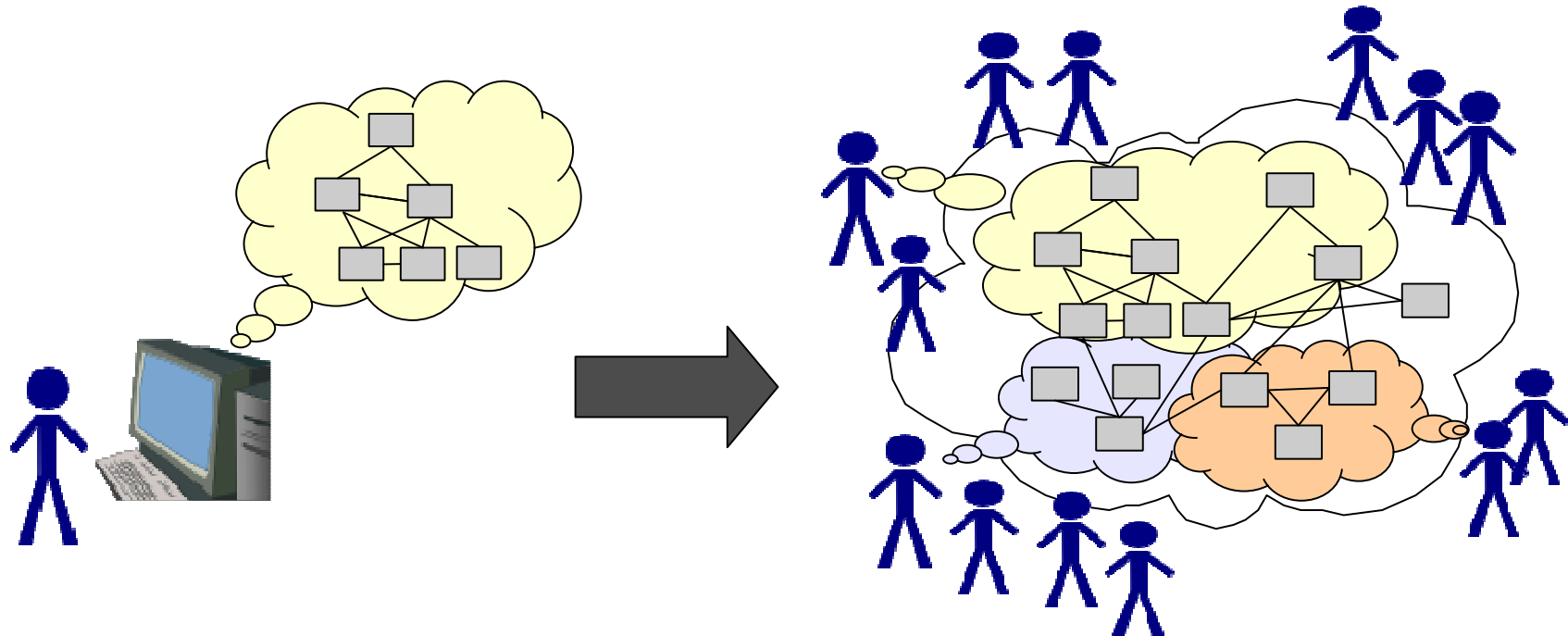


# Concurrent, Collaborative Ontology Building with Collaborative Protégé (CP)

Daniel Schober, IMBI



# Paradigm shift Collaborative Ontology Editing



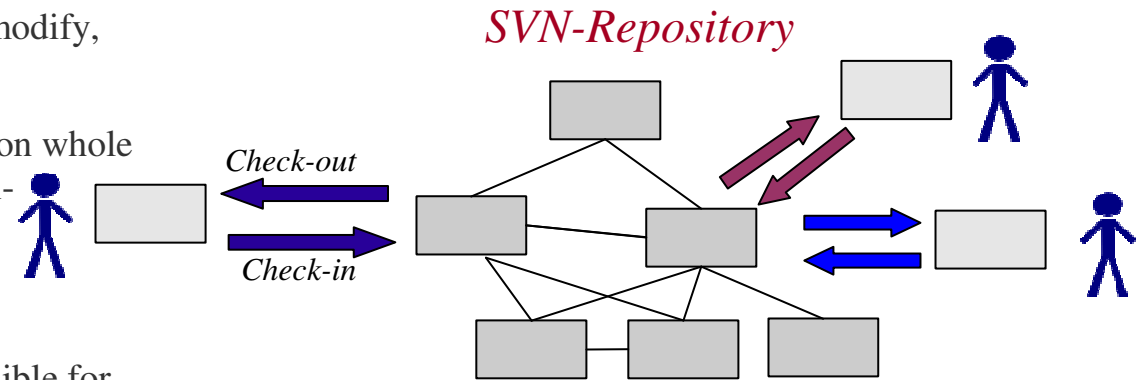
- Realize own idea
- Locally centralized
- Communication not an issue
- You know where to look and find

- Realize community consensus
- Locally distributed
- Collaboration & Communication editing, discussion & annotations
- 'Issue archeology' becomes an issue

# SVN vs. Concurrent Editing in CP

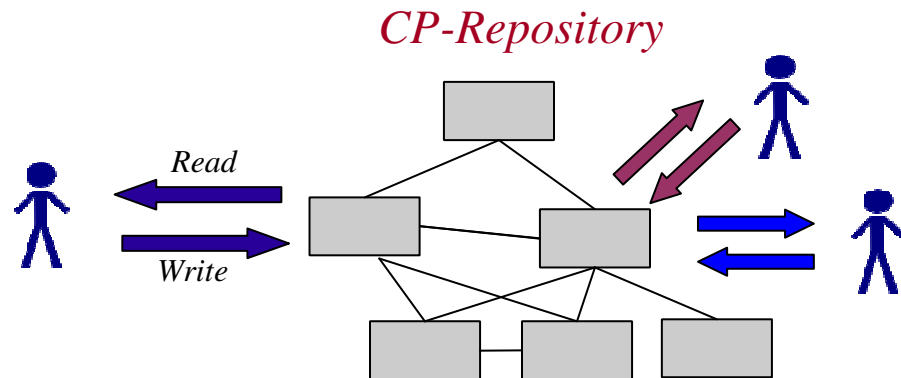
## SVN

- **Successive access** (update, lock, modify, commit local copy)
- **Complicated conflict resolution** on whole RA, even with logically non-conflicting changes
- High threshold for small changes
- Change and diff functions not feasible for owl
- **Annotations separate** from actual RU



## CP

- **Simultaneous access**
- Simple editing
- **Annotations associated to RU**



# CP Features

## Concurrent Editing

Parallel distributed Ontology Editing

## Metadata

Annotations on RUs (editorial and administrative metadata)

Annotations on Changes (annotations linked to delete actions and axiom edits)

Access rights management

## Searching

Search via user, annotation type & datestamp

## Communication

Discussion threads

Chat function (instant messaging)

Voting for decision support

# Changes Tab and Change Annotation

The screenshot displays the Protégé 3.4 beta interface. The main window is titled 'CLASS EDITOR for Bruker LC-NMR/MS platform (instance of owl:Class)'. The left sidebar shows a 'SUBCLASS EXPLORER' with a tree view of classes, including 'Bruker LC-NMR/MS platform'. The central pane shows the 'CLASS EDITOR' with a table of properties and values, and a section for 'Asserted Conditions'. The right pane is titled 'Annotations on Bruker LC-NMR/MS platform' and shows a list of annotations with columns for author and date. A 'Details' section below the list shows the author 'Kearon McNicol' and the creation date '12/16/2008 14:35:52 GMT'. A 'Description' section at the bottom right contains a photograph of a Bruker LC-NMR/MS platform. Several callout boxes highlight specific features: 'Collaborative Tabs' points to the 'Annotations' and 'Changes' tabs; 'Annotations on changes' points to the list of annotations; 'Threads' points to the 'Details' section; 'Has Annotations' points to the 'SUBCLASS EXPLORER'; and 'Hyperlinks & Pics' points to the photograph in the description.

Collaborative  
Tabs

Annotations on  
changes

Threads

Has  
Annotations

Hyperlinks &  
Pics

# Changes & Annotation Ontology (ChAO)

Classes Slots Forms Instances Queries

### CLASS BROWSER

For Project: ● annotation\_OBI

Class Hierarchy

- ▼ ● AnnotatableThing
  - ▼ ● Annotation
    - Advice (14)
    - Comment (89)
    - Example (1)
    - Explanation (1)
    - ▼ ● Proposal
      - SimpleProposal
      - ▼ ● VotingProposal
        - AgreeDisagreeVoteProposal (7)
        - FiveStarsVoteProposal
      - Question (6)
      - SeeAlso
      - ▶ ● Vote
    - ▼ ● Change
      - ▶ ● Class\_Change
      - Composite\_Change (629)
      - ▼ ● Created\_Change
        - Class\_Created (187)
        - Individual\_Created (21)
        - Property\_Created (7)
      - ▶ ● Deleted\_Change
      - ▼ ● Individual\_Change
        - DirectType\_Added
        - DirectType\_Removed
        - Individual\_Added (214)
        - Individual\_Created (21)
        - Individual\_Deleted (5)
        - Individual\_Removed (64)

### INSTANCE BROWSER

For Class: ● Advice

subject

- ◆ add metadata
- ◆ annotation notification and 'watch' functionality
- ◆ Chat
- ◆ definition
- ◆ Deprecation
- ◆ Function and Process - discuss
- ◆ identity theft
- ◆ Metadata
- ◆ metadata
- ◆ naming convention
- ◆ naming convention
- ◆ Re:
- ◆ redundancy
- ◆ Remove toy classes.

Types

- Advice

### INSTANCE EDITOR

For Instance: ◆ Remove toy classes. (Instance of ...)

Author: Daniel Schober

Created: ◆ 12/15/2008 15:59:51 GMT

Subject: Remove toy classes.

Add Internal Link To ?

Is this really a class we are likely to need? Model Scalpell instead.

#### Metrics

Summary

	System	Included	Direct	Total
Classes	15	0	63	78
Slots	34	0	30	64
Facets	10	0	0	10
Instances	0	0	5746	5746
Frames	59	0	5839	5898

Close

# CP Tool Evaluation Setting

- OntoGenesis network meeting at EBI (n=13, 2 days)
- Enrich OBI (OWL-DL)
- ‘Devices/Instruments’ branch
  - All members could contribute
  - Devices from
    - User domains
    - List provided by the Metabolomics Standard Initiative
- Feedback to CP developers

# CP Tool Evaluation Method

## Ad hoc additions under OBI (device and functions)

Duplication possible

*How are conflicts resolved ?*

## Controlled additions

Placement of devices from existing term list

*How is agreement (on subsets) coordinated ?*

## 'Agent Provocateur'

Secretly adding conflicting and false content

*How transparent are faults and nonsense edits to others ?*

## Controlled Communication

Restricted to specified channels during each editing session

Verbal shout-out, notes/discussion threads and chat

*How does CPs foster problem solving in communication ?*

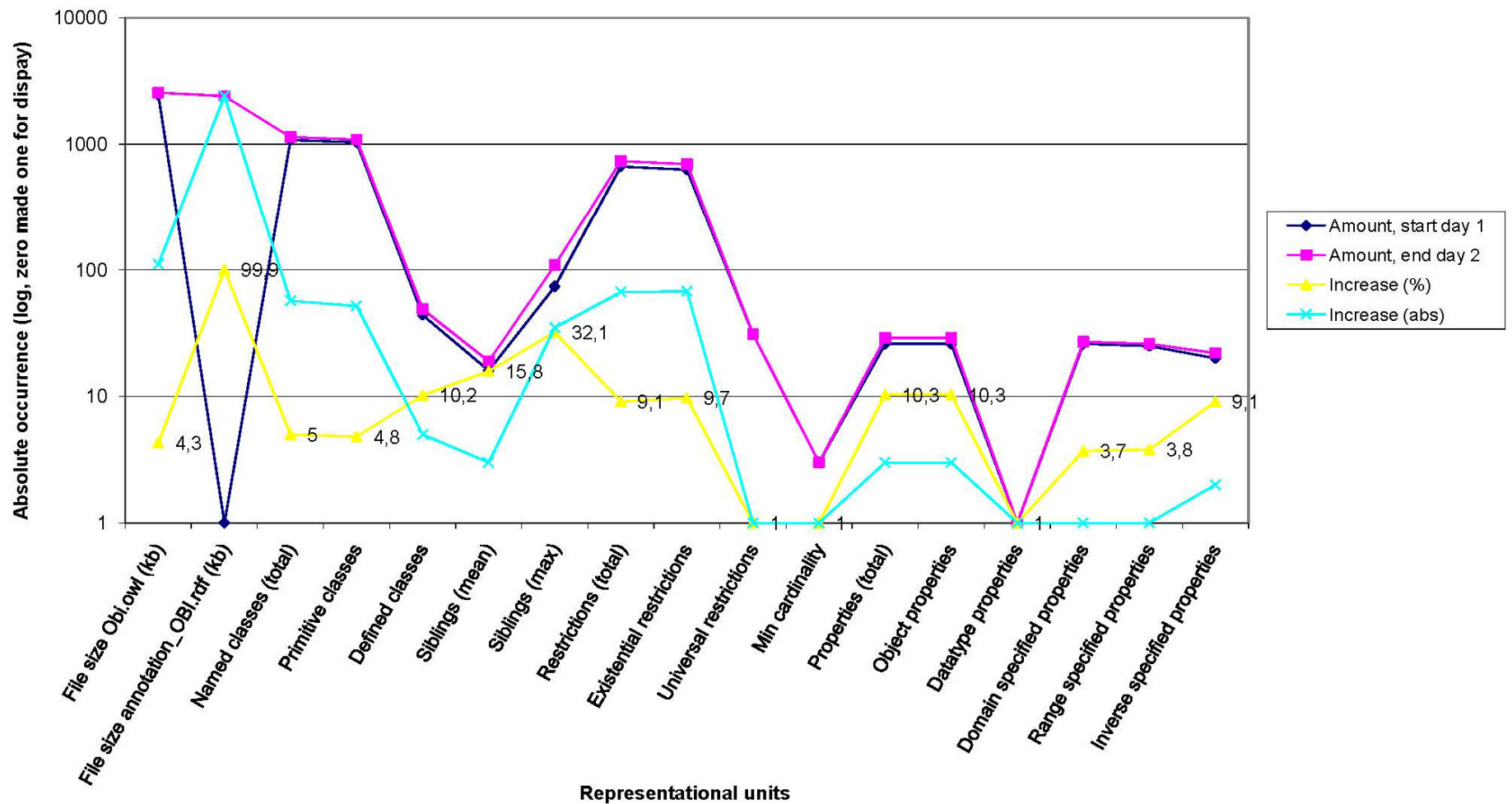


# CP Tool Evaluation Method

- Single group
  - Familiarization with CP & GUI
- Two groups
  - *Ad hoc* additions of own instruments
- Four groups
  - Add subsets of provided term list
  - Discuss, comments by other groups adding annotations
- Single group
  - Add more terms from list
  - Test communication channels
  - Deploy *Agent Provocateur*
- *Reasoning done ~ every half hour*

# Results: Increase of Ontology Size

Ontology Growth by single representational units

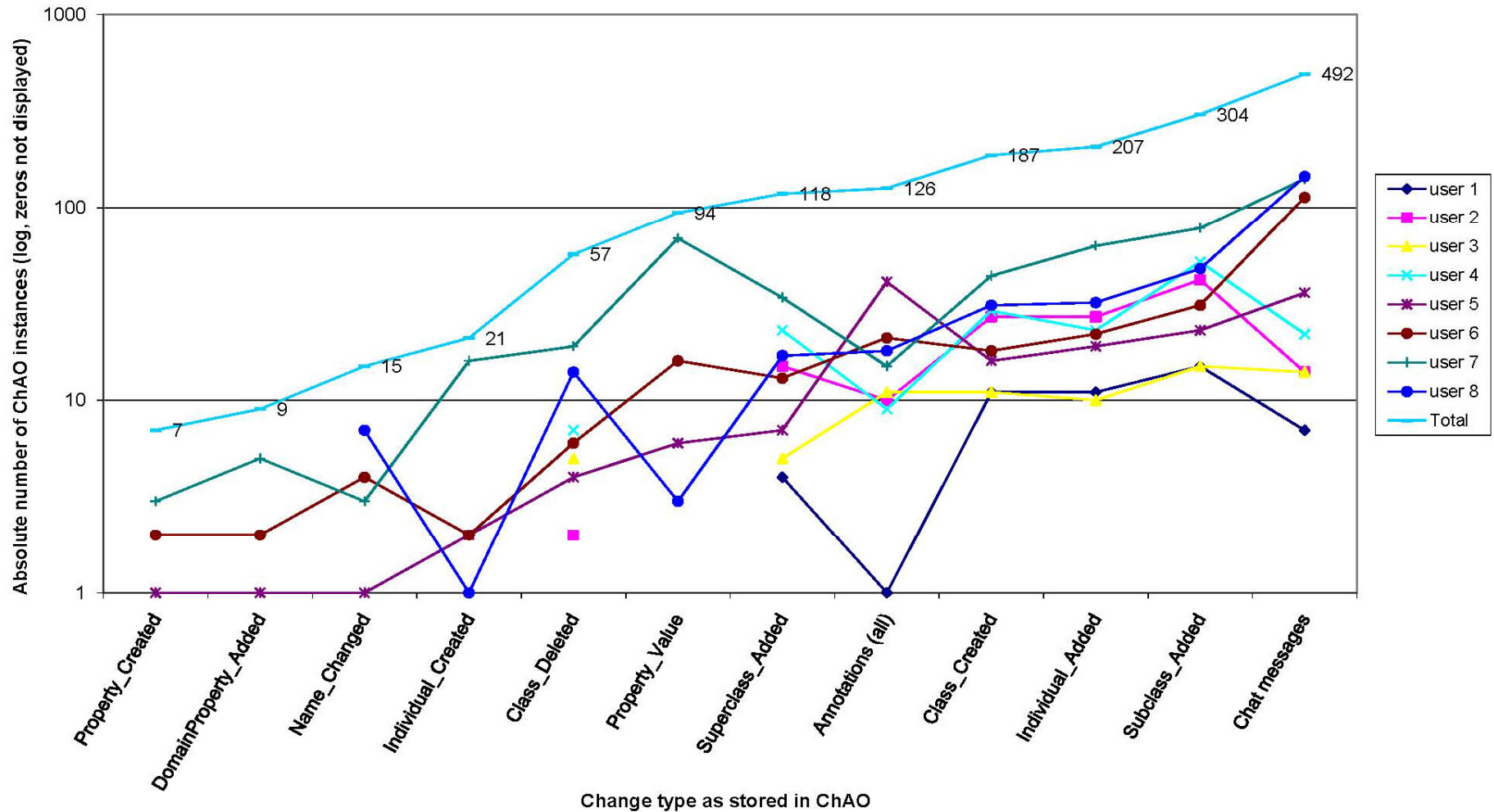


# Results: Increase of Ontology Size

- Quick setup, clear installation guide
- Metrics
  - 10.2% increase in defined classes, 4.8% in primitive classes
    - In OBI dev group primitive classes increase faster than defined classes
    - DL experienced Ontogenesis members
  - Only 3 object properties were created
    - 10.3% increase
    - Mainly re-use from OBI and RO
    - Relations used in 68 new existential restrictions (9.7% increase)
  - 4.3% increase in OBI.owl file size
    - 40 classes added, 13 refined/defined
  - 46,1 % increase in annotation\_OBI.rdf (per day)
    - 77 annotations (20 class annotations)
    - linear growth, no performance problems

# Results: Changes done per user

Actions on ontology done per user



# Results

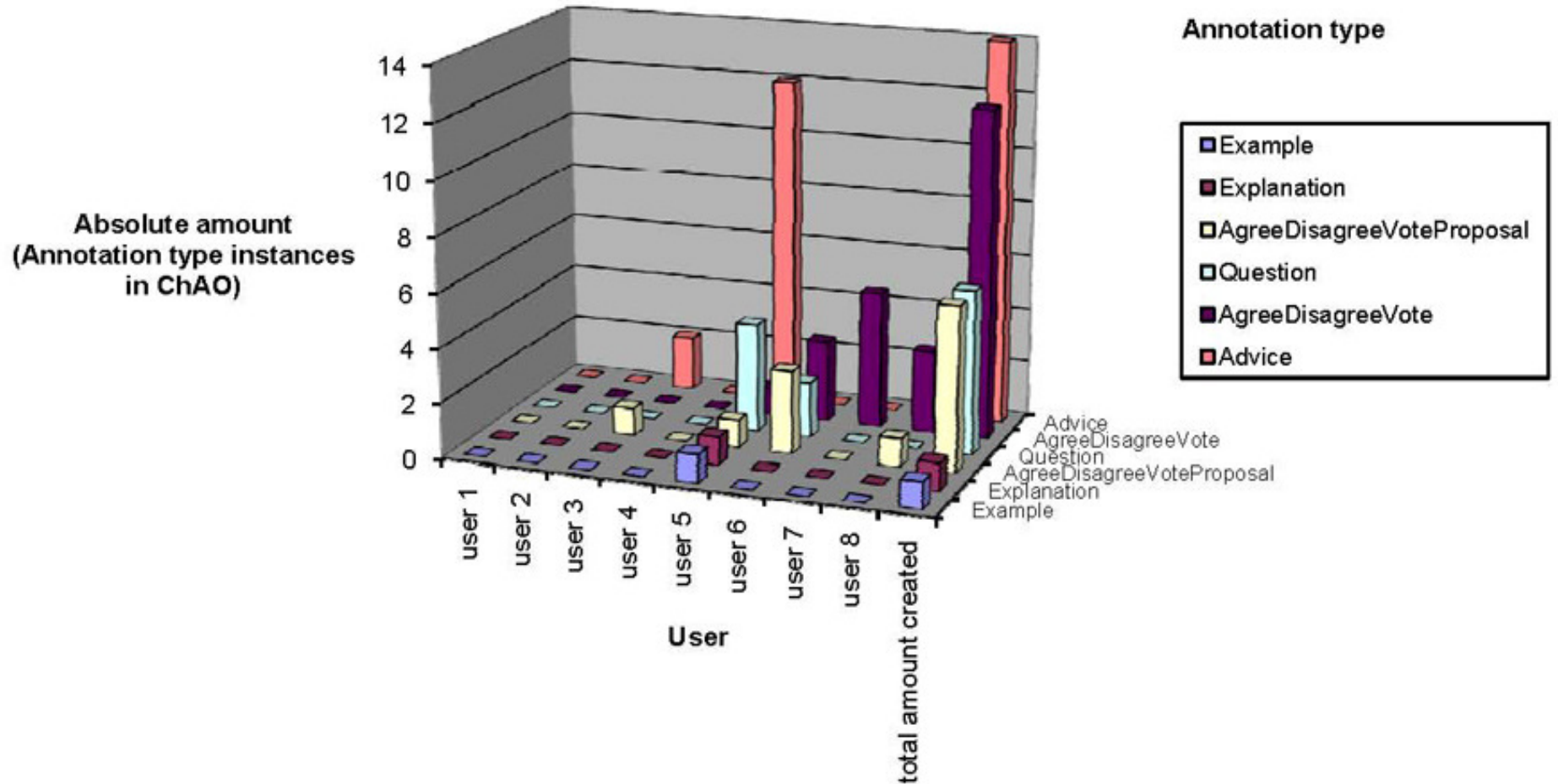
- Large differences in overall activity
  - Result of personality-structure, experience and confidence level
  - Quality of changes not yet evaluated
- Chat activity indicates overall editing activity
- Development of 'interest domains'
  - E.g. user 7 worked on relations, user 5 on annotations
- Development of 'user roles'
  - Users making comments don't necessarily implement them
  - Some users created tasks for others
    - e.g. 'add metadata', 'remove redundancy'
  - ChAO Patterns can be used to infer user roles
    - e.g. 'moderator', 'commenter', 'chatter', 'changer'
- Most classes edited by several editors (avrg. 2 per cls)
  - 13 changed classes (removed and added restrictions, changed superclasses, changed from primitive to defined)

# Results

- No power law distribution for comments per person
  - Most made ca. 10 comments, 'moderator' made 20
    - Role motivations could be Competition, Altruism, Narcissism, ...
- Discussion thread mean depth was 2.5, max depth was 5 responses
- Chat topics
  - What to work on next, modeling issues, new features & implementation
- Only 12 chats used internal hyperlinks
  - Increasing over time & CP familiarity
- Experimental helperclasses
  - '\_Kearon's collect devices by function classes', 'Frank's new meaning of function', 'asserted\_gibbon\_disco',
  - Only one user adhered to the OBI policy to indicate such play-classes with the underscore prefix (see first expl.)

# Usage of ChAO Annotation Types

Distribution of annotation types made by users



# Usage of ChAO Annotation Types

- Comment used due to 'default' setting
  - For 2 users comment was only annotation
  - Comment per class distribution followed power law
    - Few classes had 10-17 comments
    - Most classes had only 1-4 comments
- Advice and [AgreeDisagreeVotes](#) were used second abundantly
- There were a few [AgreeDisagreeVoteProposals](#) and Questions
- Example and Explanation were used seldomly
  - Distribution of annotations over the annotation types was highest among experienced users
- No annotations on changes
- No [SimpleProposal](#), [FiveStarProposal](#), [FiveStarVote](#) and seeAlso used



# Overall Performance

- GUI Updating
  - Expanding full class hierarchy in larger artefacts (took ca. 20 sec first time)
  - Opening a class with many direct subclasses will slow down clients and impair performance when done the first time
- Performance
  - Project loaded in 3 Min (on a 512MB P4 PC)
    - 2 Min for project, 1 for GUI
  - Increased by larger Heap Size
  - Removing concurrent projects from metaproject KB
  - Using DTB backend (dynamic loading)
    - Risk of data loss minimized (rollback)

# Desired Features

- RU and module locking mechanism
  - Can't prevent others from editing classes currently worked on
  - Parent class edits by unaware users can contradict definitions under construction
    - Highlight edited areas e.g. by user colour scheme
- Roll back function
  - Aid in conflict resolution
  - Undoing of deleted classes
    - Properties were found to be sub-properties of deprecated properties
    - Global change list to allow to see changes and annotations on deleted entities
- Subscription and Notification
  - Notification of changes
    - help to stay up to date and proceed faster in conflict resolution
    - E.g. a 'change view' on selected watch list items (see ICBO paper on how to implement)
    - Notification on duplicate RU labels

# Desired Features

- Planning
  - Change ontology based on vote outcomes
    - Increase development time
    - Implement using ChAO information and formalized voting outcomes
  - Issue tracker
    - A scratch pad or todo list that can be worked through and 'checked', e.g. indicating a proposed plan & what has been already realized at a certain time point
  - Connection with e.g. SF term trackers
- Chats
  - 'Retreat room'
  - Emoticons could increase transmittance of pragmatic communication aspects
  - Filter function on user names or particular ontology fragments

# Further Observations

- Annotation on RUs
  - For minor annotations providing annotation type, subject heading and value is overkill
  - Change track in ChAO KB is sometimes overly granular (overkill)
    - Users like high level abstractions, e.g Class X moved under Class C
- Communication
  - Threads and notes were misused for chats and *vice versa*
    - The latter due to the chats' instant visibility
  - Difficult to find cut off, when to move from chat to RU note or thread
  - Consequences of using wrong annotation channel
    - A user advised the group not to use an obsolete object property in a tread rather than in a note on that object property itself
      - people used the obsolete property

# Overall CP Benefits

- **Rich set of annotation properties**
  - Comment, advice, explanation, question, example, vote, ...
  - **Change-annotations** ease deprecation and versioning
- **Changes immediately visible to all clients**
  - **Use during telecons directly**
    - rather than redundantly keeping notes & implement them later
- **Centralized access to otherwise distributed contextual metadata**
  - *Issue-archaeology* much easier
- **Flexibility of ChAO metadata scheme**
  - **Annotation types can be expanded**, searched and filtered
  - Granular annotation types to suit own needs and evaluation approaches
  - **Exploit for statistics**
  - Use for proof and trust
  - Use for all non-DL add-ons, e.g. epistemology
- **Personalized views based on**
  - User roles and tasks
  - User level of expertise
  - User trust network

# Conclusions

- Rich CHAO metadata set provides audit trail of edits and decision making
- Tool in advanced stage with good performance
- Copes with complicated setups
  - Flexible enough to allow for corresponding adjustments
- Desired features
  - **Conflict resolution**, e.g. 'undo/redo' is needed, as well as transaction management
  - **Notifications** on changes to notes and threads
  - Chats to specific RUs and for specific groups
  - Feedback valuable for CP version of P4 (due next year)

# Resources and Acknowledgements

## Resources

- Ontogenesis Website
  - <http://ontogenesis.ontonet.org/moin/NetworkMeeting7>
- CP Demo
  - <http://protege.stanford.edu/doc/collab-protege>
- CP Documentation
  - [http://protege.stanford.edu/doc/collab-protege/doc/collabProtege\\_demo.pdf](http://protege.stanford.edu/doc/collab-protege/doc/collabProtege_demo.pdf)

## Acknowledgements

- Tania Tudorache, Timothy Redmond, Natasha Noy
- Robert Stevens, James Malone, Susanna Sansone, Stefan Schulz
- OBI Consortium
- DebugIT EU 7th FP ICT-2007.5.2-217139
- EBI NET-project, [www.ebi.ac.uk/net-projects](http://www.ebi.ac.uk/net-projects)

**Thank you for your attention !**





- CPs 2 parts
  - Multiuser Protégé for simultaneous editing (Client-server)
  - Collaborative Protégé for collaboration support
- Works with all backends of Protégé (OWL, frames, RDF)
- Installation
  - Install default P3.X, start server, Configure Project access rights & add users in metaproject on local host. Give port and host name on client-side to connect
- Complete change history
  - Bob created 3 classes yesterday, but today he was kind of lazy. He also tried to delete a class created by John but was not allowed to
- Users may comment on changes; for example on a class rename operation or on a change of a domain property
- WebProtégé is an open source lightweight ontology editor for the Web
- It is a Collaborative Protégé client

# CP Setup

Nature Precedings : doi:10.1038/npre.2009.3516.1 : Posted 29 Jul 2009

**Configuration of the Collaborative Protégé Server: The Metaproject**

metaproject Protégé 3.4 beta (file:/home/tania/Protege\_3.3\_beta/examples/server/metaproject.pprj, Protégé Files)

**CLASS BROWSER** For Project: metaproject

- Class Hierarchy
  - THING
  - SYSTEM-CLASS
  - Project (5)
  - User (142)
  - Group (5)
  - Operation (2)
  - GroupOperation (1)

**INSTANCE BROWSER** For Class: Project

- name
  - CKC2007
  - CKC2007 Annotations (DO NOT USE THIS)
  - Collaborative Newspaper
  - Collaborative Newspaper Annotations (DO NOT USE THIS)
  - Collaborative Pizza
  - Collaborative Pizza Annotations (DO NOT USE THIS)

**INSTANCE EDITOR** For Instance: Collaborative Pizza (instance of Project, internal name is m...)

Name: Collaborative Pizza Owner: tania

Location: examples/collaborativePizza/collaborativePizza\_localhost.owl.pprj

Description: The collaborative pizza project

**Allowed Group/Operation**

allowed Group	allowed Operation
World	Read, Write

**Annotation Project**

- Collaborative Pizza Annotations (DO NOT USE THIS)

Callouts:

- Projects available on the server (points to Class Browser)
- Policies on operations/p project (points to Allowed Group/Operation table)
- Associated annotation project (points to Annotation Project section)
- Project location on server (points to Location field)



# CP Setup

CollabProtegeTutorial.pdf - Adobe Acrobat Pro

File menu -> Open ... -> Server

The server host name (might be different during demo)

Create new user

Open Project

File: Host Machine Name  
smi-protege.stanford.edu:5200

User Name  
Guest

URL: Password  
\*\*\*\*\*

Server: Newuser

OK Cancel

Web-start and applet demo: <http://smi-protege.stanford.edu/collab-protege/>

23

Nature Precedings : doi:10.1038/npre.2009.3516.1 : Posted 29 Jul 2009

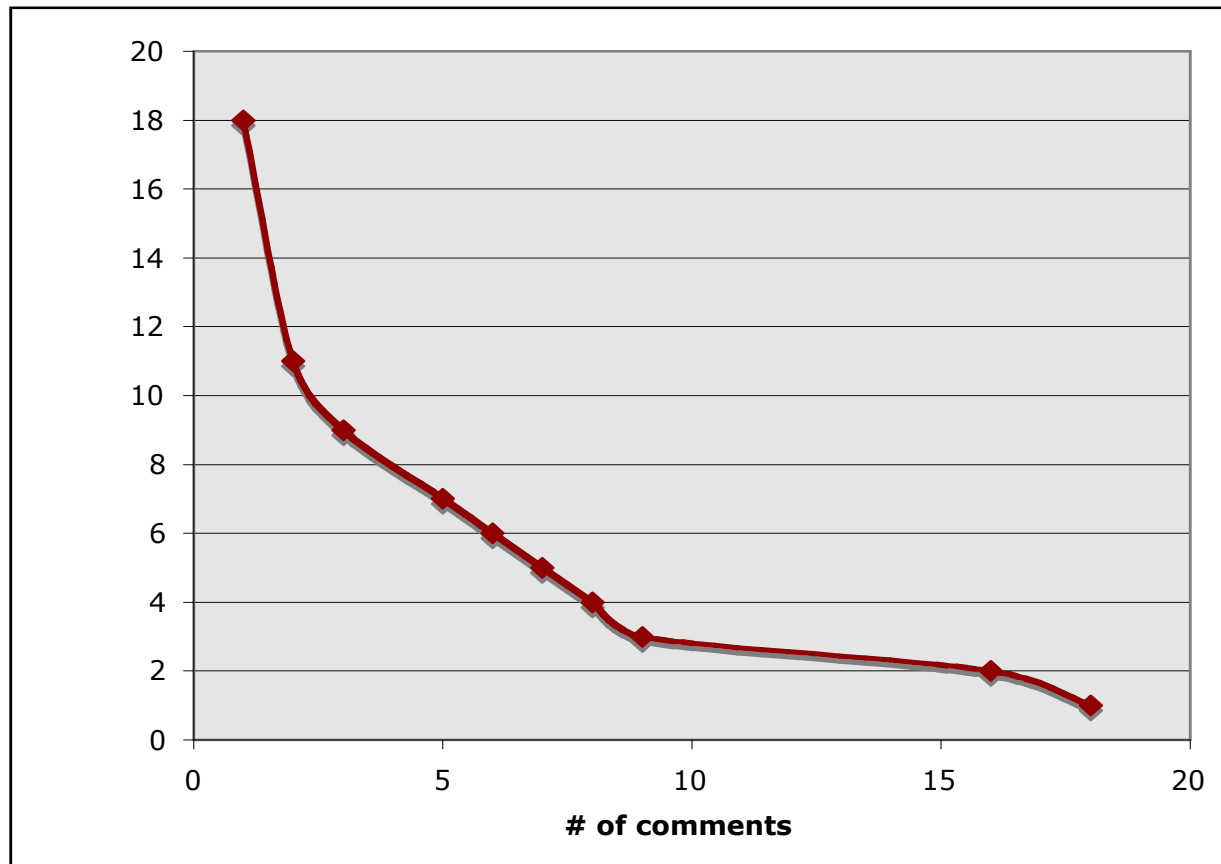
# Changes & Annotation Ontology (ChAO)

The screenshot displays the ChAO software interface, which is divided into several main sections:

- CLASS BROWSER:** Located on the left, it shows a hierarchical tree of classes. The root is 'AnnotatableThing', which includes 'Annotation' (with sub-classes like Advice, Comment, Example, Explanation, Proposal, and VotingProposal) and 'Change' (with sub-classes like Class\_Change, Created\_Change, Deleted\_Change, and Individual\_Change). The 'Composite\_Change' class is currently selected.
- INSTANCE BROWSER:** Located in the top middle, it shows a list of instances for the selected class. The instances listed include '\_Kearon's collect devices' and multiple 'Add Equivalent Class -- A' entries.
- INSTANCE EDITOR:** Located on the right, it provides a detailed view of the selected instance. It includes fields for 'Context', 'ApplyTo', 'AssociatedAnnotations', 'SubChanges', and 'Action'. The 'SubChanges' section is expanded, showing a list of changes such as 'Superclass Added: Chromatography, Nuclear mag...', 'Subclass Added: http://purl.obofoundry.org/obo/C...', 'Superclass Added: '(has\_part, has\_part)' some 'N...', and 'Subclass Added: @5\_1e8f0807\_51f7\_4fd0\_9a82...'. The 'Action' field is set to 'Composite\_Change'.

# Power law distribution in comments per class

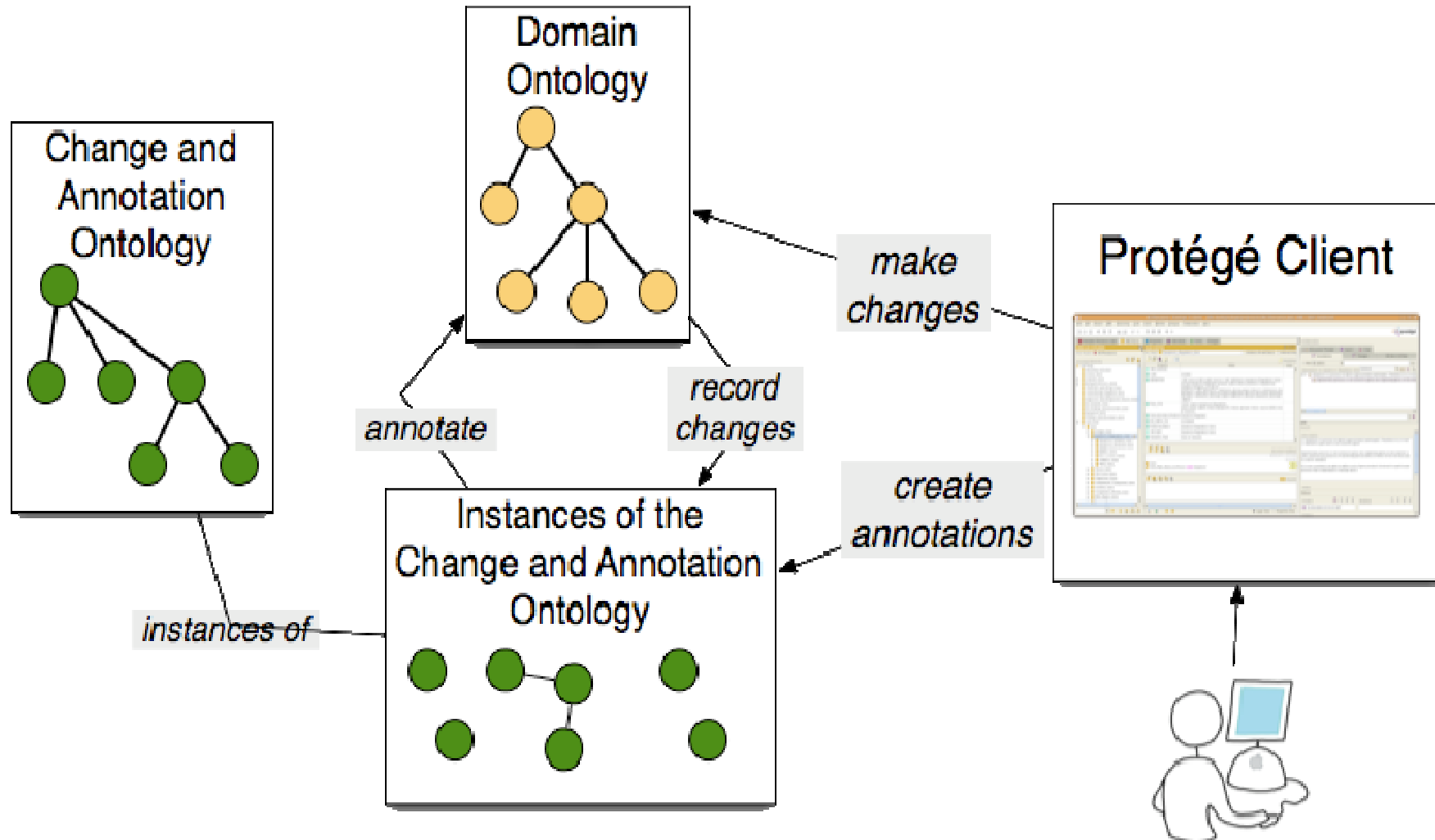
- a few classes with large number of annotations (> 15 each)
- a large number of classes with only one annotation



## Sone additional Stats

- The ratio of created to deleted classes was 2,1 for user7, 2,2 for user 8, 2,3 for user 3, 3 for user 6, 4 for user 5, 4,1 for user 4 and 13,5 for user 2
  - Ratio smaller in users that generally made more changes (outlier user 4), than in more 'careful' users

# Collaborative Protégé Components







# ChaO Instances

annotation\_obi Protégé 3.3 beta (file:\C:\Program%20Files\Protege\_3.3\_beta\obi\branchDevelopment\trunk\annotation\_obi.pprj, RDF Files)

File Edit Project Window Help

Classes Slots Forms Instances Queries Ontoviz Classes & Instances

**CLASS BROWSER**

For Project: ● annotation\_obi

Class Hierarchy

- Change
  - Class\_Change
    - Annotation\_Change
      - Annotation\_Added (30)
      - Annotation\_Modified (7)
      - Annotation\_Removed (26)
    - Class\_Created (3)
    - Class\_Deleted (2)
    - DisjointClass\_Added
    - Documentation\_Added
    - Documentation\_Removed
    - DomainProperty\_Added
    - DomainProperty\_Removed (12)
    - Subclass\_Added (7)
    - Subclass\_Removed (6)
    - Superclass\_Added (4)
    - Superclass\_Removed (4)
    - TemplateSlot\_Added
    - TemplateSlot\_Removed
  - Composite\_Change (41)
  - Created\_Change
    - Class\_Created (3)
    - Individual\_Created
    - Property\_Created
  - Deleted\_Change
    - Class\_Deleted (2)
    - Individual\_Deleted
    - Property\_Deleted
  - Individual\_Change
    - DirectType\_Added
    - DirectType\_Removed
    - Individual\_Added (3)
    - Individual\_Created
    - Individual\_Deleted
    - Individual\_Removed (2)

Superclasses

- Annotation\_Change

**INSTANCE BROWSER**

For Class: ● Annotation\_Added

context

- Annotation Added: alternative\_term: " to class: OBI\_400112
- Annotation Added: alternative\_term: 'liquid\_handler' to class: OBI\_400112
- Annotation Added: alternative\_term: 'liquid\_handling\_instrument' to class: OBI\_400112
- Annotation Added: definition: " to class: OBI\_400132
- Annotation Added: definition: 'An instrument used for automated liquid transfer and handl
- Annotation Added: definition: 'This is a helper class / bin to collect things that will be mod
- Annotation Added: definition\_editor: 'DS: Is this class justified? Its a unnamed class. If so
- Annotation Added: definition\_editor: 'DS: Is this class justified? Its a unnamed class. If so
- Annotation Added: editor\_note: " to class: OBI\_400020
- Annotation Added: editor\_note: " to class: OBI\_400044
- Annotation Added: editor\_note: 'DS: a collection tube is no instrument. This is yet another
- Annotation Added: editor\_note: 'DS: Is this class justified? (has currently only one subcl
- Annotation Added: editor\_note: 'DS: Is this class justified? (has currently only one subcl
- Annotation Added: editor\_note: 'DS: Is this class justified? Its a unnamed class. If so, put
- Annotation Added: editor\_note: 'DS: not an instrument,' to class: OBI\_400057
- Annotation Added: editor\_note: 'DS: Put under liquid\_handler if justified,' to class: OBI\_48
- Annotation Added: editor\_note: 'Is this class justified? (has currently only one subclass)'
- Annotation Added: editor\_note: 'Unclearity if this needs to be cardinal\_part\_of or just part\_
- Annotation Added: preferred\_term: '?liquid\_handler' to class: OBI\_400112
- Annotation Added: preferred\_term: 'cardinal\_part\_of\_fluidic\_system' to class: OBI\_4000
- Annotation Added: preferred\_term: 'part\_of\_electronic\_device' to class: OBI\_400035
- Annotation Added: rdfs:label: " to class: OBI\_400132
- Annotation Added: rdfs:label: 'collected\_relations' to class: OBI\_400132
- Annotation Added: rdfs:label: '?#en ?collection\_tube' to class: OBI\_400020
- Annotation Added: rdfs:label: '?#en ?flow\_cytometer' to class: OBI\_400044
- Annotation Added: rdfs:label: '?#en ?liquid\_handler' to class: OBI\_400112
- Annotation Added: rdfs:label: '?#en cardinal\_part\_of\_fluidic\_system' to class: OBI\_4000
- Annotation Added: rdfs:label: '?#en part\_of\_electronic\_device' to class: OBI\_400035

Types

**CLASS EDITOR**

For Class: ● Annotation\_Added (instance of :STANDARD-CLASS)

Name: Annotation\_Added

Documentation:

Constraints:

Role: Concrete ●

Template Slots

Name	Cardinality	Type	Other F
action	single	String	
applyTo	single	Instance of Ontology_Cor...	
associatedAnnotations	multiple	Instance of Annotation	inverse-slot=annotates
associatedProperty	single	Instance of Ontology_Pro...	
author	single	String	
context	single	String	
partOfCompositeChan...	single	Instance of Composite_Cf...	inverse-slot=subChanges
timestamp	single	Instance of Timestamp	

start Draft... 3 W... Wind... Prote... anno... Docu... EndN... 17:35

# Webprotege (lightweight, configurable)

The screenshot displays the WebProtégé web interface within a Mozilla Firefox browser window. The browser's address bar shows the URL: `http://smi-protege:8888/edu.stanford.bmir.protege.web.WebPro`. The interface includes a navigation menu with tabs for 'My WebProtégé', 'Ontology Metadata', 'Classes', 'Properties', and 'Individuals'. The main content area is titled 'Ontology: NCIThesaurus 0712' and features a 'Class tree' on the left, a 'Notes for Cancer\_Gene' section in the center, and a 'Properties for Cancer\_Gene' table at the bottom. A configuration menu is open on the right side, showing options for 'My WebProtege', 'Ontology Metadata', 'Classes', 'Properties', 'Individuals', 'Discussions', and 'Visualization'. The 'Properties for Cancer\_Gene' table is as follows:

Property	Value
rdfs:label	Cancer Gene
code	C19540
Preferred_Name	Cancer Gene
Semantic_Type	Gene or Genome
Synonym	Cancer Gene
Synonym	Cancer-Promoting Gene

<http://bmir-protege-dev1.stanford.edu/webprotege/WebProtege.html>

# Collaborative Protégé components

## Ontology editor component

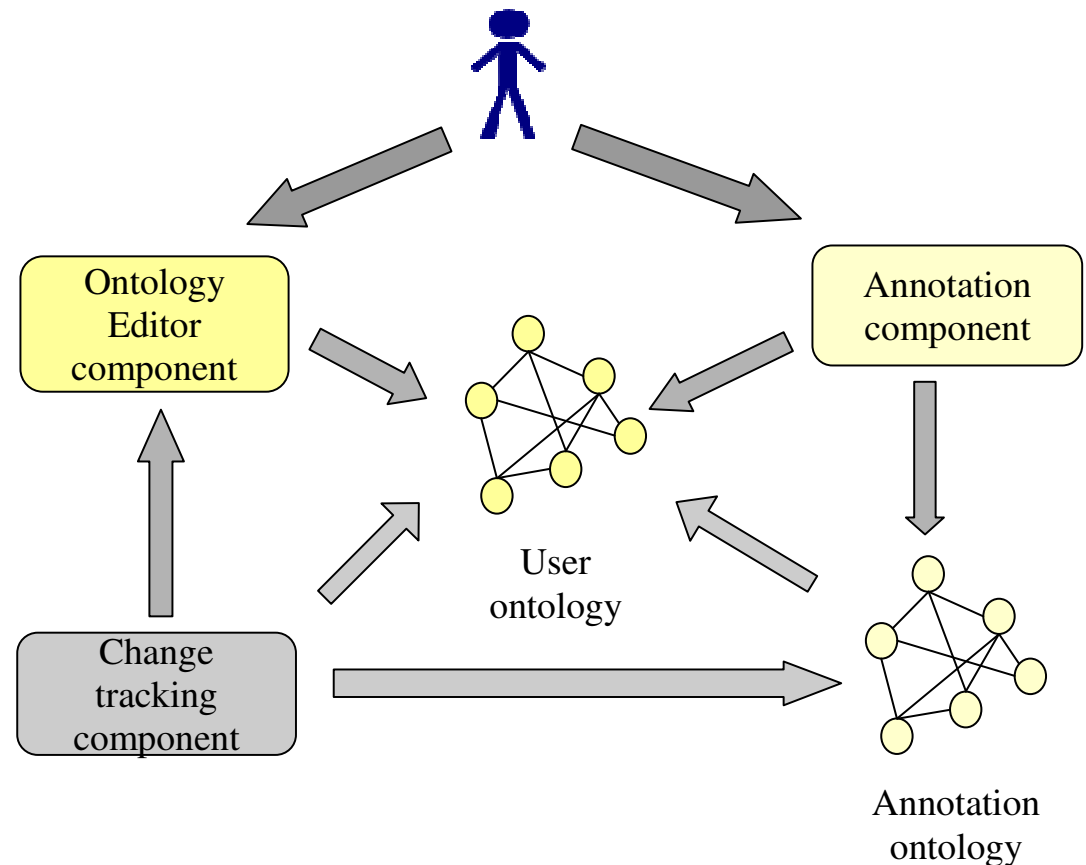
- basic ontology editing functionalities

## Annotation component

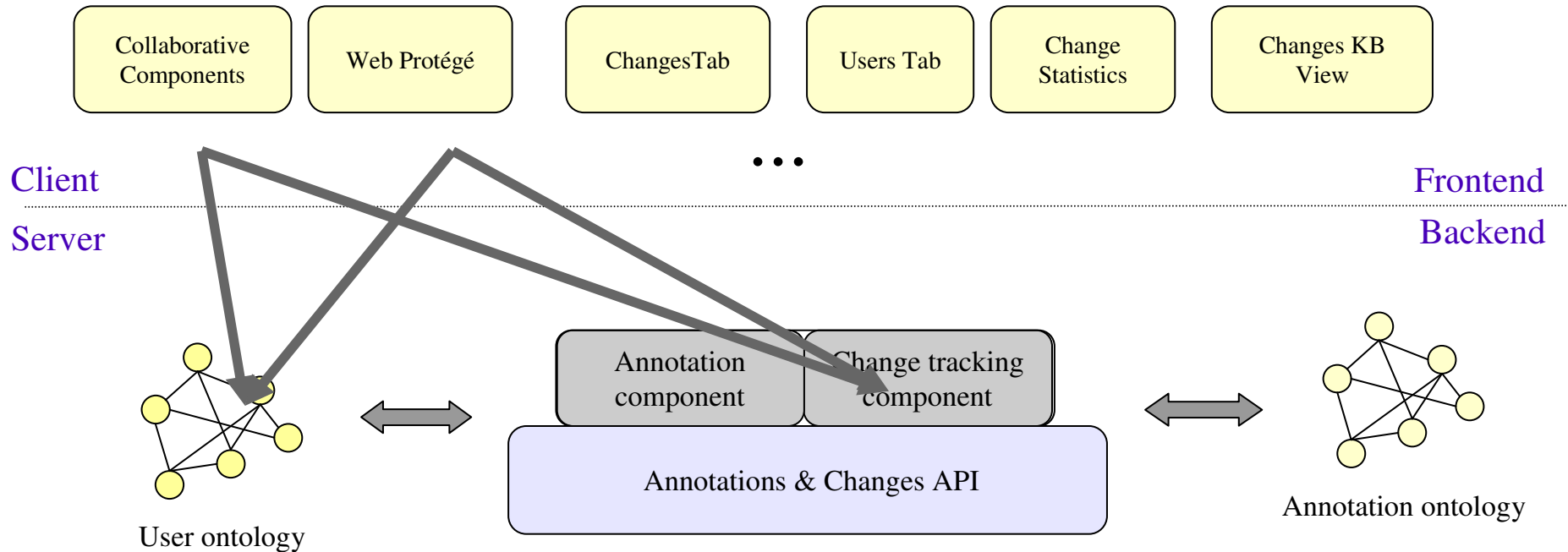
- user ontology is annotated with annotation instances from the Annotation ontology

## Change tracking component

- changes are stored as instance of the Annotation ontology



# Client-server architecture



Applications and components on the client side use common **Changes & Annotations API** to manipulate annotations and changes

Supports OWL, RDF(S), Frames, as well as back-ends: database, XML

Java-based, plug-in architecture