

ProteomeCommons.org

Annotations and Project Management Recourse

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

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
- Detailed permissions within groups
- Internal messaging system between users
- Powerful search capabilities

Poster PDF and Additional Information:
The ProteomeCommons.org and Tranche Repository community resources are primarily sponsored by the NCCR grant #P41-RR018627 and the NCI subcontract #27XS115.

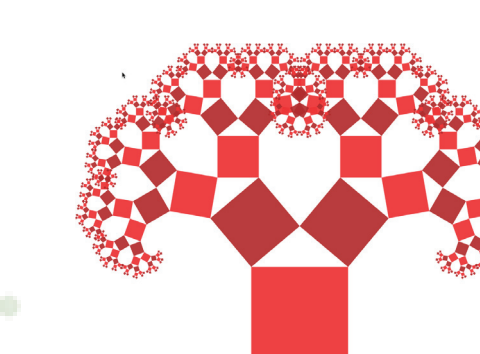
Please contact markgj@umich.edu for any questions or comments regarding these projects.

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](https://proteomecommons.org)



Tranche Project

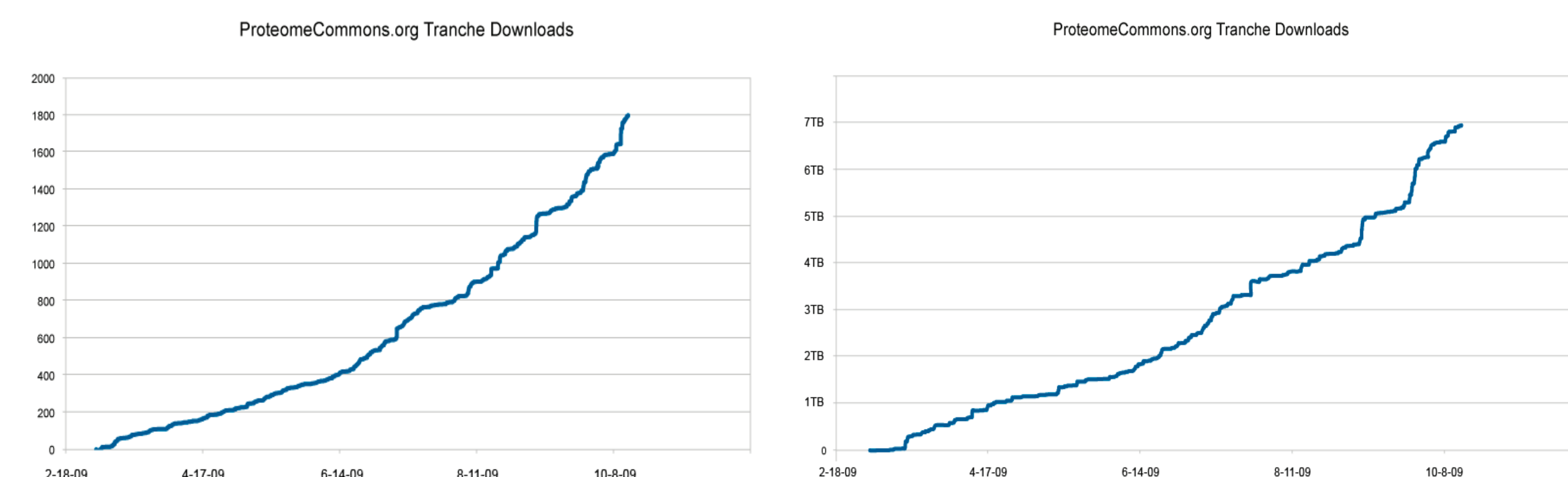
ProteomeCommons.org Tranche Repository Statistics

Files On-line.....**12.6 Million**
Servers On-line.....**16**
Total Data On-line.....**10.7 TB**
Members.....**391**

Growth by year

2008 ff
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.