

Annotation and Project Management Resource Integrated with the ProteomeCommons.org Tranche Repository

James Hill¹; Bryan Smith¹; Mark Gjulkich¹; Panagiotis G Papoulias¹; Jayson Falkner²; Philip Andrews¹
¹University of Michigan, Ann Arbor , MI; ²SOSI, Portland, OR

Introduction

Annotation has been a difficult problem in the field of proteomics with compliance and ease of use being two of the major concerns. Controlled vocabularies, ontologies, and file formats are in various stages of completion - adding to the challenges. As these community standards are finalized, however, it becomes feasible to develop standardized annotation tools. The re-engineered ProteomeCommons.org applies concepts from social networks to project and annotation management. It has also integrated with the ProteomeCommons.org Tranche Repository and, through the Proteomexchange, will be integrated with other proteomics resources including PRIDE, Peptide Atlas, and TheGPMdb. Dissemination through Tranche increases your citation opportunities.

Key Points

- Dissemination management system for project deliverables
- Data storage and dissemination through a Tranche repository
- Collaborative annotation of data sets
- Data is replicated to prohibit loss of files using Tranche
- Open-source and free to use
- Meets NIH requirements for dissemination of data sets
- Provides MIAPE compliant annotations

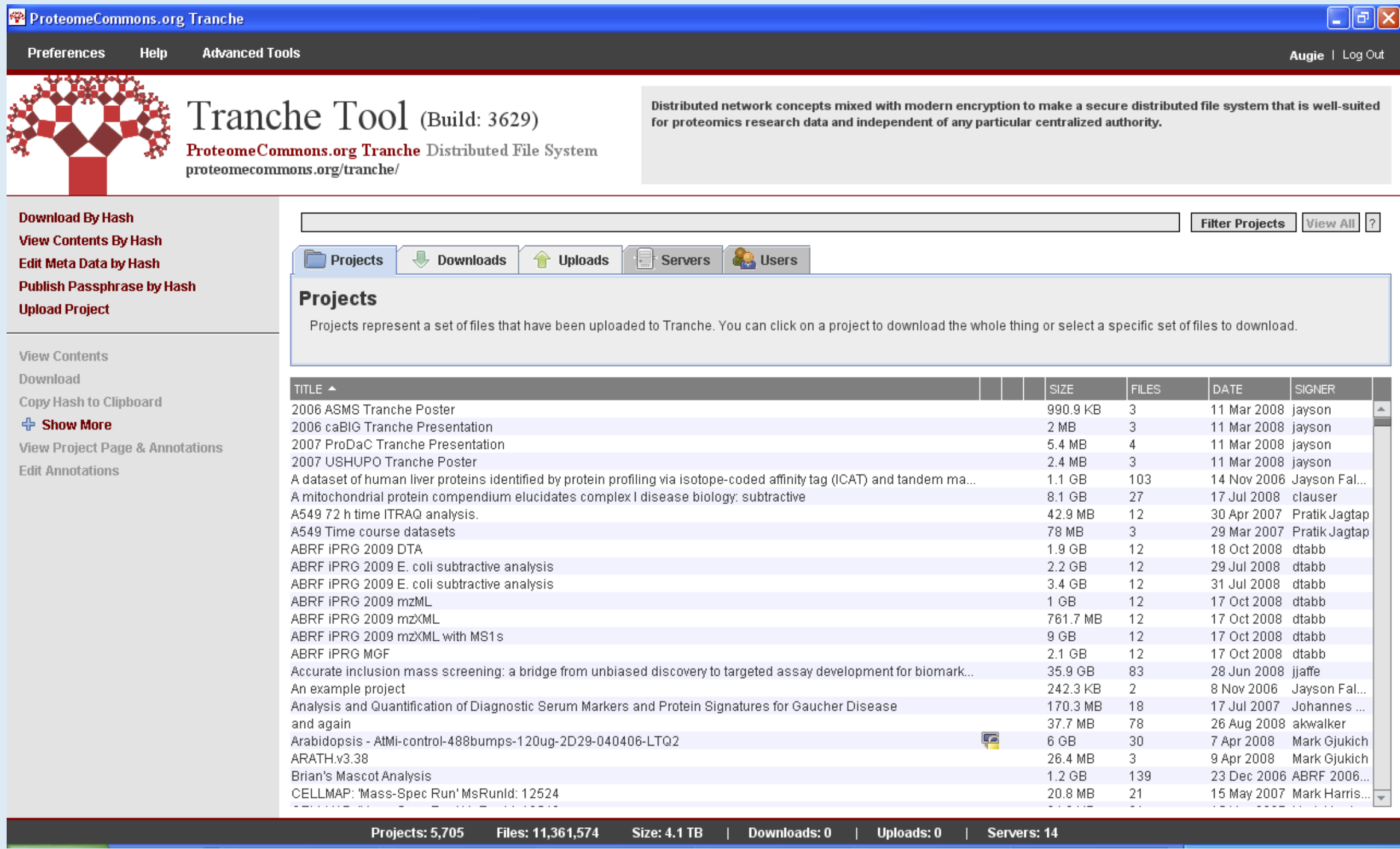
Availability

This project is on-line now and requires a web browser to use. All code for both the Tranche Distributed Repository project and the Tranche Annotations project is open-source and free to use both commercially and non-commercially.
Link: <https://proteomecommons.org>

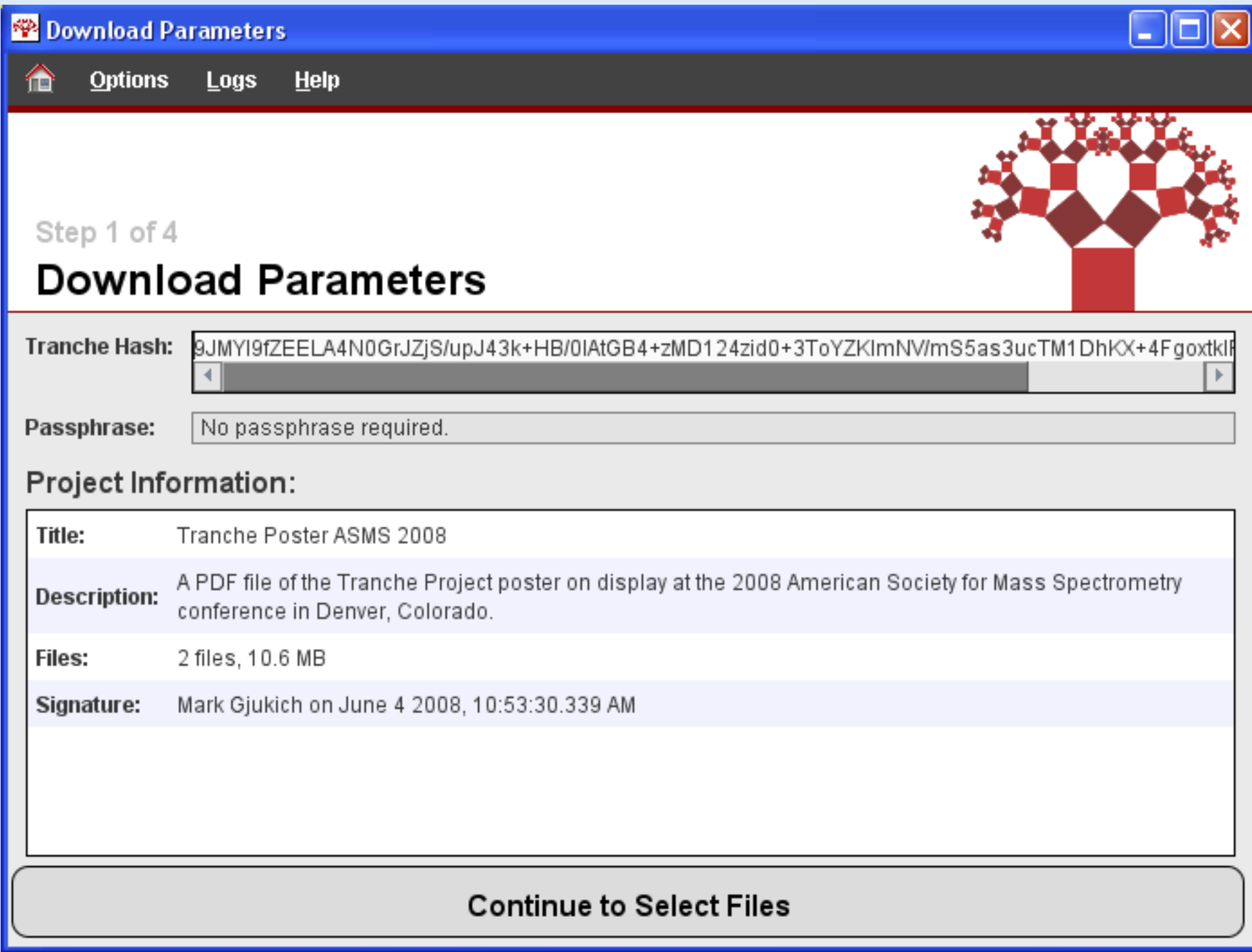
Methods

Tranche Repository

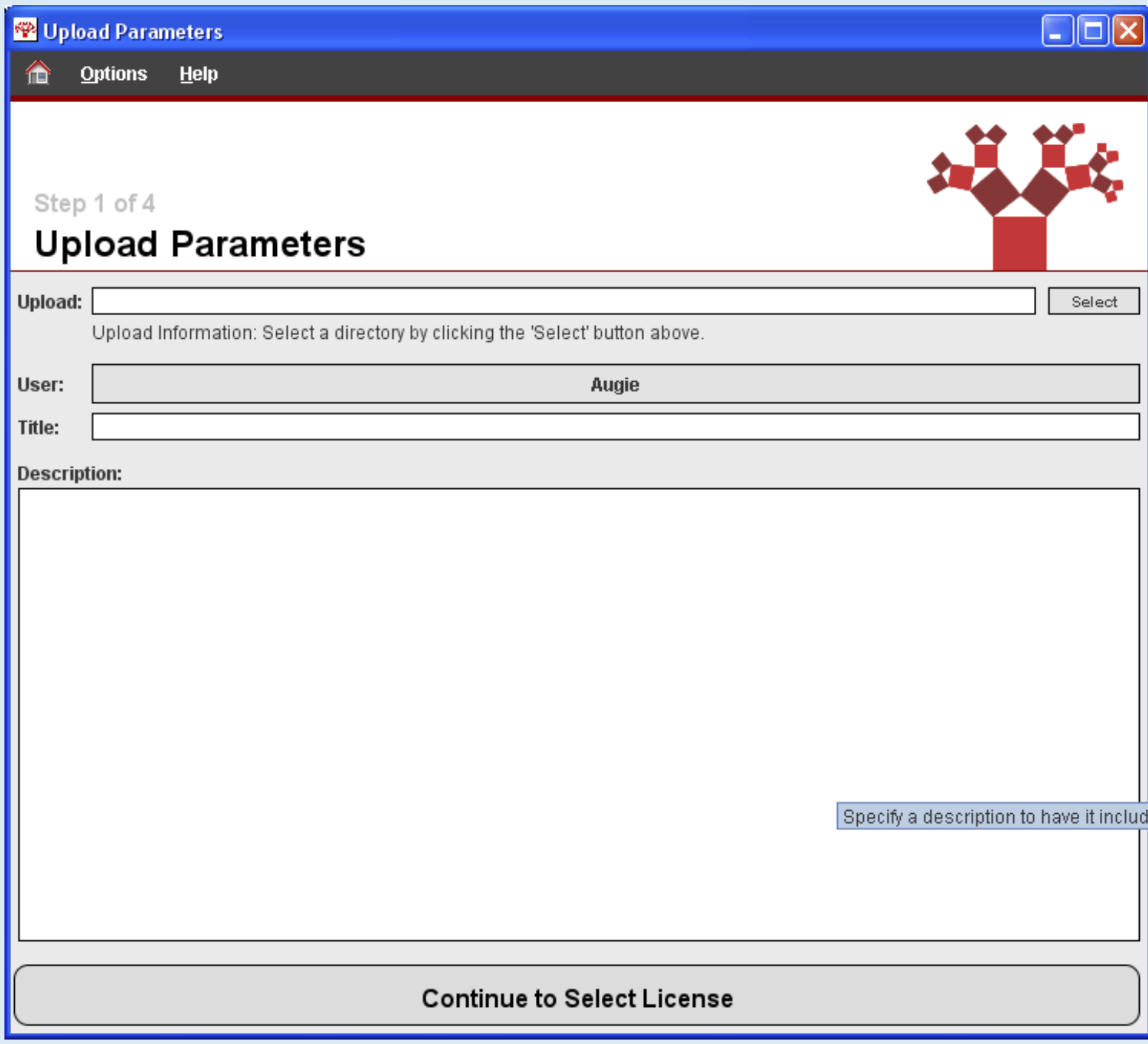
- Data storage and dissemination
- Can handle very large data sets
- Verifies that data hasn't changed since publication
- Open access to data
- Secure sharing of private data and means of public dissemination
- Easy-to-Use



Tranche Advanced GUI - The primary interface for working with a Tranche repository. Anonymous users can use this to browse and download data sets, while registered users can manage existing data sets and upload new data sets.



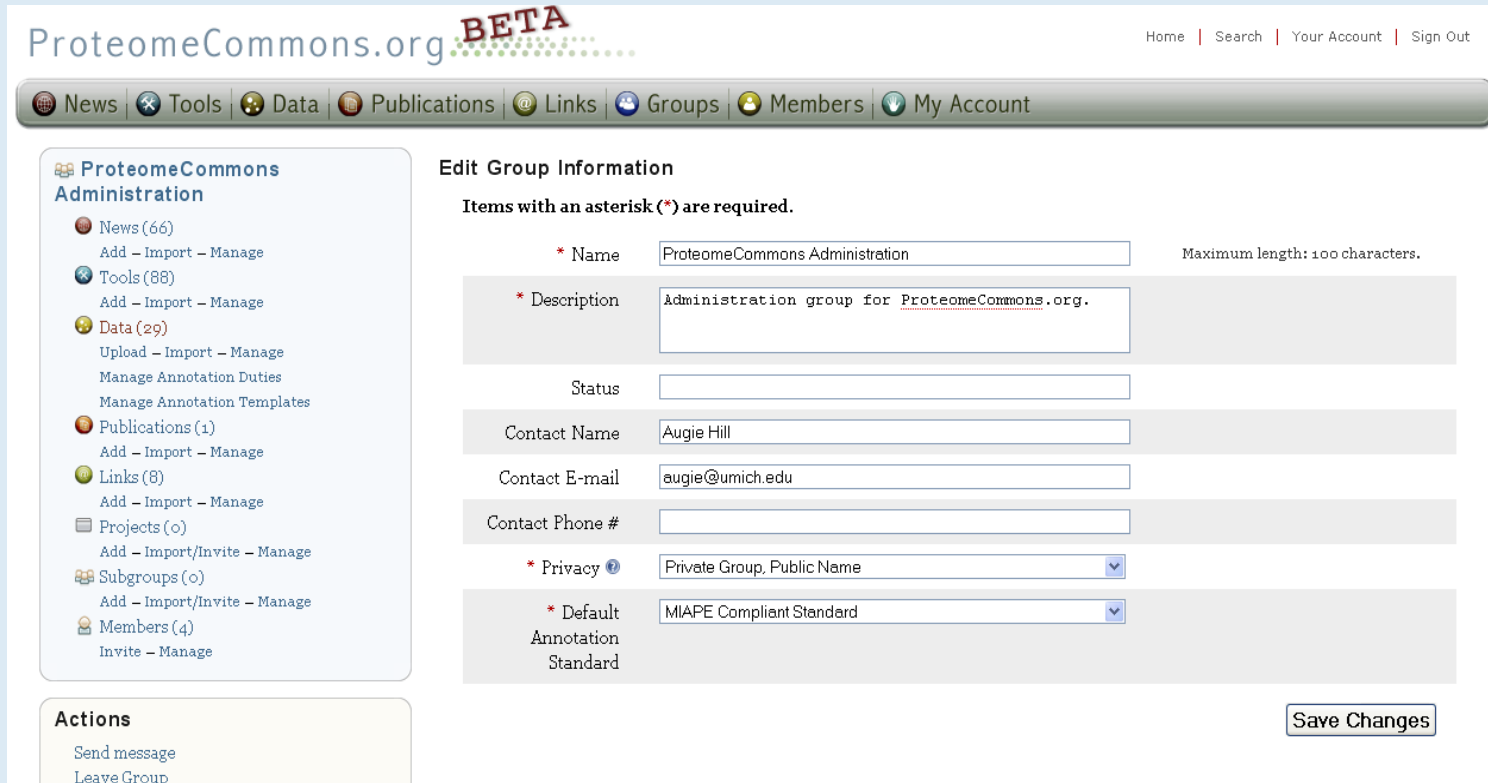
Tranche Download GUI - The interface for downloading from a Tranche repository. Users can select a subset of files to download instead of the full data set.



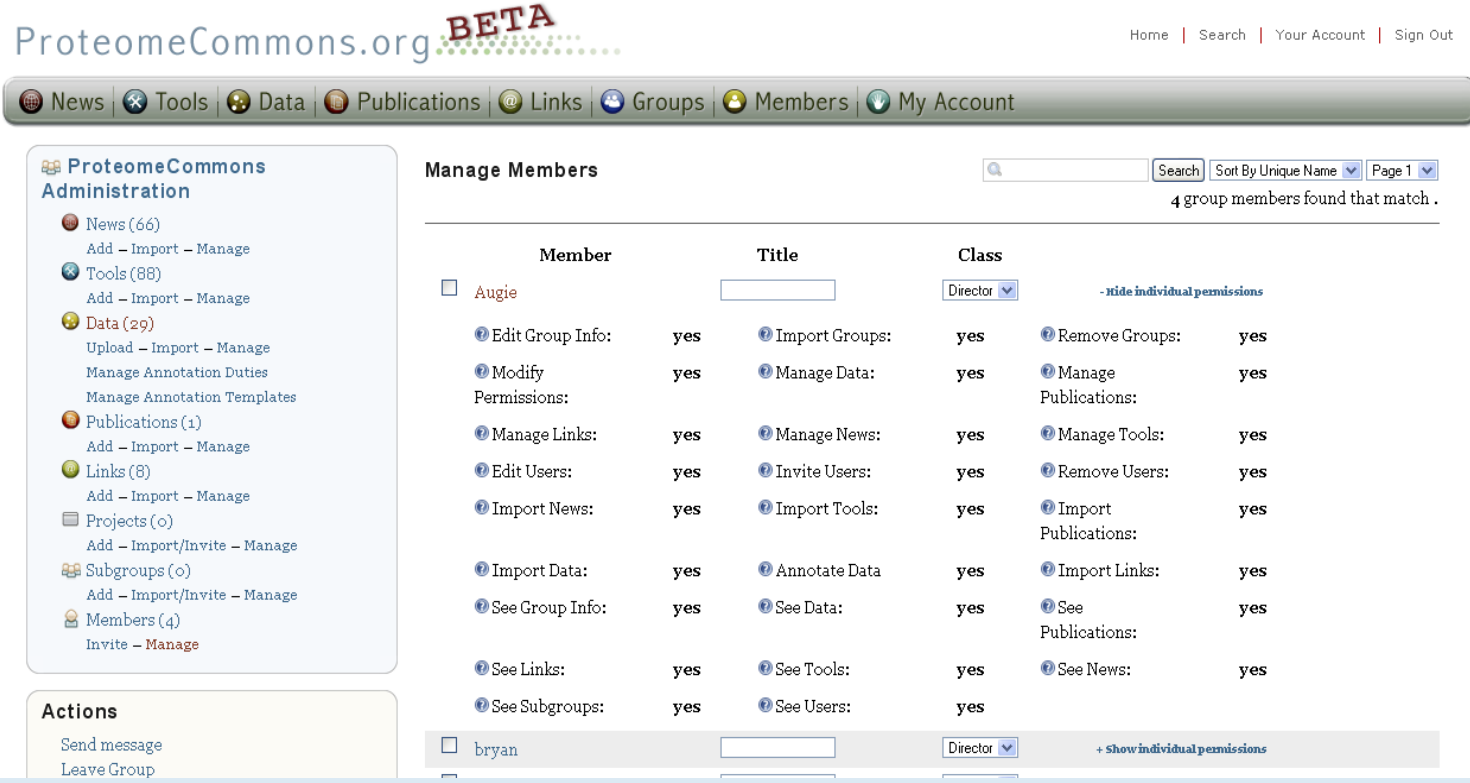
Tranche Upload GUI - The interface for uploading to a Tranche repository. Only registered users can upload. Data uploaded to a Tranche repository will be downloaded exactly as it is uploaded.

Annotation and Project Management

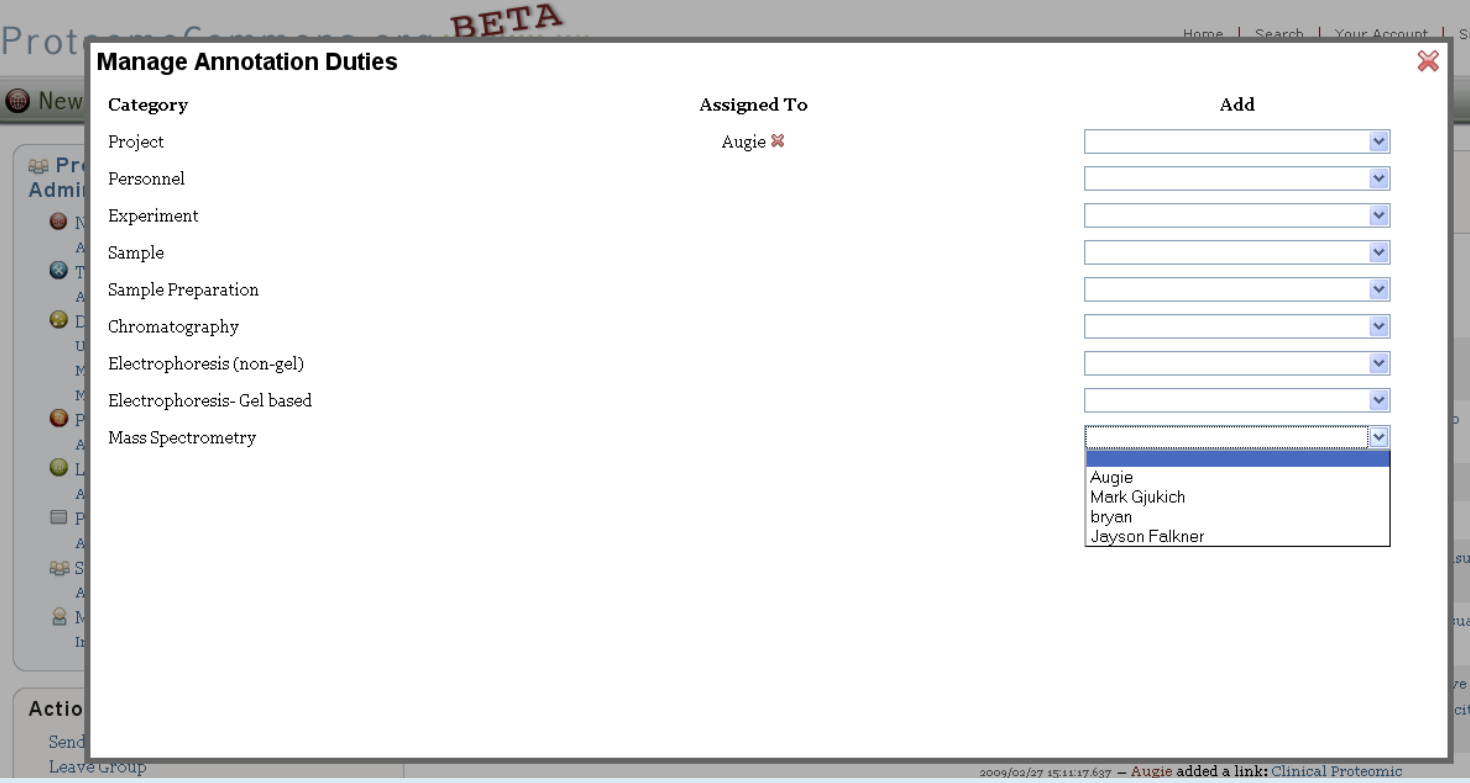
- Annotate data sets that exist within the Tranche repository
- Assign particular sections of annotations to domain experts
- Group leaders can assign annotation responsibilities
- Annotations database submitted for caBIG Silver compatibility review
- Tranche data management: Make Public, Hide, Delete
- Users can work independently and/or with a Group



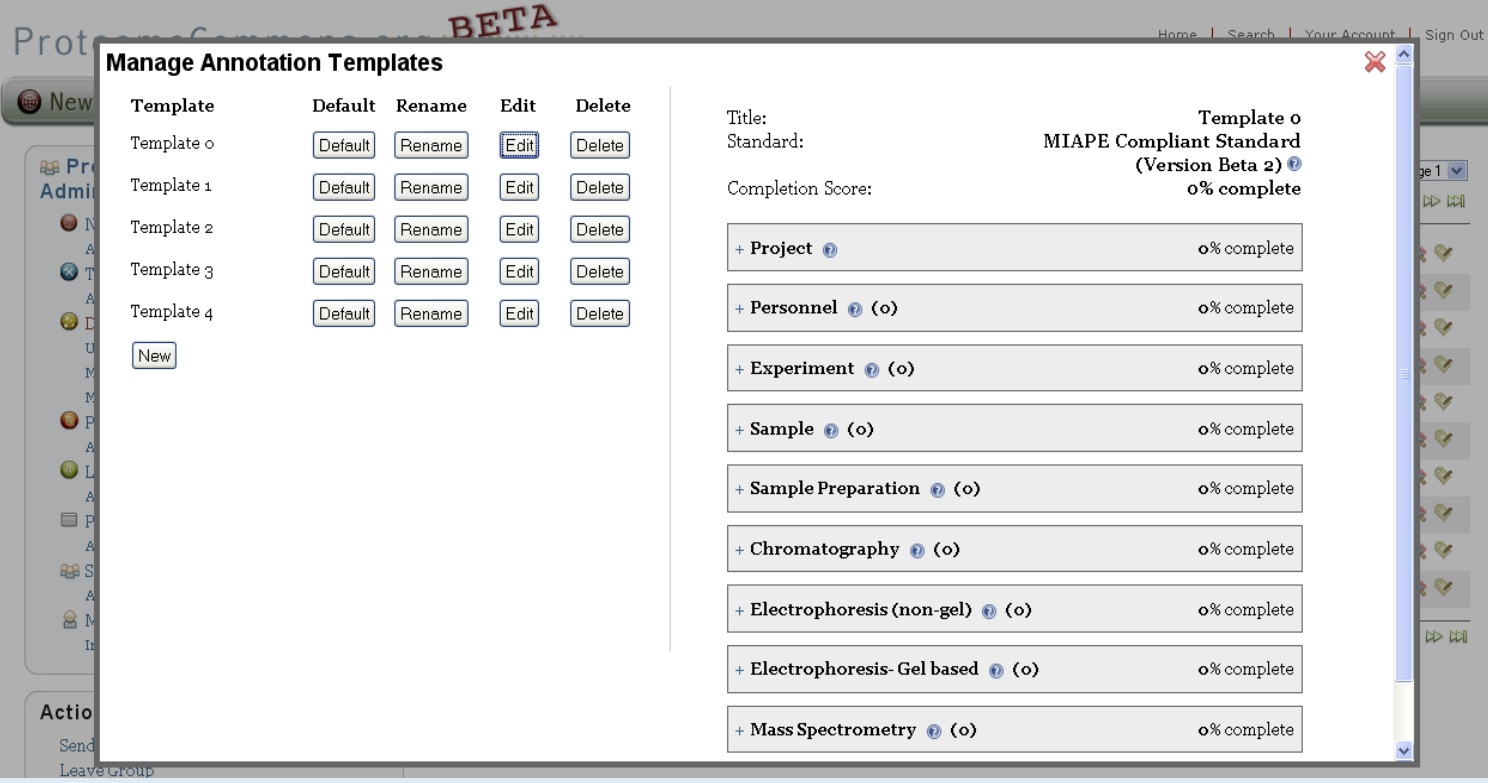
Users can create and modify groups, then invite and set permissions for other users.



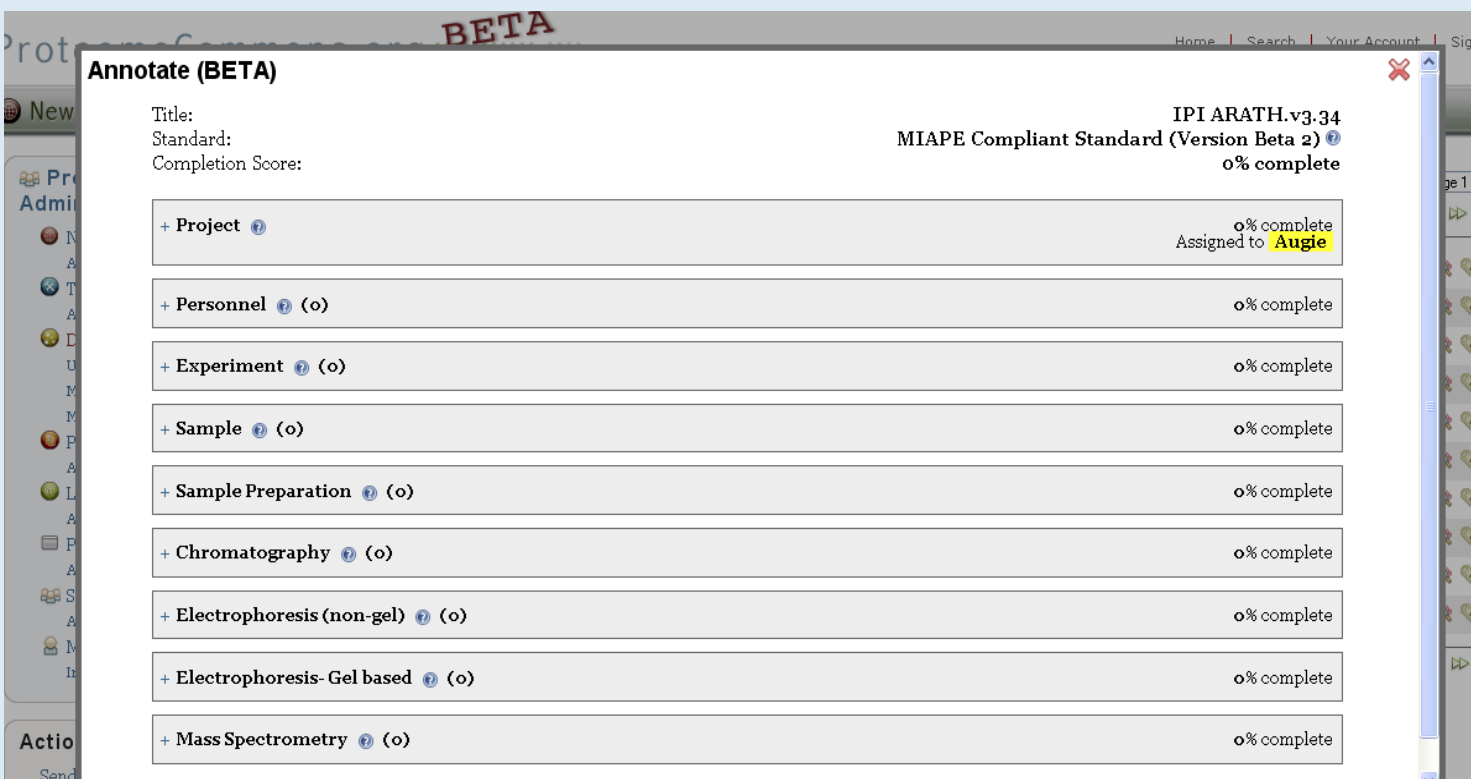
Managing permissions for members of the group can be done by assigning a permission class or setting each permission individually.



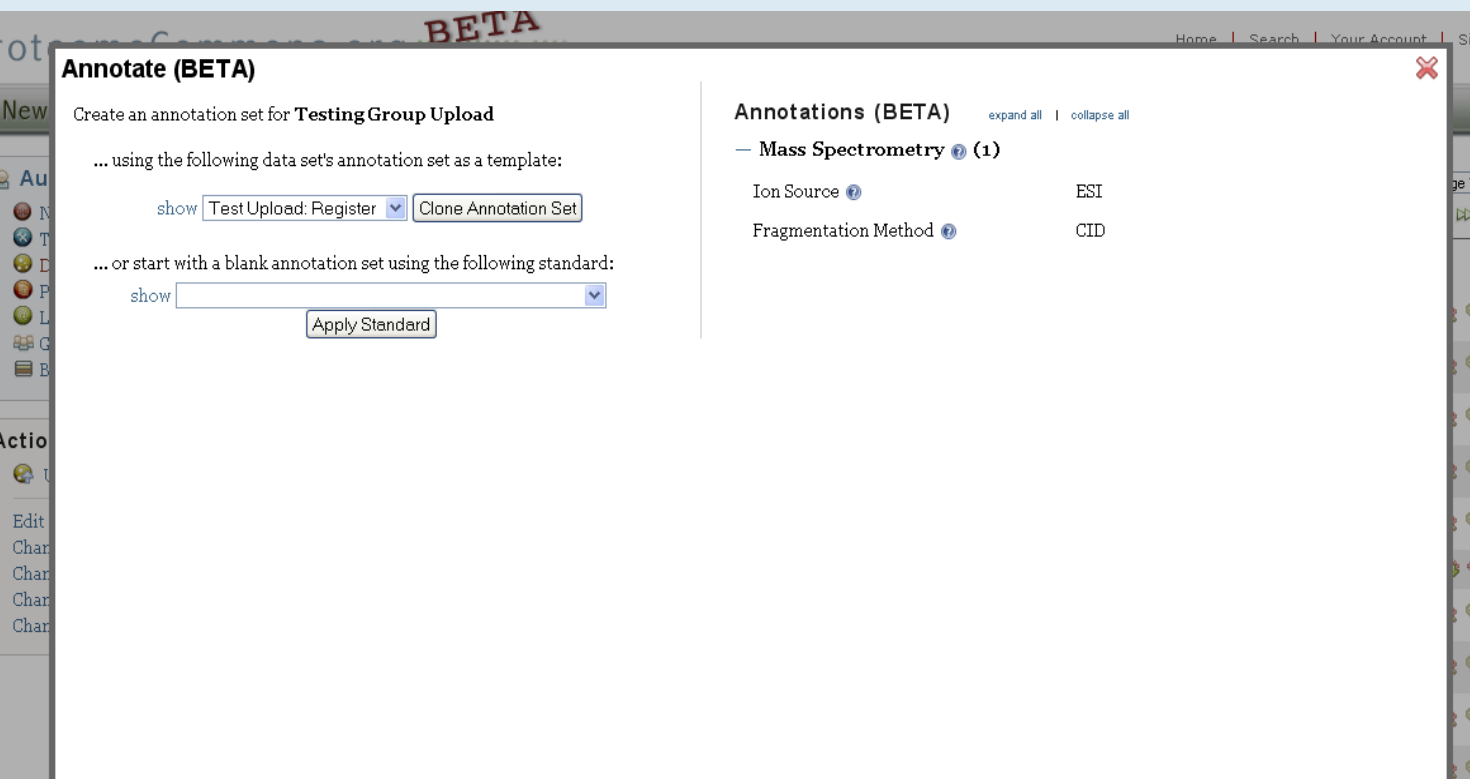
Assign sections of data set annotations to members of the group. On the left are the names of the annotation sections, in the middle are the members assigned to that section, and on the right are the members that are not assigned to the section.



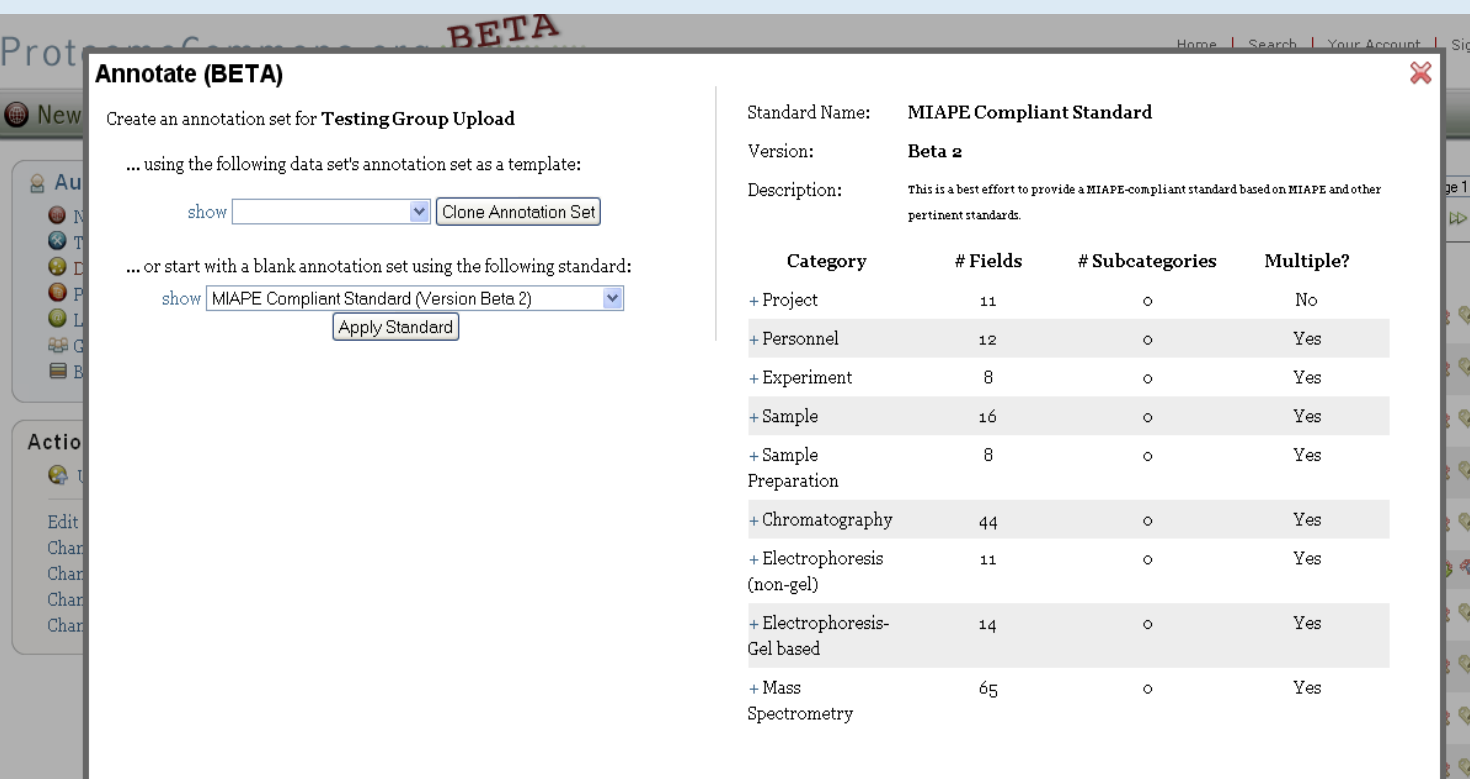
Groups and projects can create and manage templates of annotation sets. This saves time in annotating many data sets with similar properties.



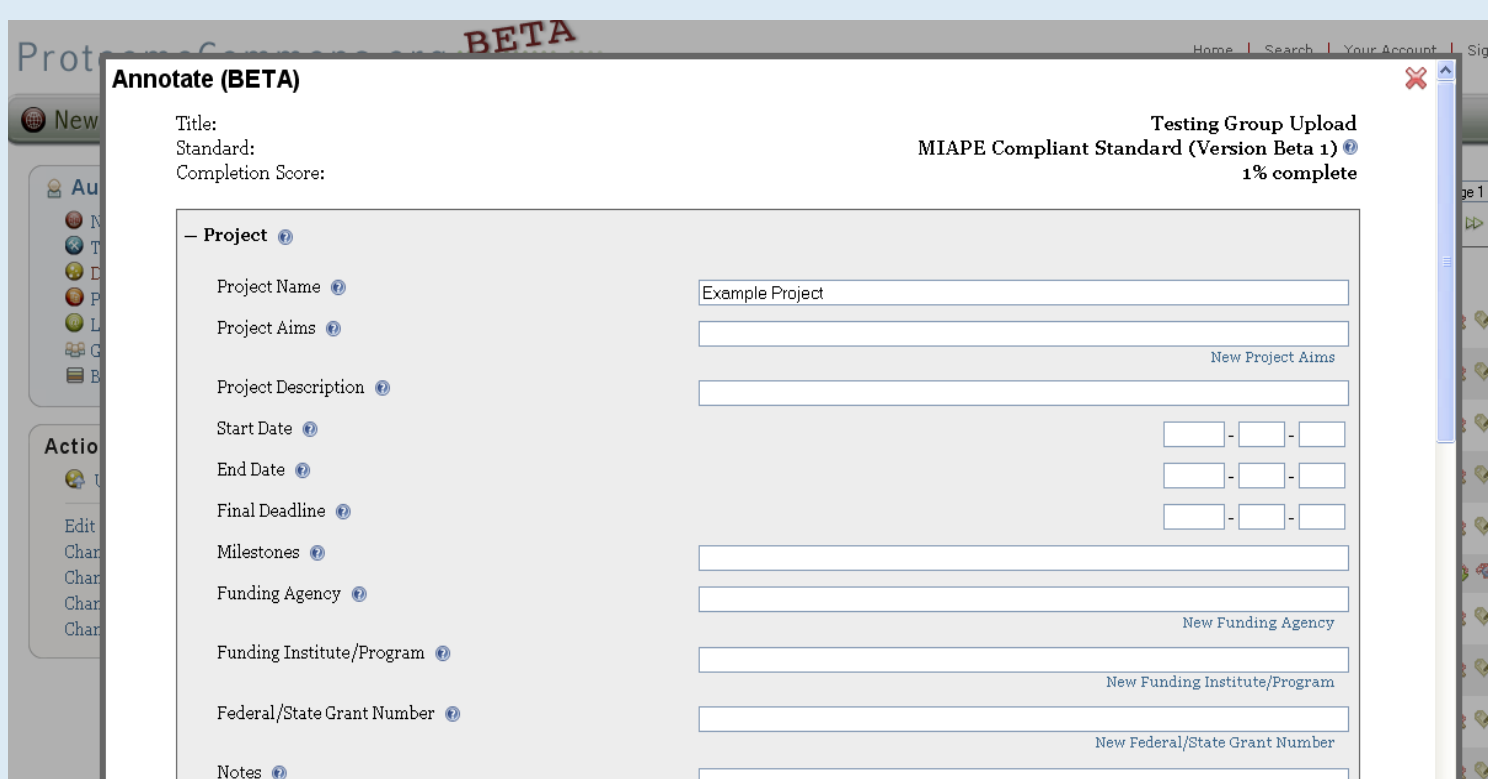
Members of groups can be assigned sections of annotations they are responsible for.



When a user begins to annotate a data set, they can either start with a blank form, start from a previously created template, or clone a previously annotated data set.



Members can view a breakdown of the annotation set and categories.



Annotation can be completed and edited using simple web-based forms online. Confirmation of percent complete is auto-updated and can be monitored for progress.

Statistics and Usage

The Tranche project was started in 2006 and has grown to include millions of files and participating groups around the world.

<https://trancheproject.org>
<https://proteomecommons.org>

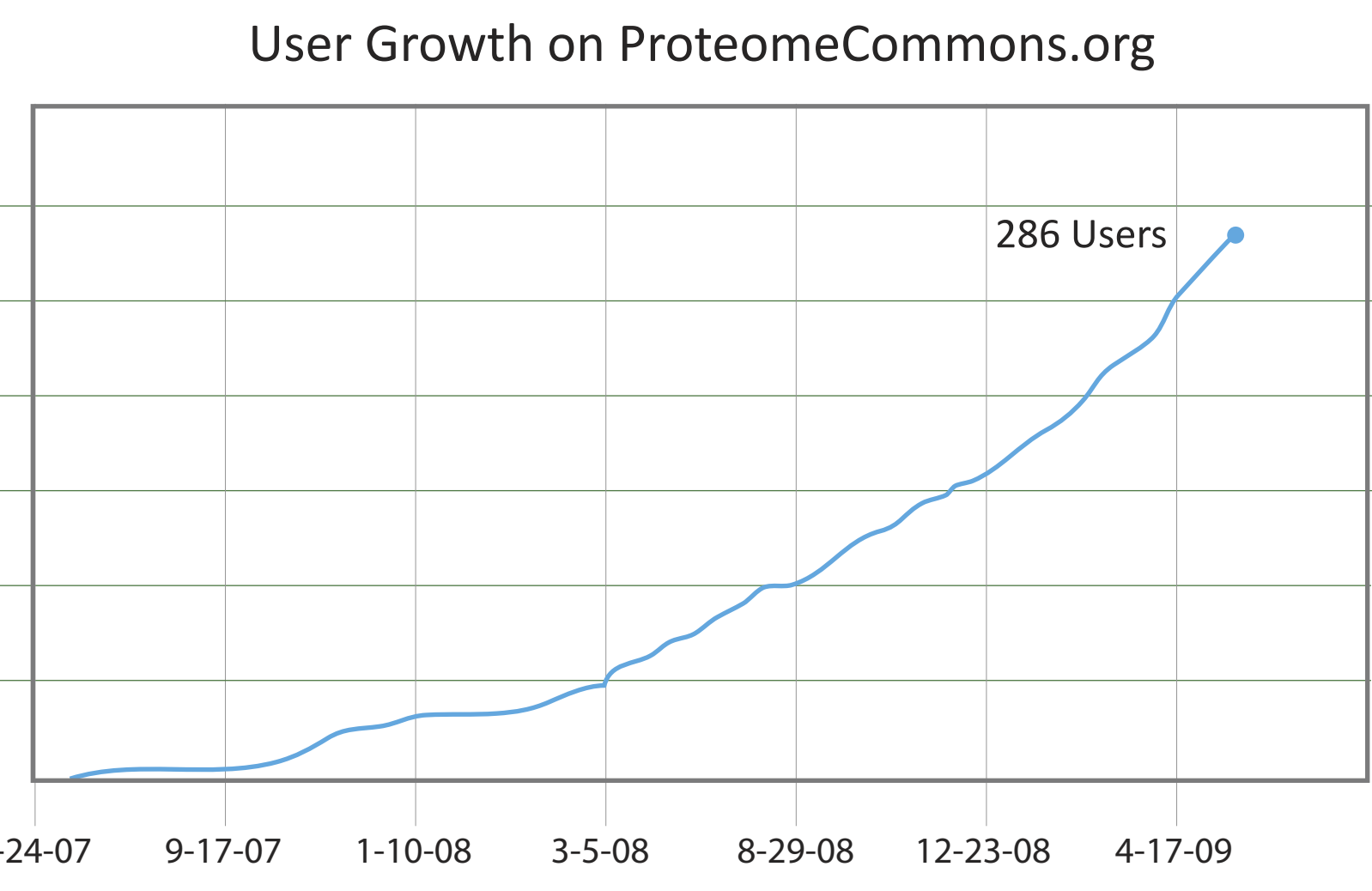
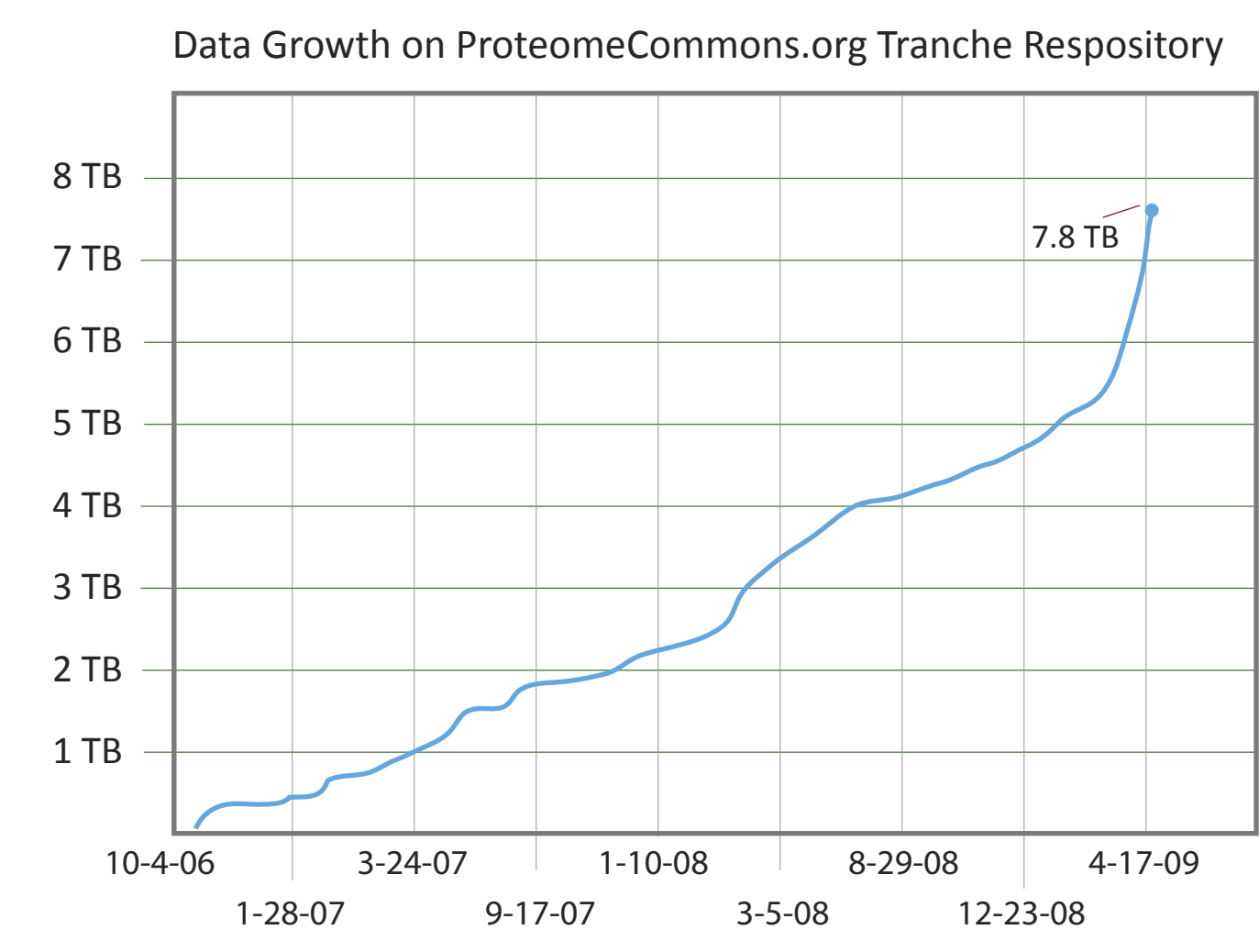


ProteomeCommons.org Tranche Repository Statistics

Files On-line..... **12.5 Million**
Servers On-line..... **15**
Total Data On-line..... **9 TB**
Members..... **291**

Growth by year
2008
Data Sets On-line: 5,435
Total Size: 3 TB
Files Online: 11.2 Million
2007
Data Sets On-line: 3,496
Total Size: 890 GB
Files Online: ~9 Million

Growth Charts



Conclusion

Data sets are given meaning when they can be put into a context. Using the ProteomeCommons.org project and annotation management resource, users can upload data sets and annotate them in collaboration with their colleagues.

Along with being able to host all the raw data associated with an experiment is the need to annotate these data with information about the experiment. ProteomeCommons.org provides a mechanism for letting annotations be linked to raw data, which is intended for future use with upcoming proteomics annotation standards. This provides a practical option for researchers to meet recently proposed requirements for publishing data.

The ProteomeCommons.org and Tranche Repository community resources are primarily sponsored by the NCRR grant #P41-RR018627 and the NCI subcontract #27XS115. Please contact markgj@umich.edu for any questions or comments regarding these projects.

Poster PDF and Additional Information:
<https://proteomecommons.org>