Collaborative Ontology Development

Tania Tudorache

Protégé Short Course

Stanford University March 31, 2017

Collaborative Ontology Development

- <u>Collaboration</u>: 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!

2 parts: Editing and Publishing

Editing





- "Internal" development team
- Tools for editing ontologies
- Short development cycles

Publishing



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

What your team does

How others use your ontology

Outline

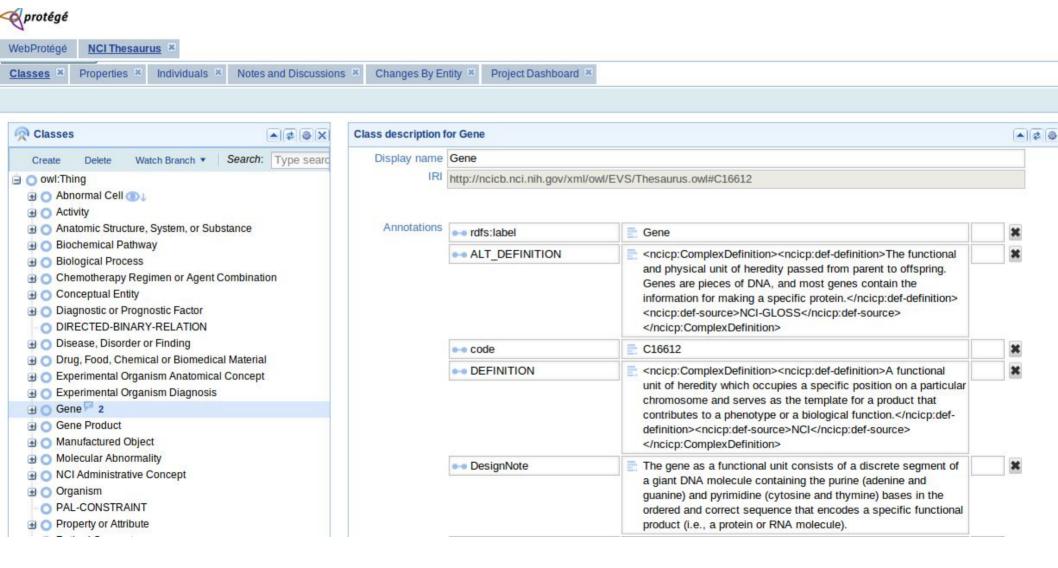
- Use cases of collaborative development
- Collaboration infrastructures and more use cases
- Collaboration support in WebProtégé

Outline

- Use cases of collaborative development
- Collaboration infrastructures and more use cases
- Collaboration support in WebProtégé

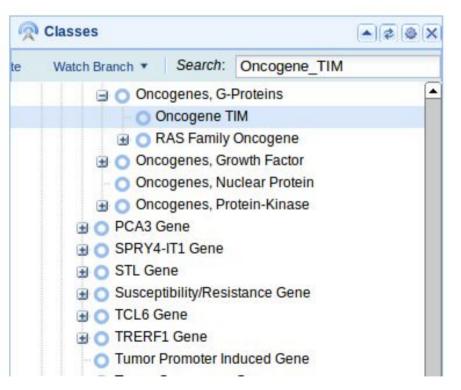
Use case: The NCI Thesaurus collaborative development process

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



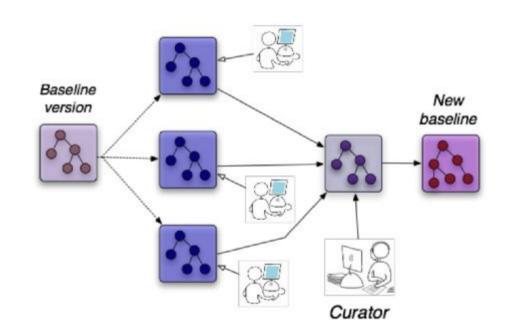
```
Equivalent To

('Oncogenes, G-Proteins'
and Allele_In_Chromosomal_Location some 7q33-q35
and Gene_Plays_Role_In_Process some 'Signal Transduction'
and Gene_Found_In_Organism some Human)

SubClass Of
SubClass Of Ancestor Class
Gene
Gene_Plays_Role_In_Process some Tumorigenesis
Oncogene
Gene_Plays_Role_In_Process some 'Signal Transduction'
'Cancer Gene'
Gene_Found_In_Organism some Human
Gene_Plays_Role_In_Process some Tumorigenesis
```

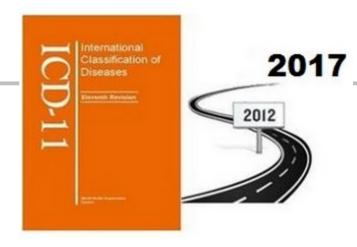
Use case: The NCI Thesaurus collaborative development process (cont.)

- 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



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

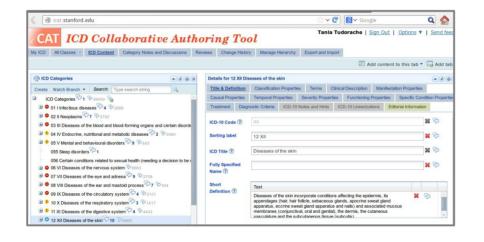




ICD-11 process today









- 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

Outline

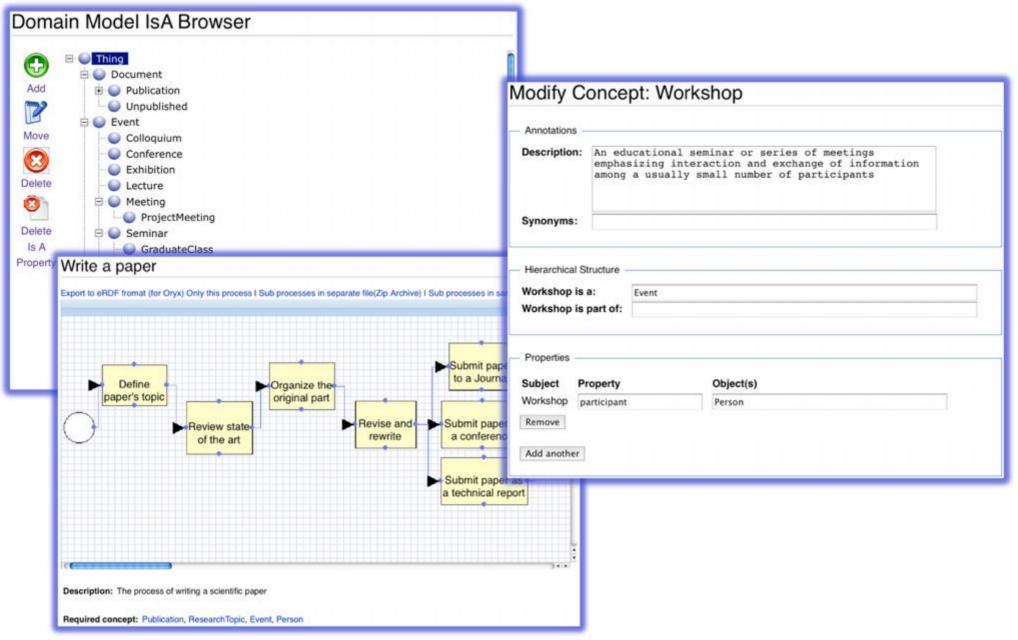
- Use cases of collaborative development
- Collaboration infrastructures and more use cases
- Collaboration support in WebProtégé

Collaboration Infrastructures: Wikis

- Wikis are well known; Wikipedia
- Semantic Wikis add semantic extensions to the wiking 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 I on page Apoptosis: [[OType:Category]] Individual(Apoptosis, type(Category)) Individual(Apoptosis value(CC-isa Biological_process)) 2 on page Apoptosis: [[CC-isa::Biological_process]] 3 on page HvSUT2: [[Realizes:: function = Sugar transporter activity; Individual(Realizes-0 type(Realizes)) process = Glucose transport]] Individual(Realizes-0 value(Realizes-subject HvSUT2)) Individual(Realizes-0 value(Realizes-process Glucose_transport)) SubClassOf(Realizes gfo:Relator)) 4 on page Realizes: [[has-argument:: name = function; type = OType:Function_category]] ObjectProperty(Realizes-function domain(Realizes) range(Function_category))

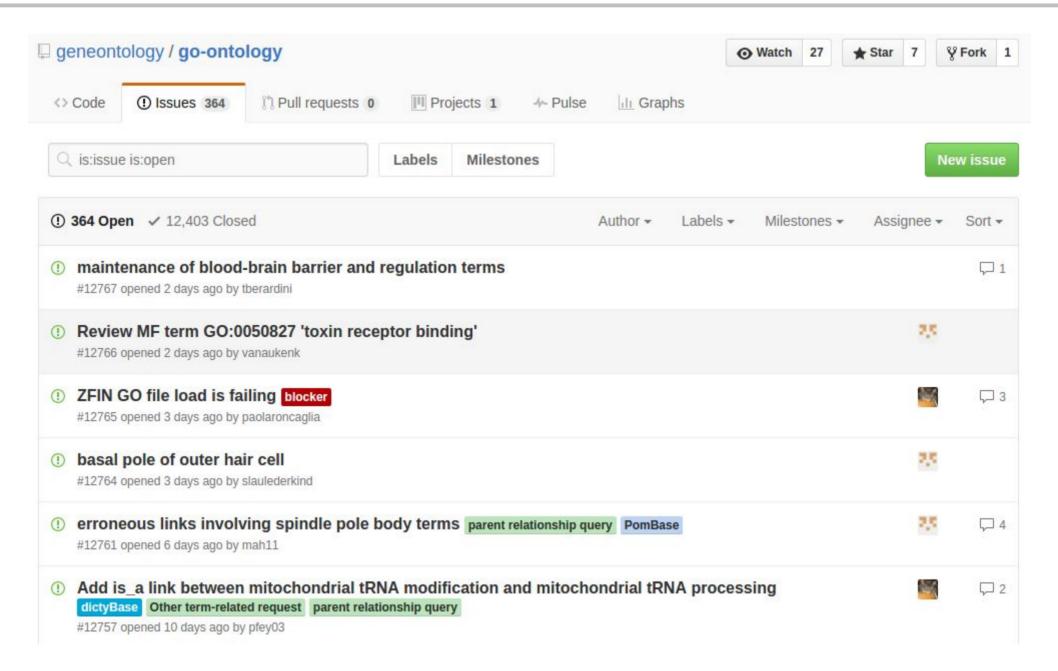


Source: Hoehndorf, Robert, et al. "BOWiki: an ontology-based wiki for annotation of data and integration of knowledge in biology." BMC bioinformatics 10.Suppl 5 (2009): S5.

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



Other collaboration processes

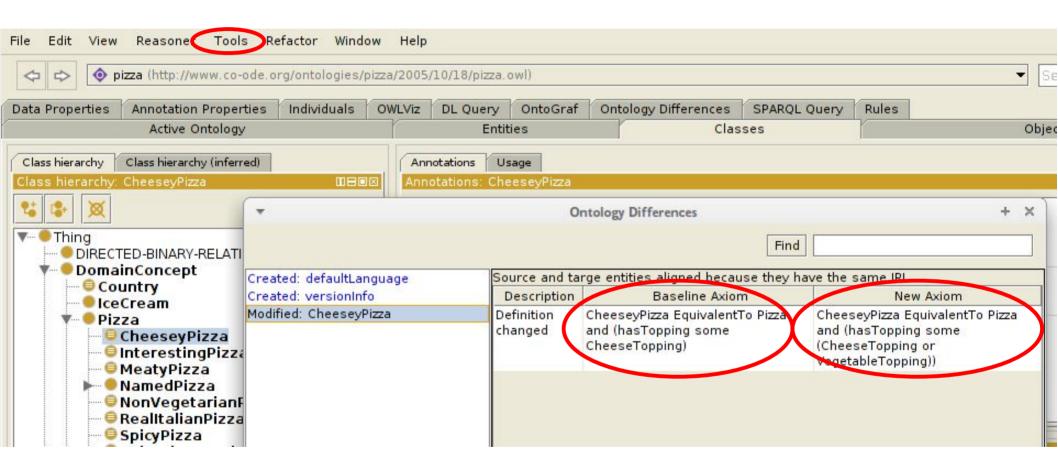
- Use source control repositories (e.g., SVN)
 - Text based mechanisms
 - Hard to merge local copies in the shared copy
- <u>Locking mechanisms</u> (lock parts of an ontology for editing)
- Use specialized (domain dependent) ontology repositories, e.g., <u>BioPortal</u>

Why using textual diffs doesn't work



Textual diff: 588 lines, including the "only" change:

Protégé diff support: Tools menu → Compare Ontologies



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

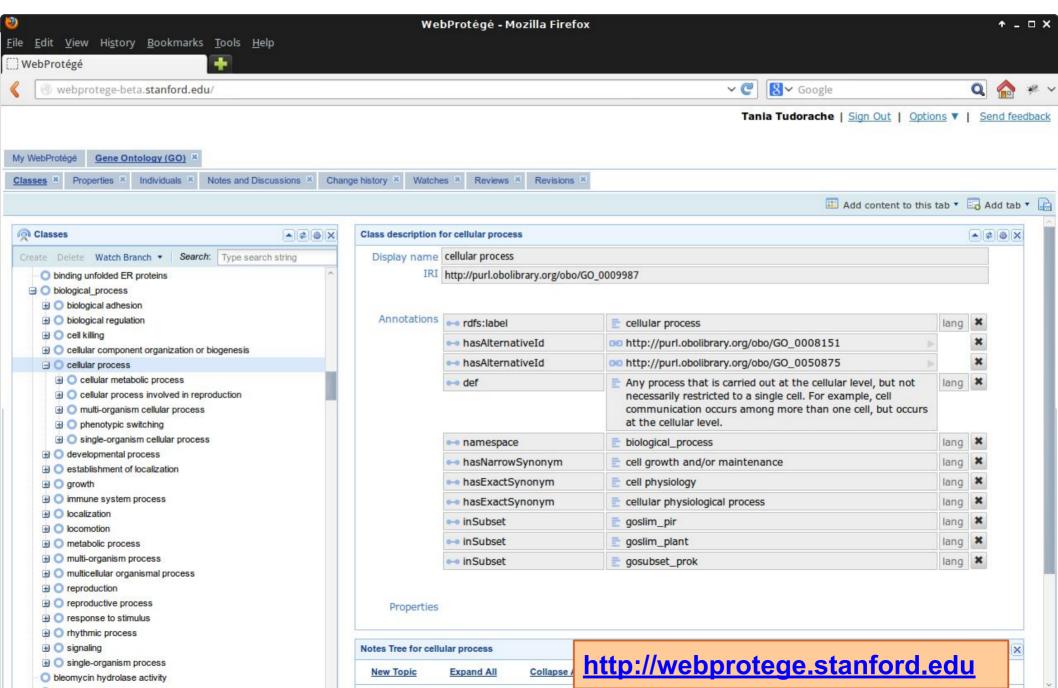
Outline

- Use cases of collaborative development
- Collaboration infrastructures and more use cases
- Collaboration support in WebProtégé

WebProtégé – quick overview

- Free, open source collaborative ontology development environment for the Web
- Google docs for ontologies; over 27,000 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)

WebProtégé – simplified editing interface



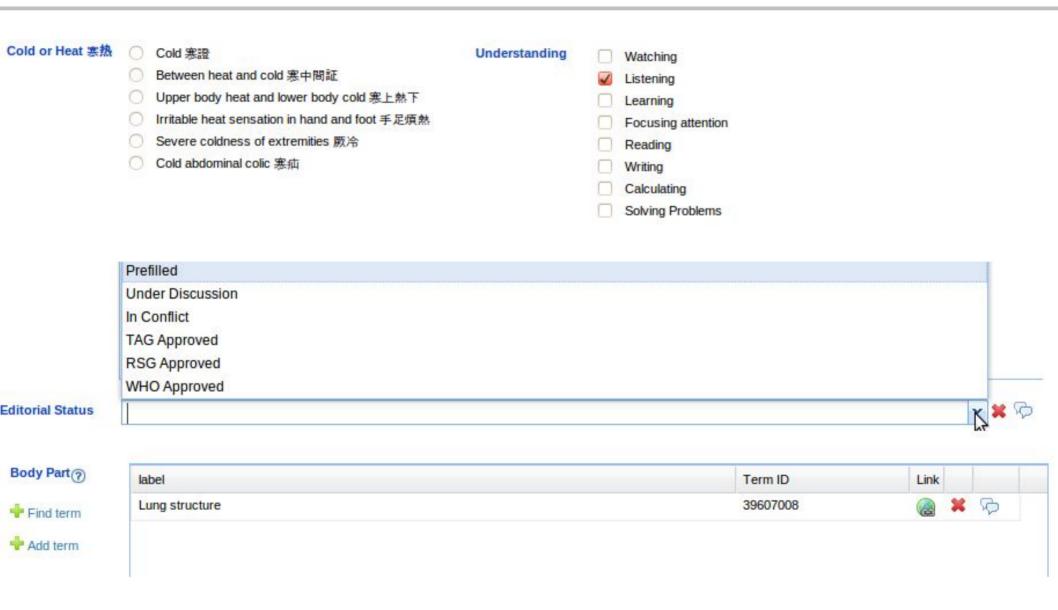
Custom entry forms for editing the ontology content

- Easy to create user interfaces for the <u>domain experts</u>
- Use <u>common entry forms</u>, but still keep the ontology "intelligence" behind it
- A form widget (e.g., text field) is linked to a property in the ontology
- Easy to create <u>custom forms</u> with different views for different users
- Hides complex ontology stuff

Form configuration in WebProtégé



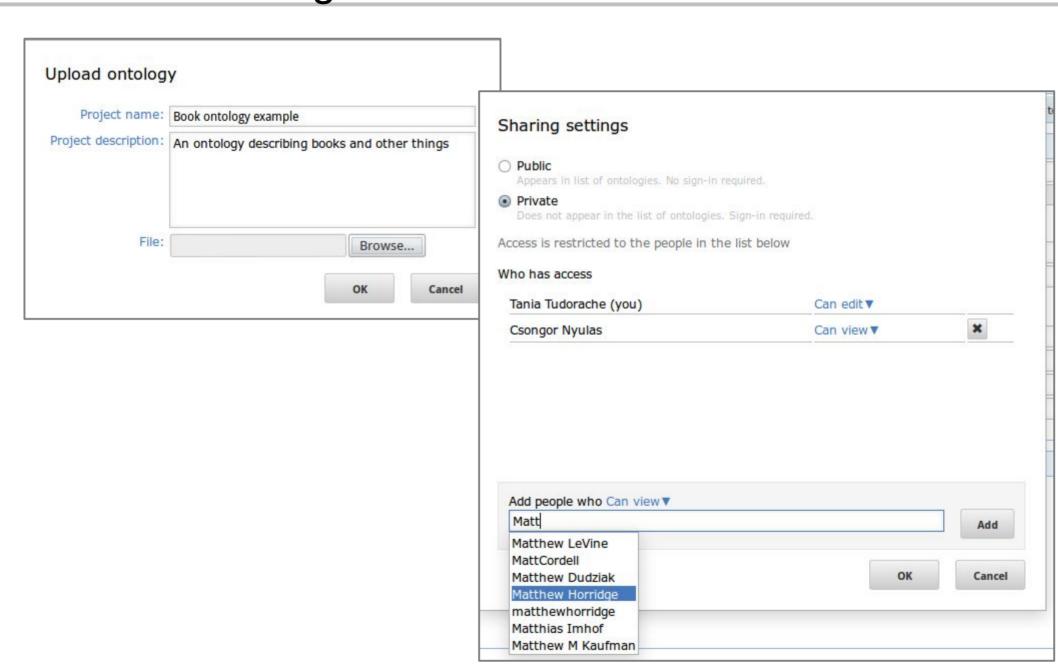
Examples of form-based editing



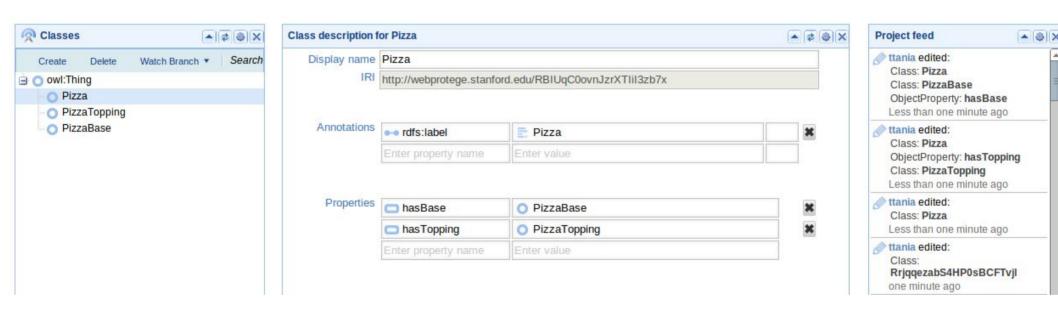
Collaboration Features in WebProtégé

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

Uploading and Sharing an ontology – similar to Google Docs

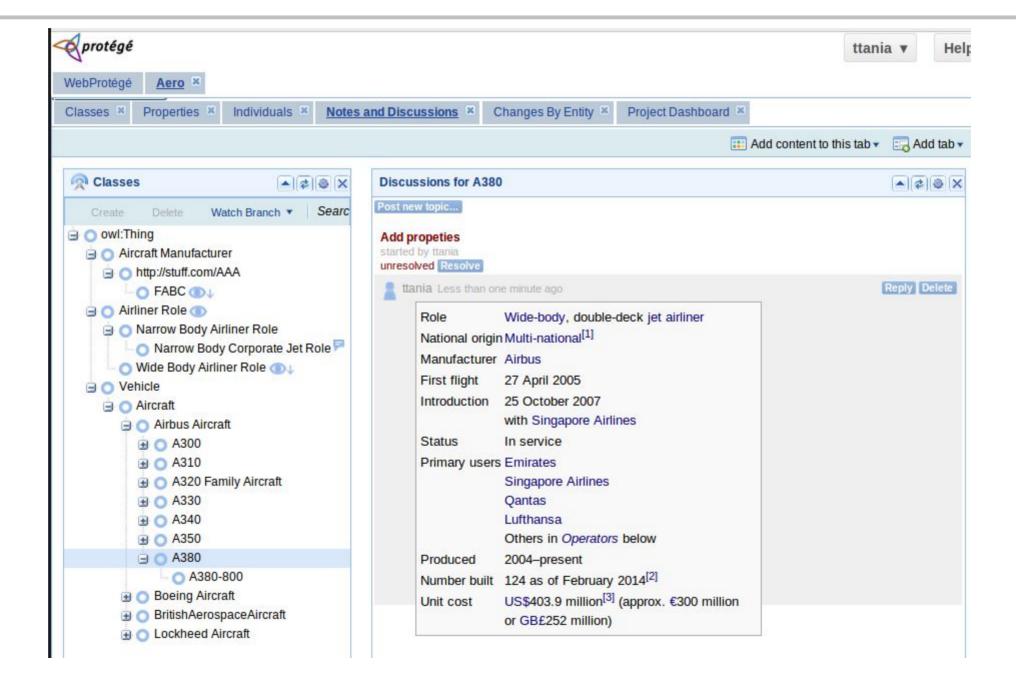


Project feed: see what other online users are changing in the ontology

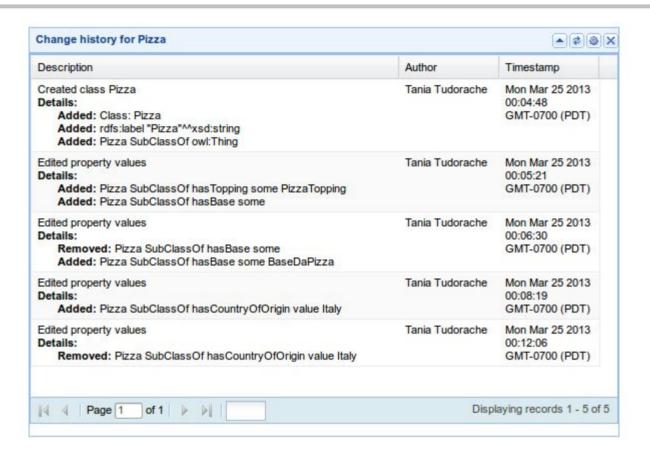


Feed updates as new changes are happening

Notes and Discussions

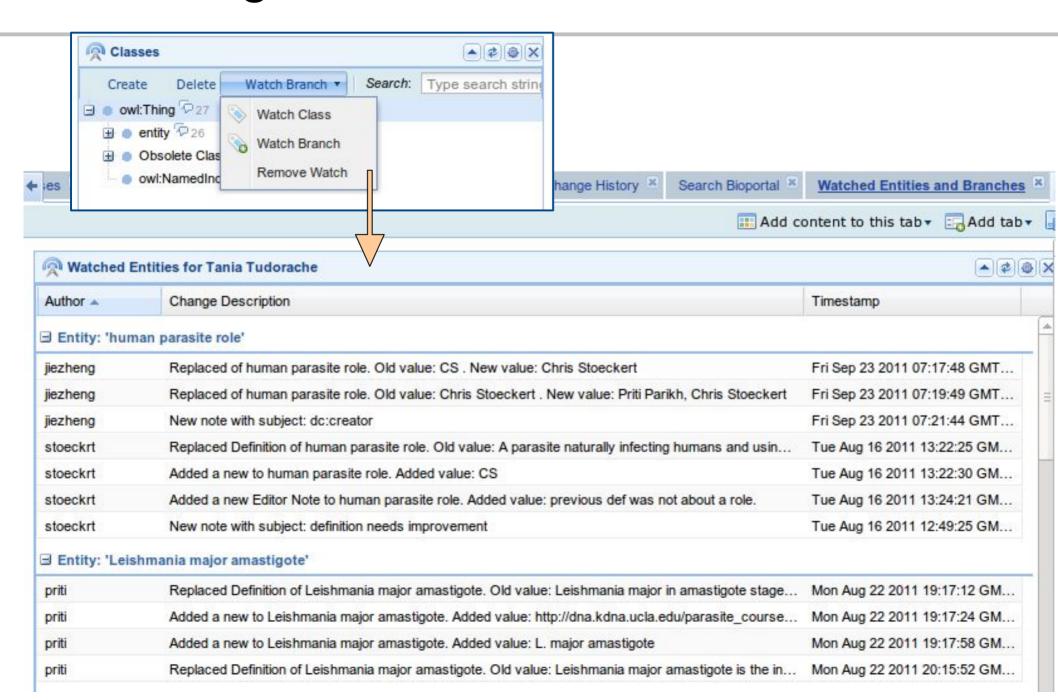


Change tracking in WebProtégé



- All changes are tracked on the server side
- User friendly description of changes:
 - Who did what, when and on what entity
- Changes are stored in a structured log that can be automatically processed

Watching entities and branches



Downloading revisions

Revisions				
Revision	Date	Number of change	Author	Download
116	26th March	2	M Horridge	0
115	26th March	2	M Horridge	0
114	10th March	1	M Horridge	0
113	17th March	2	M Horridge	0
112	17th March	3	M Horridge	0
111	7th March	1	guest	0
110	12th March 2013	(1)	M Horridge	0

Download any snapshot in time of the ontology

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
- <u>WebProtégé simplified user interface</u>: "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