contributors:

Roles:

- writer
- reviewer
- proof-reader

Structure

Style

Workflow

Quality Assurance

Publishing

2 parts: Editing and Publishing

Editing

- “Internal” development team
- Tools for editing ontologies
- Short development cycles

What your team does

Publishing

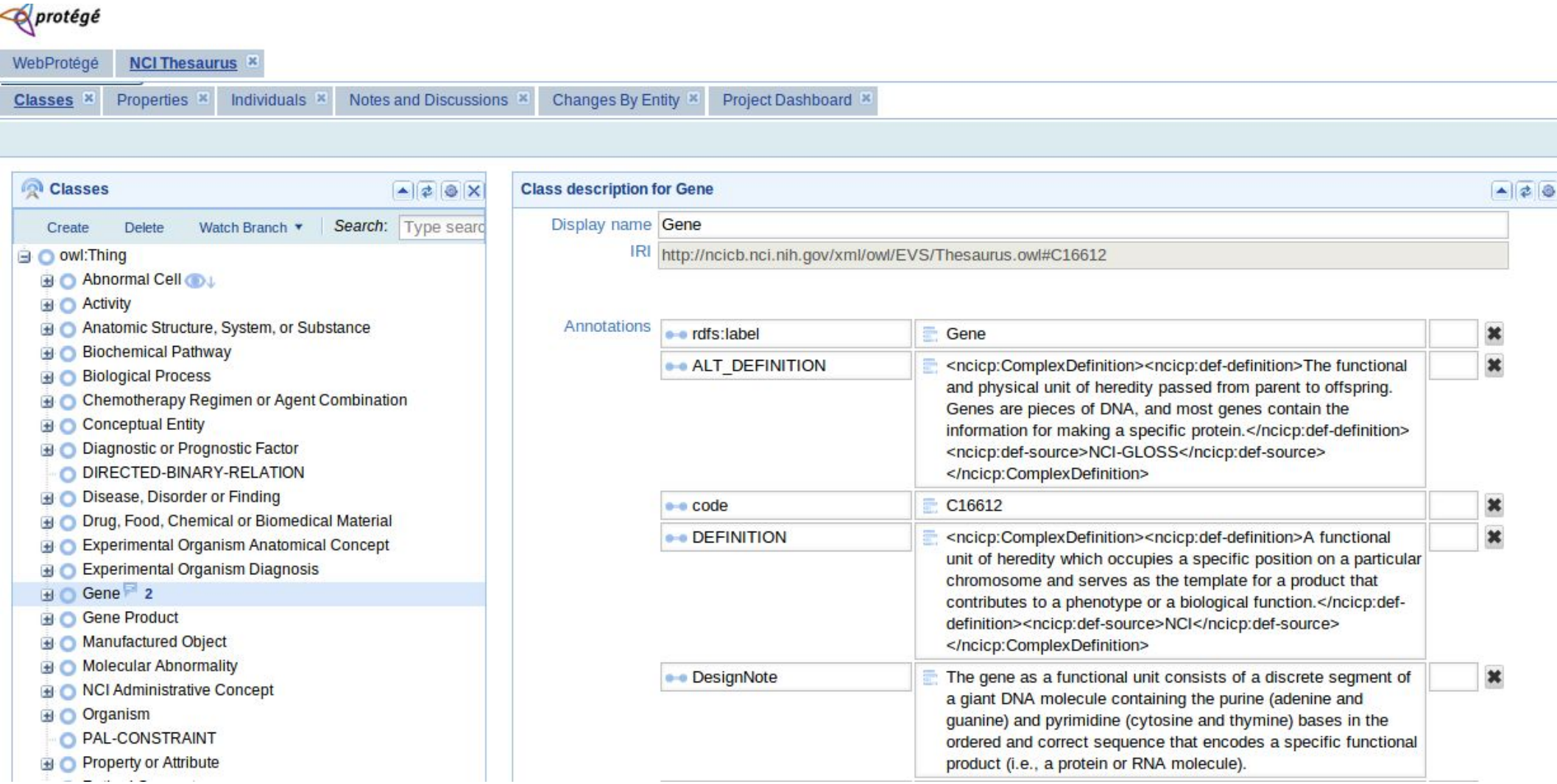
- External release of ontologies for public/production use
- Tools for supporting versioning, public feedback and reviewing
- Set release cycles

How others use your ontology

Use cases of collaborative development

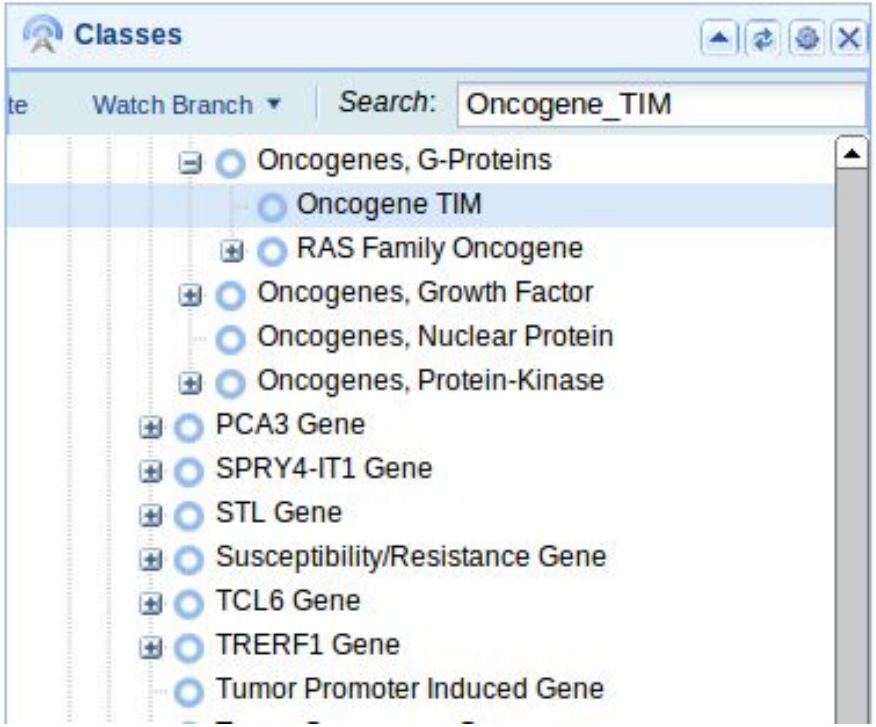
Use case: The NCI Thesaurus collaborative development process

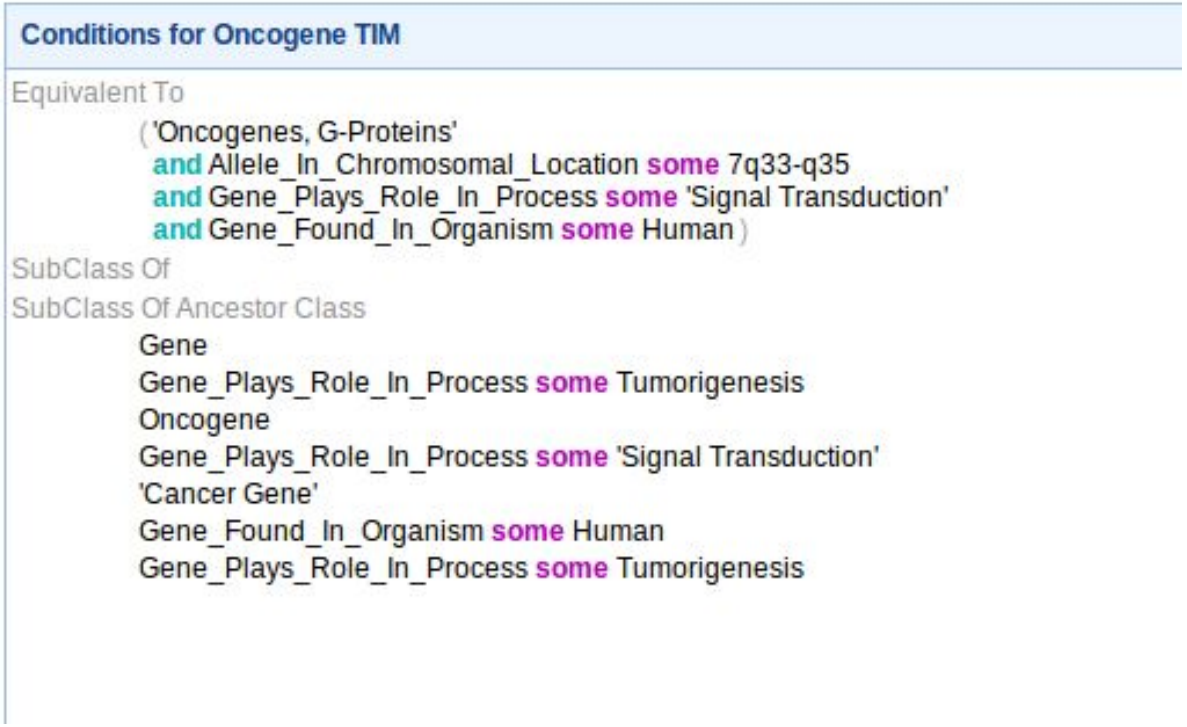
NCI Thesaurus: Reference ontology for cancer biology, translational science, and clinical oncology



Use case: The NCI Thesaurus collaborative development process (cont.)

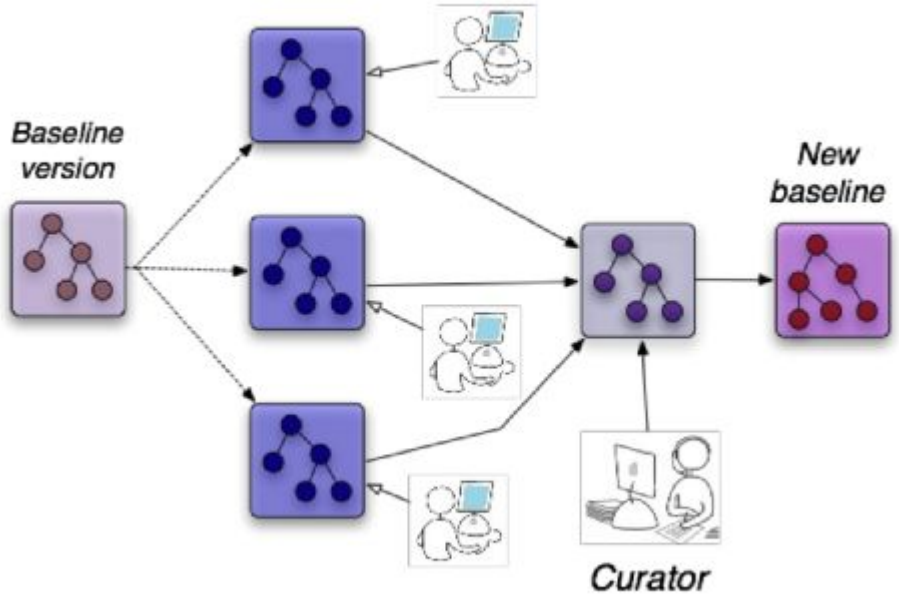
NCI Thesaurus: Reference ontology for cancer biology, translational science, and clinical oncology





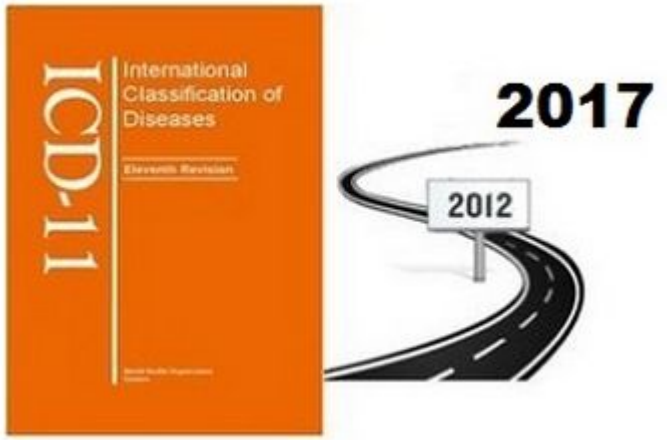
Use case: The NCI Thesaurus collaborative development process (cont.)

- Simultaneous editing in Protégé clients
- Custom UI for restricting user input and enforcing business rules
- Development cycle begins after baseline
- ~20 full-time editors making changes; 1 lead editor approves the changes, and assigns new tasks
- Released version on NCI website and BioPortal

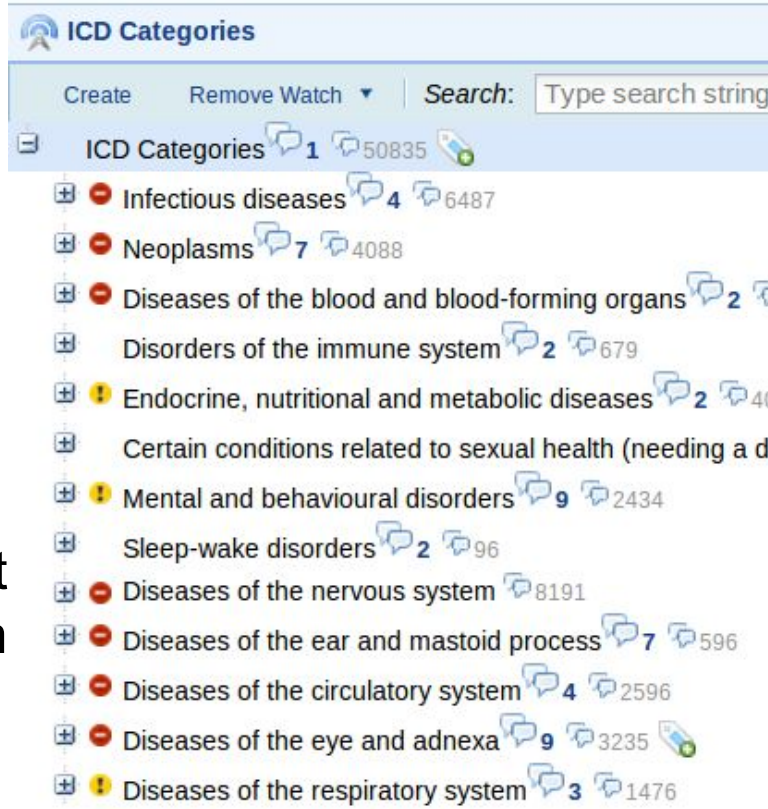


http://bioportal.bioontology.org/ontologies/NCIT/?p=summary

Use case: WHO's ICD



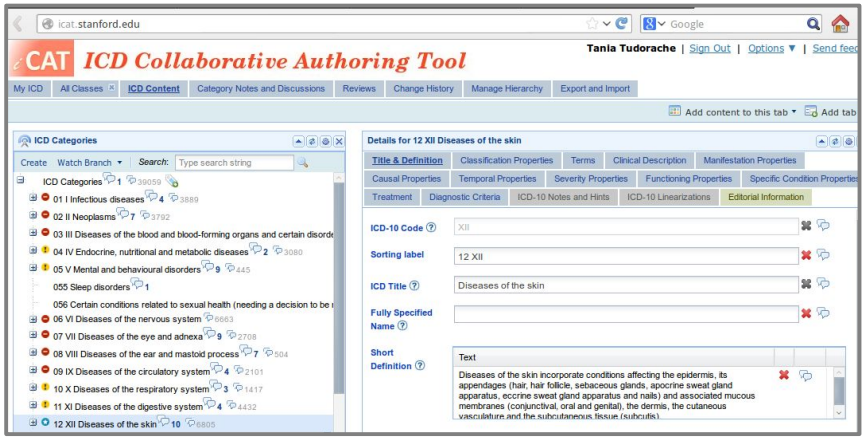
- ICD – International Classification of Diseases
- Developed by the World Health Organization (WHO)
- Current revision in use: ICD-10, development work on ICD-11
- Over 10.000 categories used for coding, billing, statistics, policy making all over the world
- Collaborative and international effort
- Current version: published as books
- Goal for the new version: use a more formal representation and published in electronic format use Web-based collaboration and social platform for editing



Construction of ICD-10: Revision Process in the 20th Century

- 8 Annual Revision Conferences (1982 - 89)
- 17 – 58 countries participated
  - 1- 5 person delegations
  - Mainly health statisticians
- Manual curation
  - List exchange
  - Index was done later
- "Decibel" Method of discussion
- Output: Paper Copy
- Work in English only
- Limited testing in the field

ICD-11 process today



- ICD-11 is an OWL ontology edited collaboratively in WebProtégé
- Over 250 domain experts from around the world
- Organized in groups, which edit different parts of the ontology

ICD-11 process today (cont.)

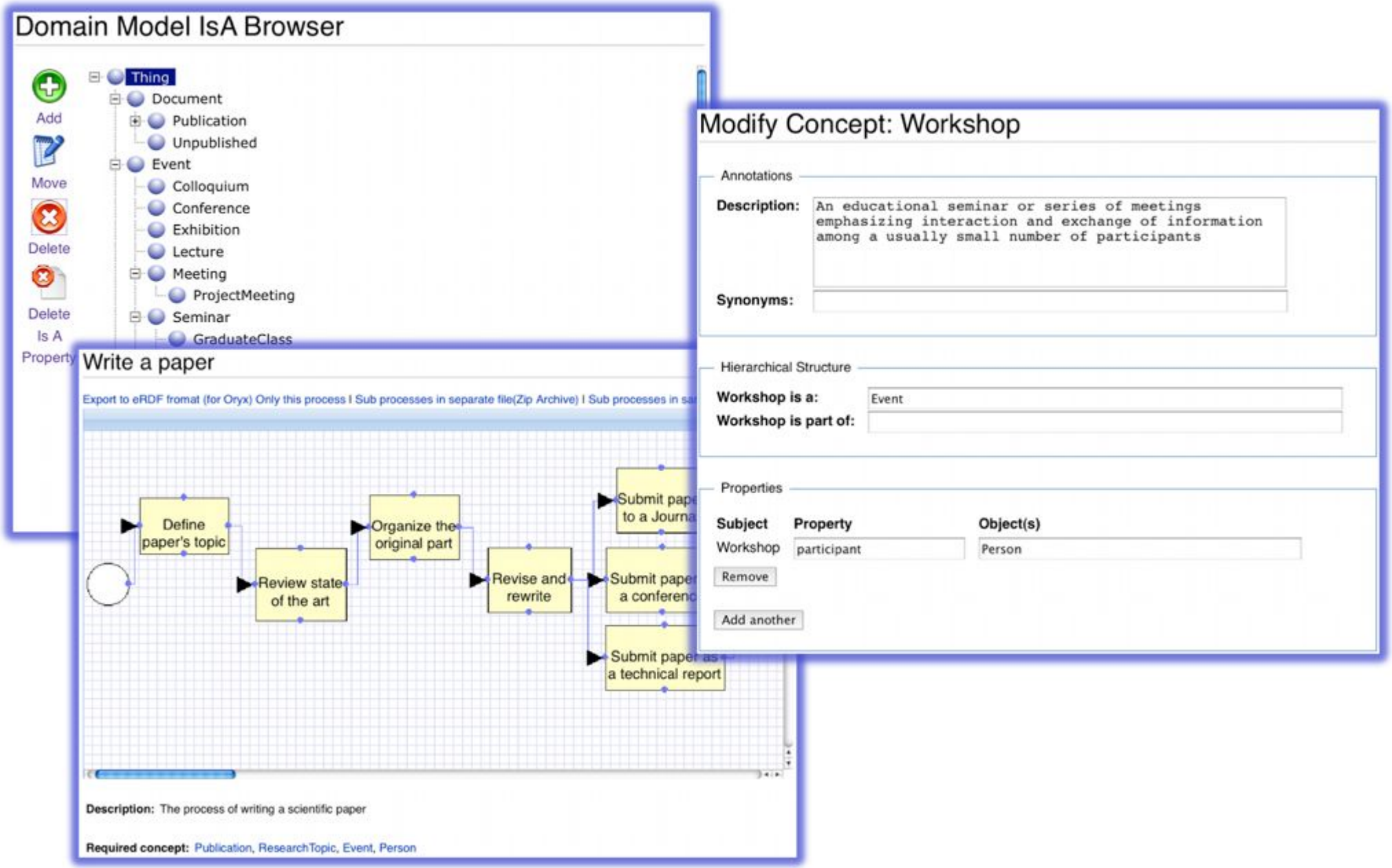
- Each night a snapshot of the commonly edited ontology is published in a public platform to encourage feedback from the larger community
  http://apps.who.int/classifications/icd11/browse/f/en
- Editorial workflow
- Centrally overseen by WHO
- Peer-reviewed process for the content and structure
- WebProtégé used as the collaborative ontology development platform

Collaboration infrastructures and more use cases

Collaboration Infrastructures: Wikis

- Wikis are well known; Wikipedia
- Semantic Wikis – add semantic extensions to the wiki platforms
- Assign a wiki page to an entity in the ontology
- Usually focused on filling in a knowledge base, less on the classes
- Export/import RDF

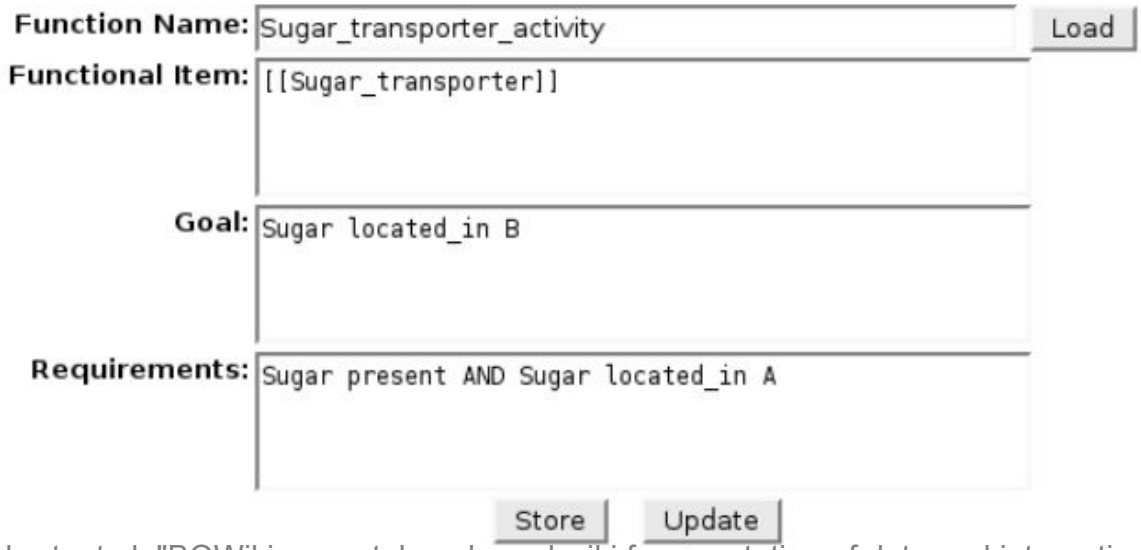
Semantic Wiki: MoKi



Source: https://moki.fbk.eu/website/userfiles/image/entmod.png

The challenge with wikis

BOWiki syntax	OWL abstract syntax
Examples	
1 on page Apoptosis: [[OType:Category]]	Individual(Apoptosis, type(Category))
2 on page Apoptosis: [[CC-isa::Biological_process]]	Individual(Apoptosis value(CC-isa Biological_process))
3 on page HvSUT2: [[Realizes:: function = Sugar_transporter_activity; process = Glucose_transport]]	Individual(Realizes-0 type(Realizes)) Individual(Realizes-0 value(Realizes-subject HvSUT2)) Individual(Realizes-0 value(Realizes-process Glucose_transport))
4 on page Realizes: [[has-argument:: name = function; type = OType:Function_category]]	SubClassOf(Realizes gfo:Relator)) ObjectProperty(Realizes-function domain(Realizes) range(Function_category))

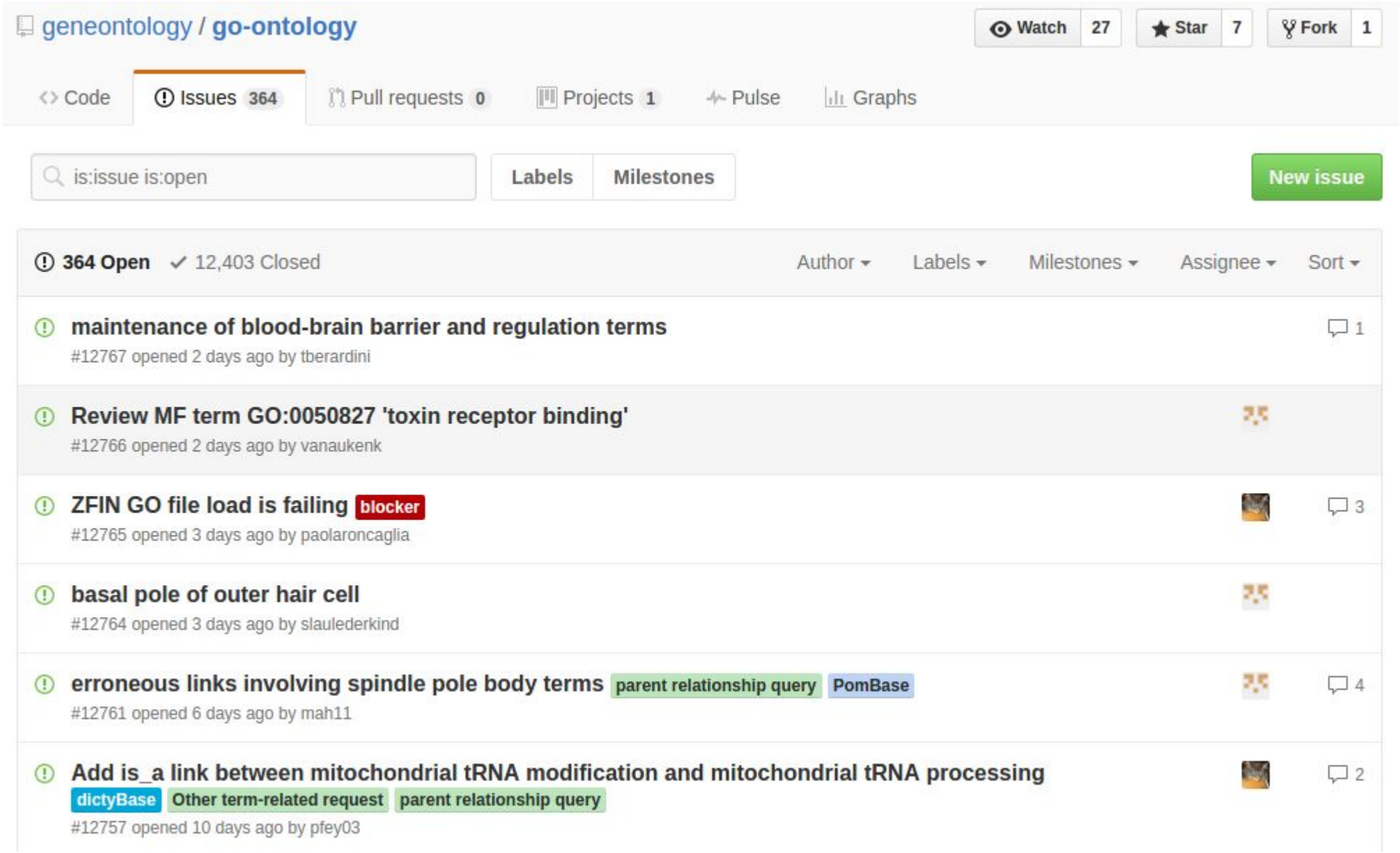


Source: Hoehndorf, Robert, et al. "BOWiki: an ontology-based wiki for annotation of data and integration of knowledge in biology." BMC bioinformatics 10.Suppl 5 (2009): S5.

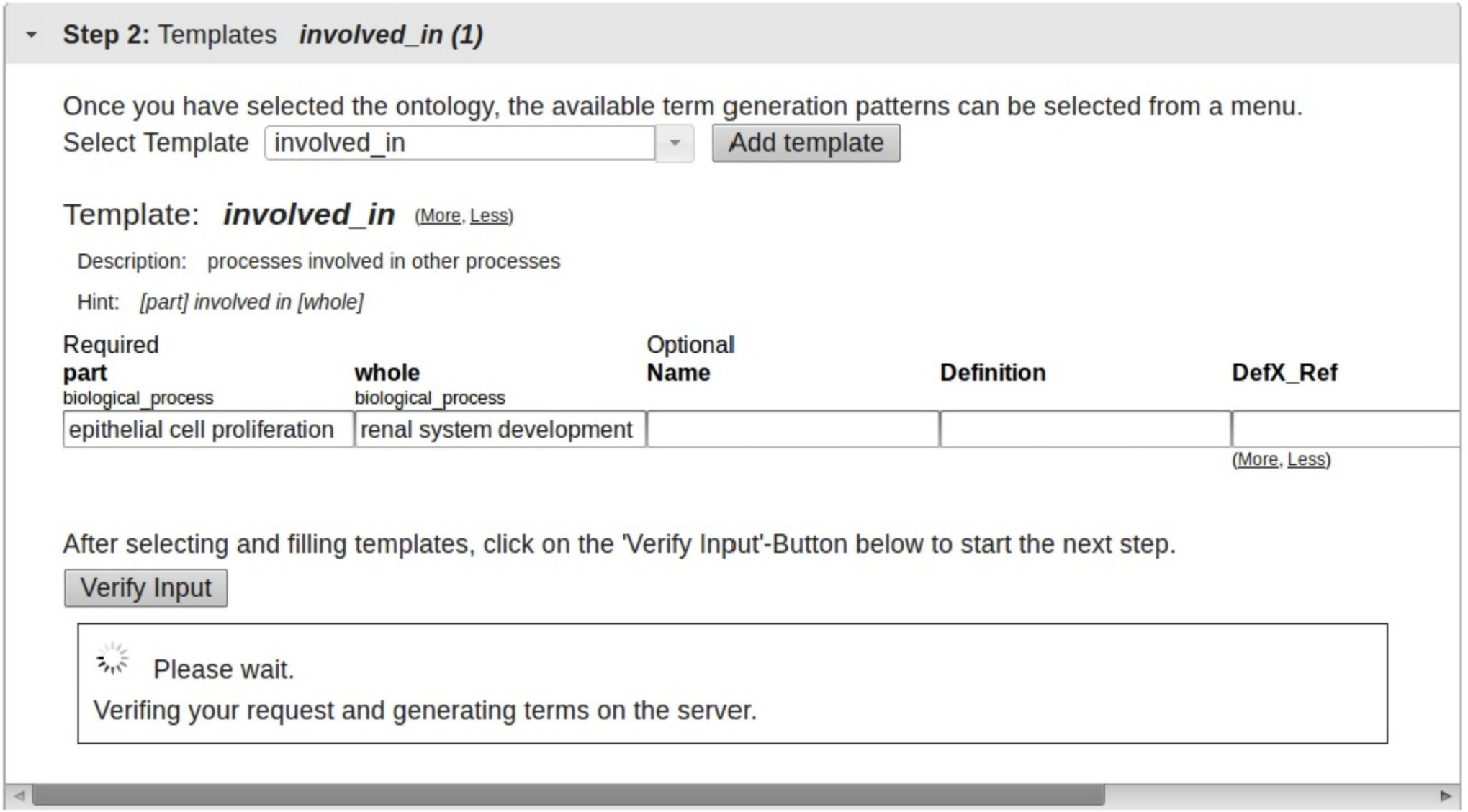
Project management platforms, content management platforms or issue trackers

- Some collaboration projects reuse Web-based platforms built for software and other types of projects
- Example: GitHub, SourceForge, GForge, RedMine, WordPress
- Issue trackers: Mostly for managing change proposals or reviews
- Content Management platforms: tagging articles with existing vocabularies (WordPress has some semantic extensions)

Gene Ontology Issue Tracker in GitHub



Gene Ontology - TermGenie



Other collaboration processes

- Use source control repositories (e.g., SVN)
  - Text based mechanisms
  - Hard to merge local copies in the shared copy
- Locking mechanisms (lock parts of an ontology for editing)
- Use specialized (domain dependent) ontology repositories, e.g., BioPortal

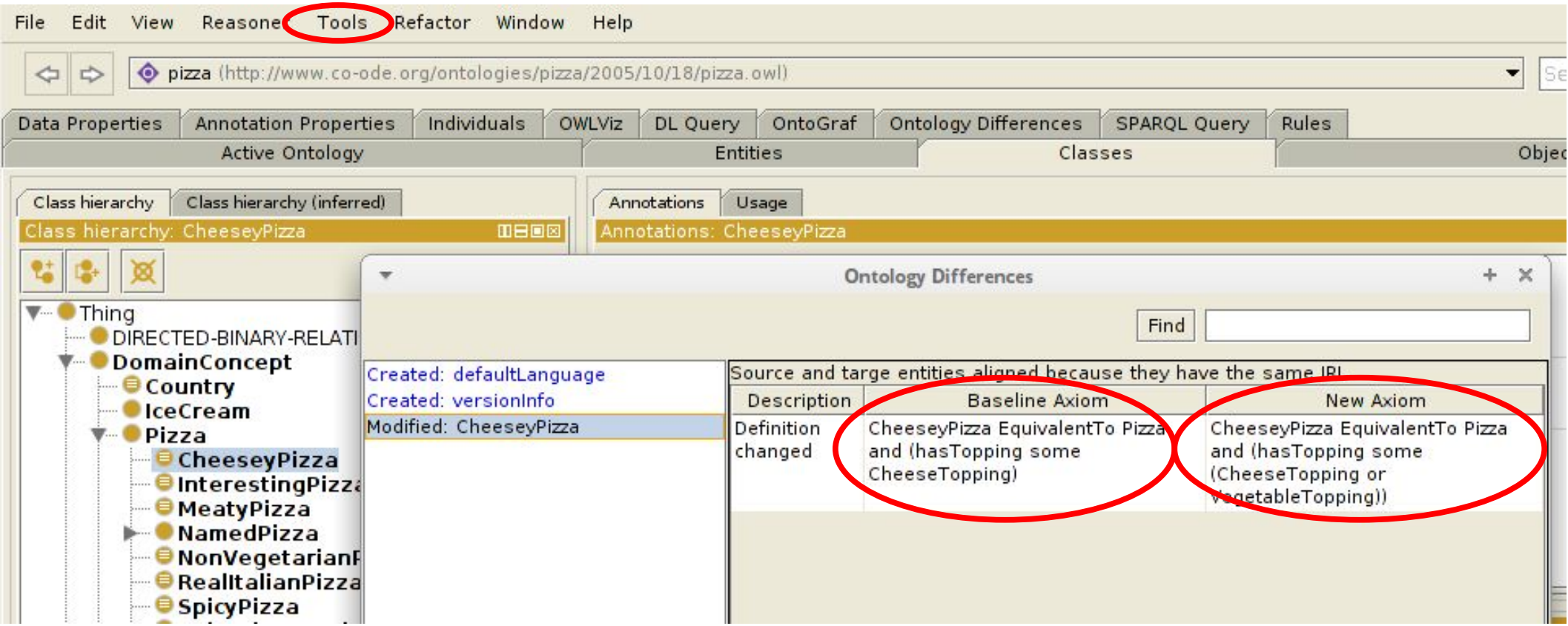
Textual diffs can be tricky

Textual diff: 588 lines, including the “only” change:

```
648c648,655
<    <owl:someValuesFrom rdf:resource="http://www.co-ode.org/ontologies/pizza/2005/10/18/pizza.owl#CheeseTopping"/>
---
>    <owl:someValuesFrom>
>      <owl:Class>
>       <owl:unionOf rdf:parseType="Collection">
>         <rdf:Description rdf:about="http://www.co-ode.org/ontologies/pizza/2005/10/18/pizza.owl#CheeseTopping"/>
>         <rdf:Description rdf:about="http://www.co-ode.org/ontologies/pizza/2005/10/18/pizza.owl#VegetableTopping"/>
>       </owl:unionOf>
>      </owl:Class>
>    </owl:someValuesFrom>
```

Protégé diff support:

Tools menu → Compare Ontologies



See demo screencast at: https://www.youtube.com/watch?v=JzMNDfy4jcg

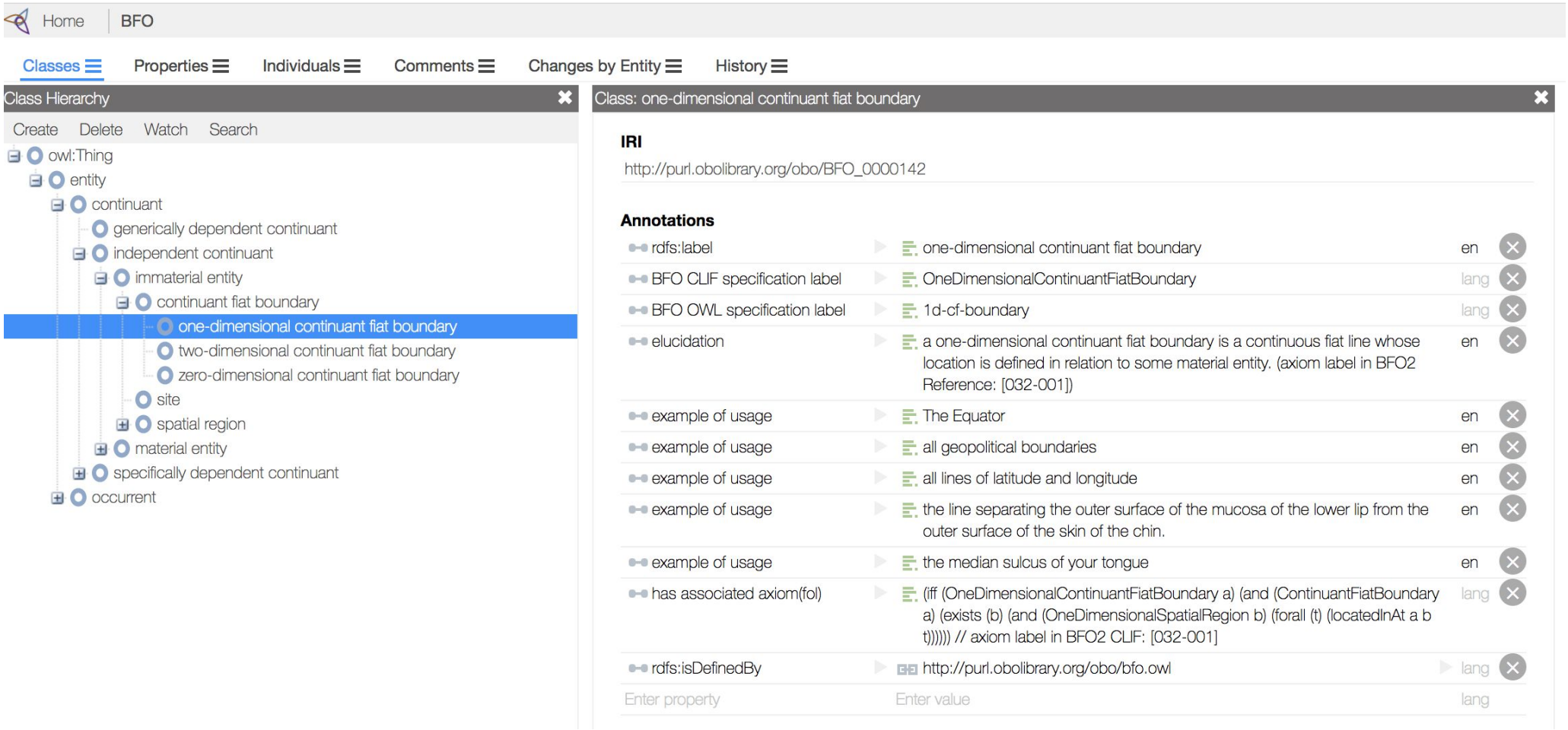
Collaboration support in WebProtégé

WebProtégé – quick overview

- **Free, open source collaborative ontology development environment for the Web**
- *Google docs* for ontologies; over 30,000 ontologies submitted or created by users
- OWL 2 ontologies
- A default simple editing interface
- Full change tracking and revision history
- Collaboration tools such as, sharing and permissions, threaded notes and discussions, watches and email notifications
- Customizable user interface
- Customizable Web forms for application/domain specific editing
- Multiple formats for upload and download of ontologies (supported formats: RDF/XML, Turtle, OWL/XML, OBO, and others)

http://protegewiki.stanford.edu/wiki/WebProtege

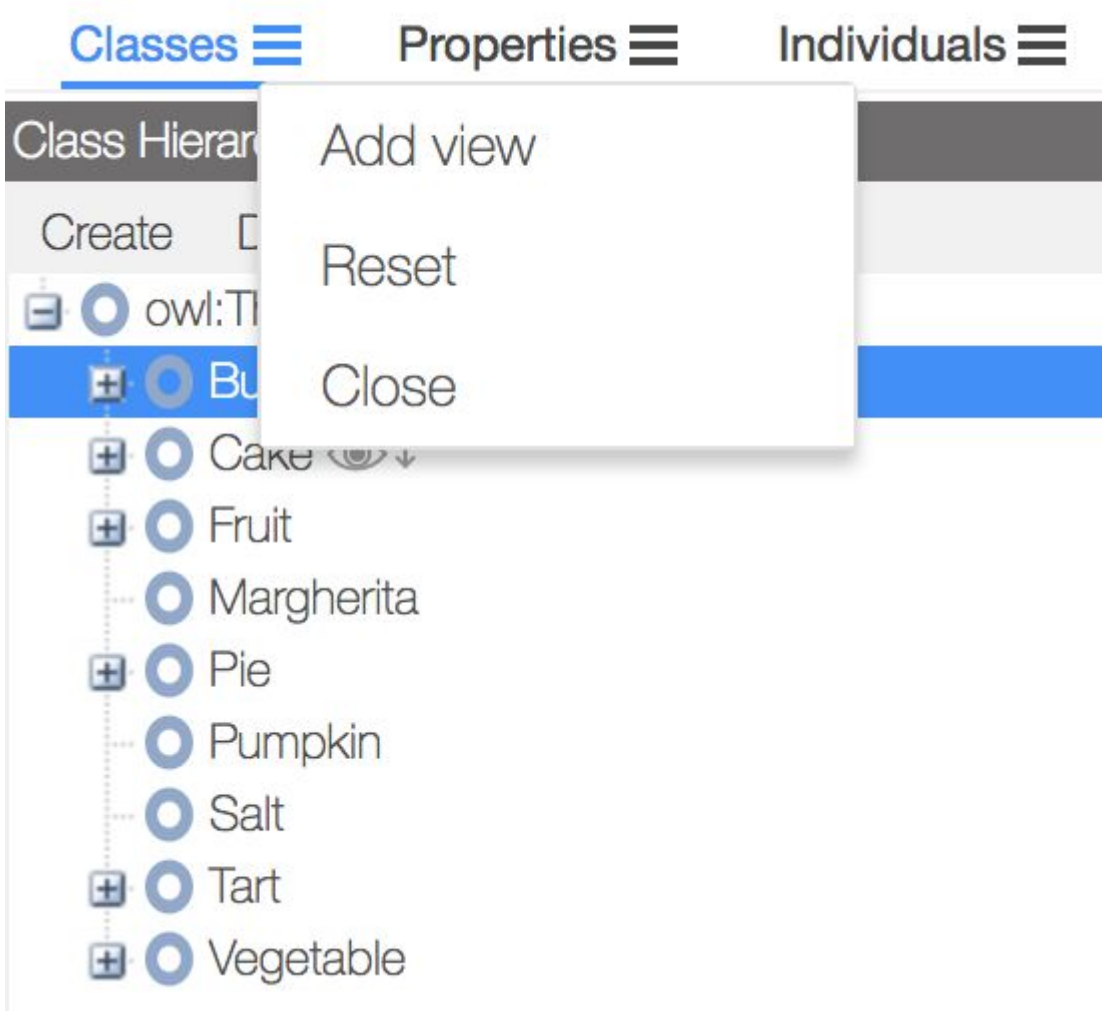
WebProtégé – simplified editing interface



http://webprotege.stanford.edu

Customizable Interface

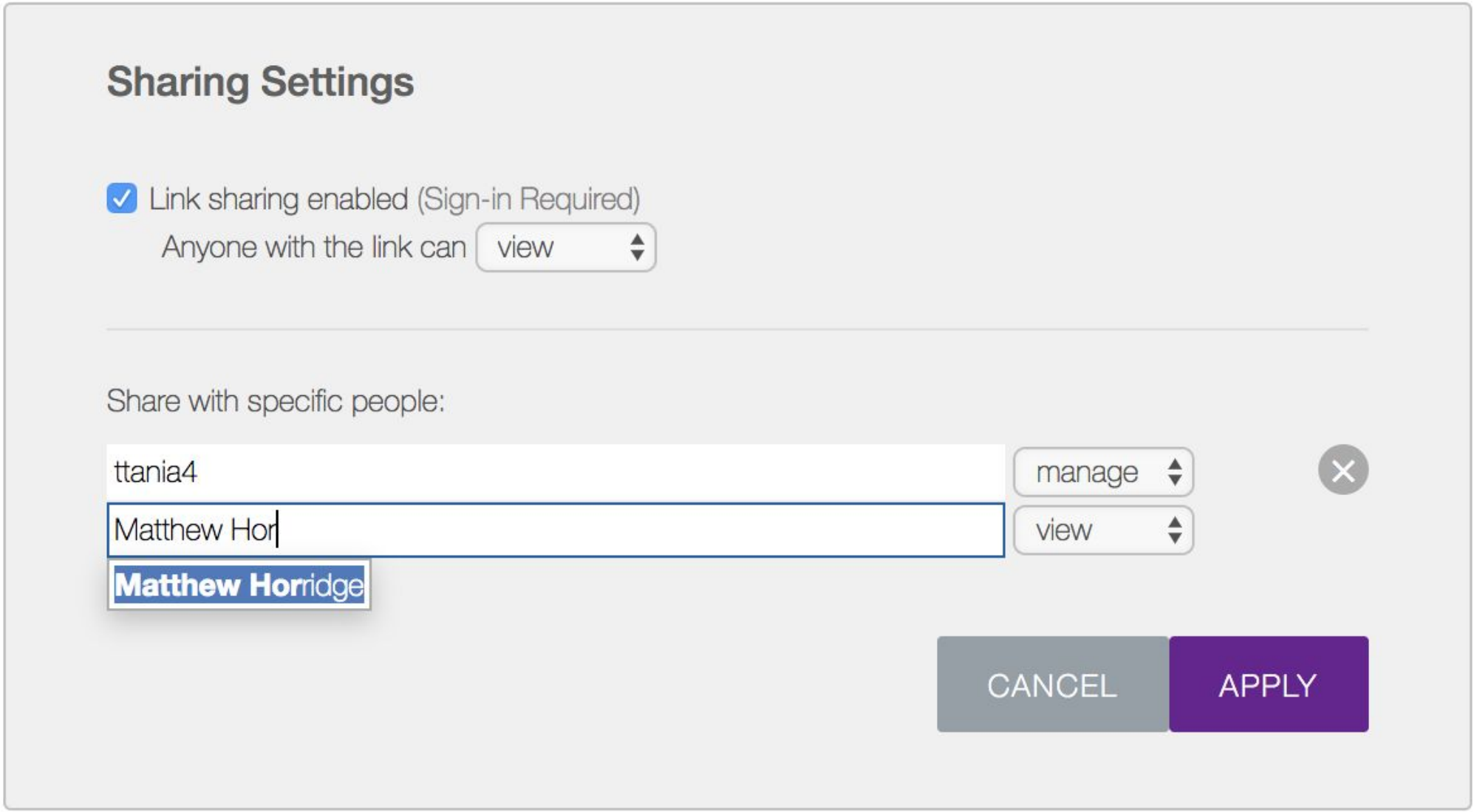
- Add new tab
- Add views to tabs
- Layout is saved and restored in the next session



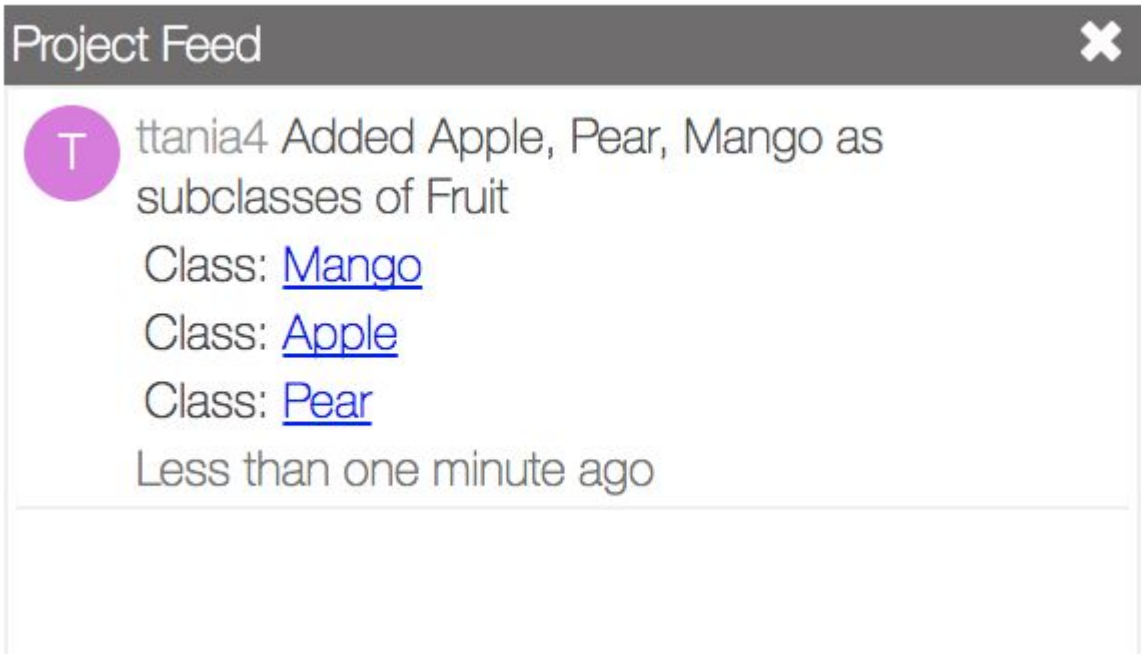
Collaboration Features in WebProtégé

- Sharing ontologies
- Simultaneous editing
- Change tracking
- Threaded discussions for ontology entities and changes (notes, discussions, proposals, reviews)
- Watching ontology entities and branches and notifications
- Upload and sharing of ontologies
- Download/revert any revision of the ontology
- Access policies
- User interface customization

Uploading and Sharing an ontology

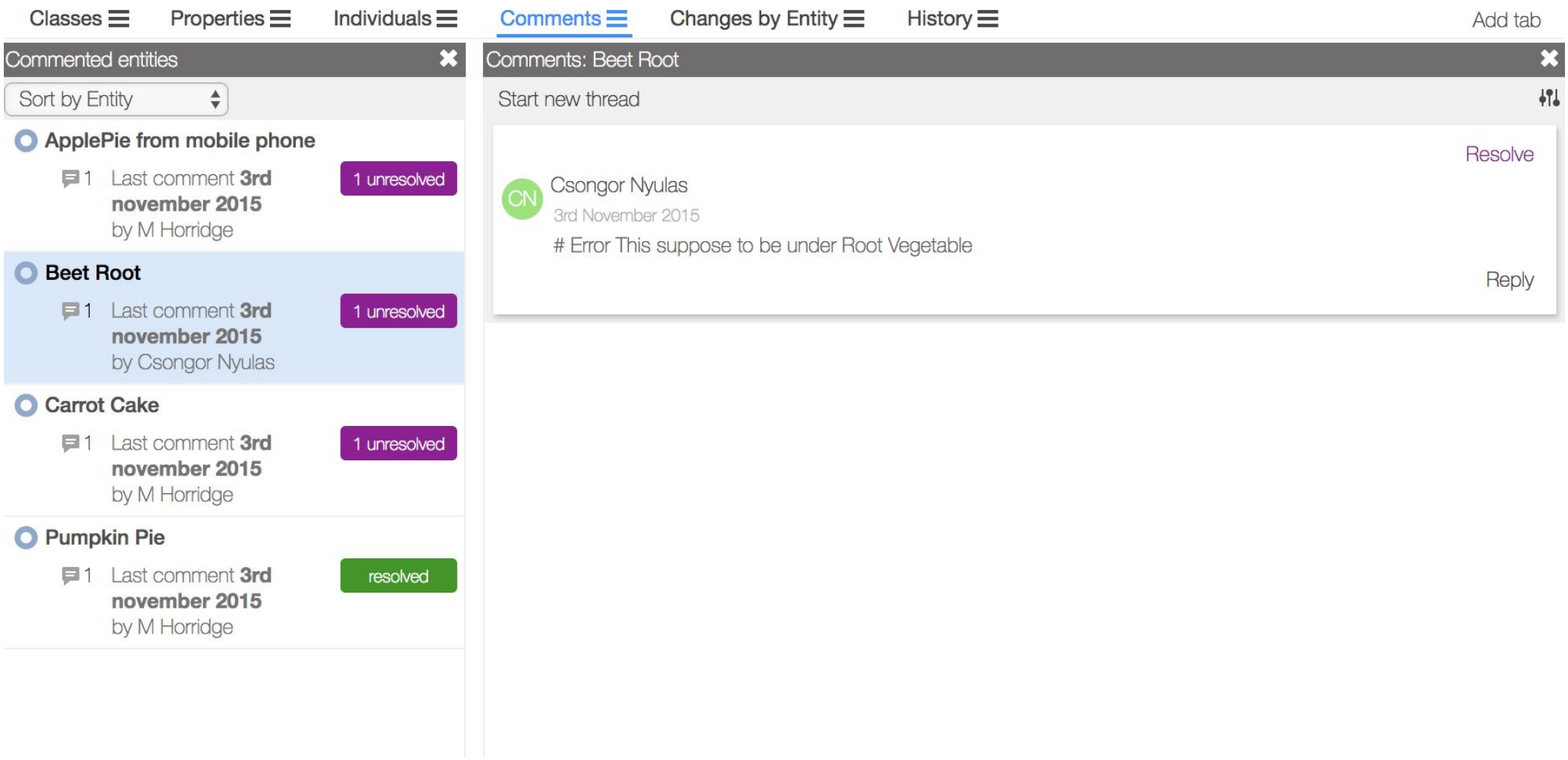


Project feed: see what other online users are changing in the ontology

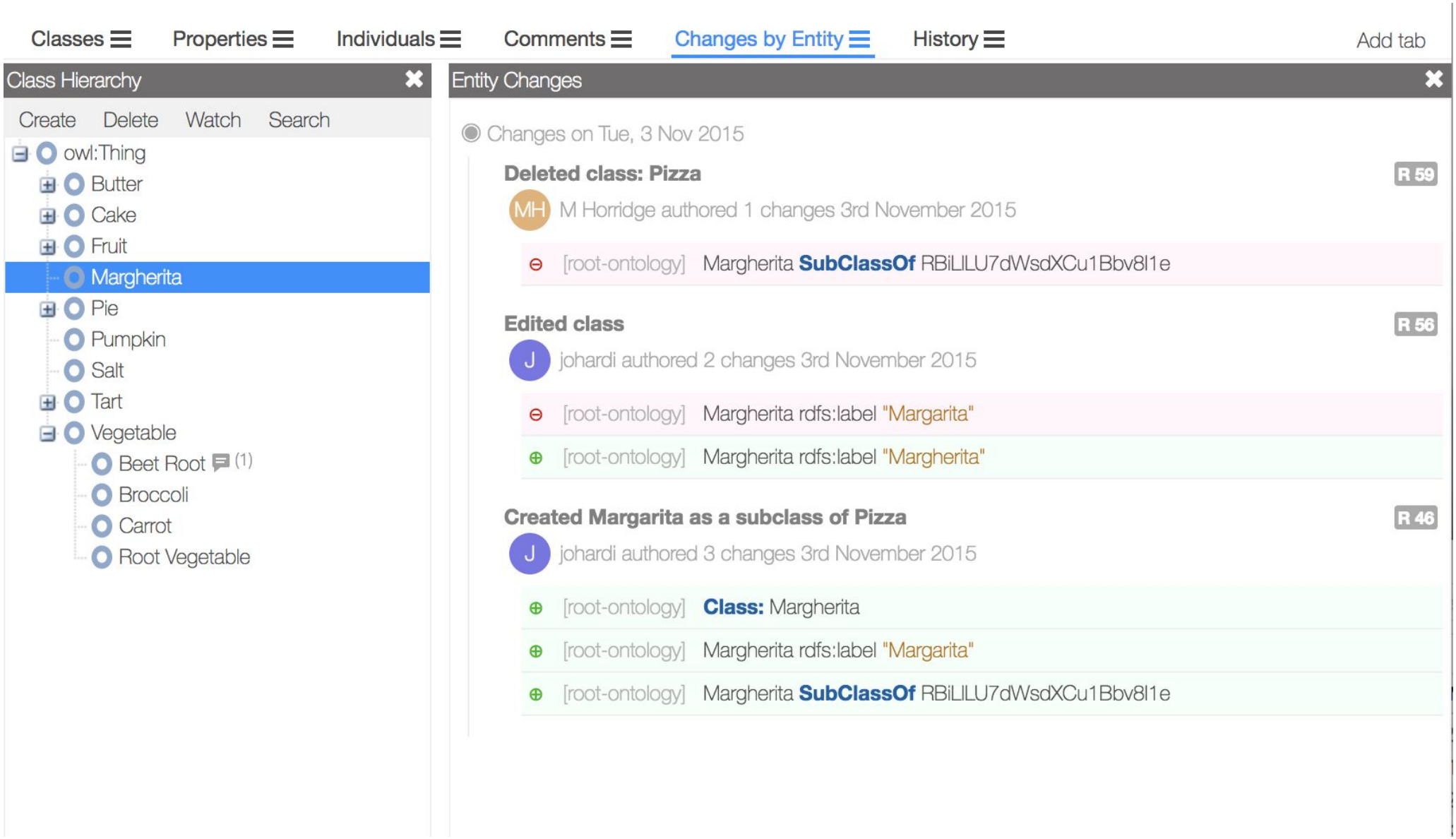


Feed updates as new changes are happening

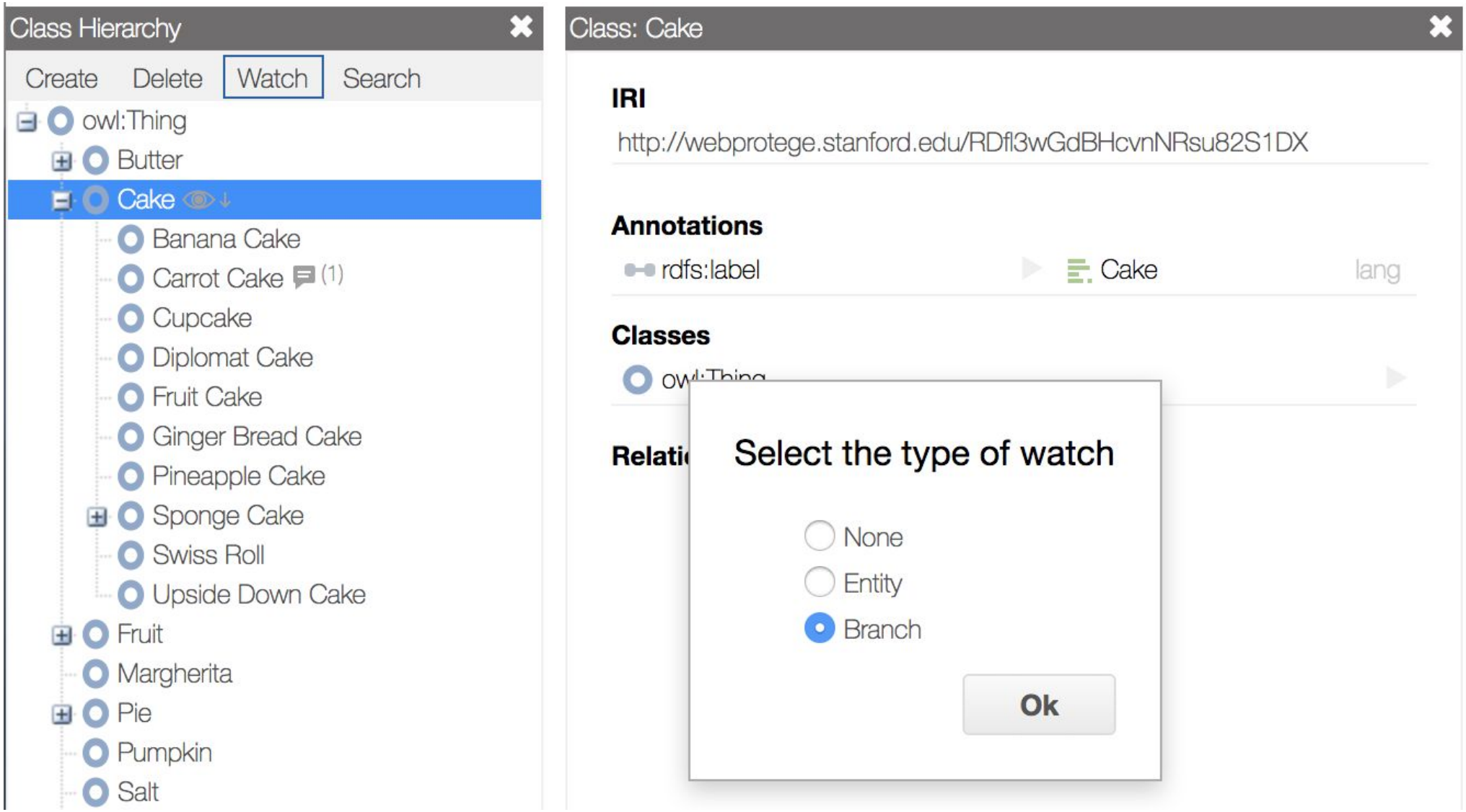
Notes and Discussions



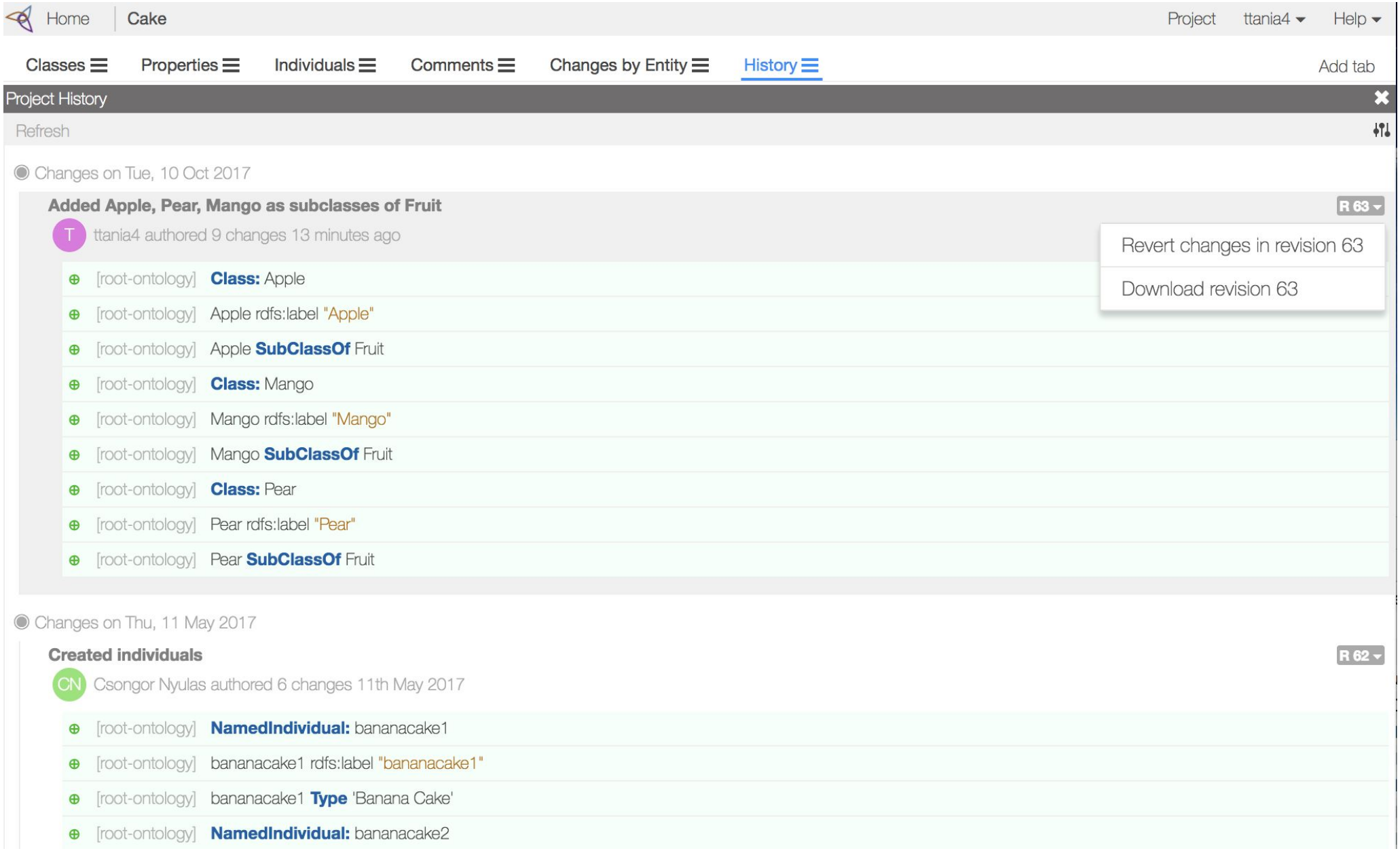
Change tracking



Watching entities and branches



History


Home
Cake
Project
ttania4
Help
Classes
Properties
Individuals
Comments
Changes by Entity
History
Add tab
Project History
Refresh
Changes on Tue, 10 Oct 2017
Added Apple, Pear, Mango as subclasses of Fruit
R 63
T
ttania4 authored 9 changes 13 minutes ago
Revert changes in revision 63
Download revision 63
[root-ontology] Class: Apple
[root-ontology] Apple rdfs:label "Apple"
[root-ontology] Apple SubClassOf Fruit
[root-ontology] Class: Mango
[root-ontology] Mango rdfs:label "Mango"
[root-ontology] Mango SubClassOf Fruit
[root-ontology] Class: Pear
[root-ontology] Pear rdfs:label "Pear"
[root-ontology] Pear SubClassOf Fruit
Changes on Thu, 11 May 2017
Created individuals
R 62
CN
Csongor Nyulas authored 6 changes 11th May 2017
[root-ontology] NamedIndividual: bananacake1
[root-ontology] bananacake1 rdfs:label "bananacake1"
[root-ontology] bananacake1 Type 'Banana Cake'
[root-ontology] NamedIndividual: bananacake2

Resources

- Online WebProtégé server: http://webprotege.stanford.edu
- Mailing list: http://protege.stanford.edu/support.php
- WebProtege on GitHub (sources and issue tracker): https://github.com/protegeproject/webprotege/
- WebProtégé documentation:http://protegewiki.stanford.edu/wiki/WebProtege
- WebProtégé simplified user interface: "Simplified OWL Ontology Editing for the Web: Is WebProtégé Enough?." Horridge, Matthew, et al., The Semantic Web–ISWC 2013. Springer Berlin Heidelberg, 2013. 200-215.
- WebProtégé paper: “WebProtégé: A Collaborative Ontology Editor and Knowledge Acquisition Tool for the Web”, Tania Tudorache, Csongor Nyulas, Natalya F. Noy, Mark A. Musen, Semantic Web Journal (SWJ) 4 (Number 1 / 2013), 89 - 99
- WebProtégé in use: “Will Semantic Web Technologies Work for the Development of ICD-11?”, T. Tudorache, S. M. Falconer, C. I. Nyulas, N. F. Noy, M. A. Musen. The 9th International Semantic Web Conference, ISWC 2010 (In-Use track), Shanghai, China, Springer. Published in 2010. http://bmir.stanford.edu/file_asset/index.php/1646/BMIR-2010-1427.pdf
- Other References: http://protegewiki.stanford.edu/wiki/WebProtege#References