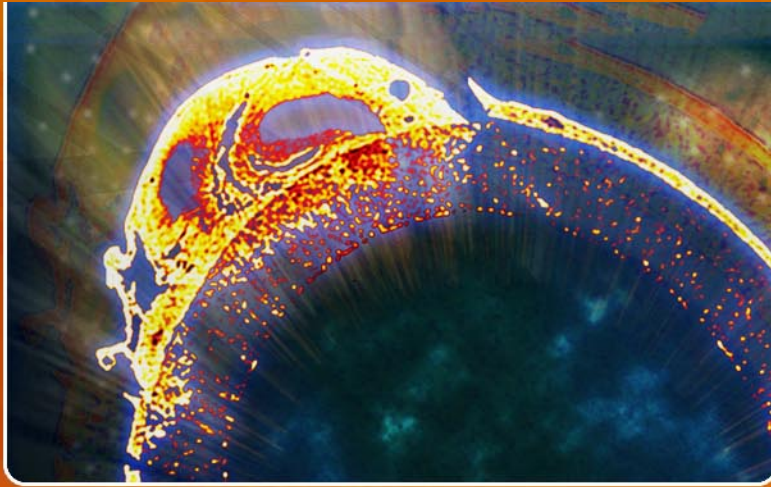




**Pacific Northwest**  
NATIONAL LABORATORY

*Proudly Operated by Battelle Since 1965*



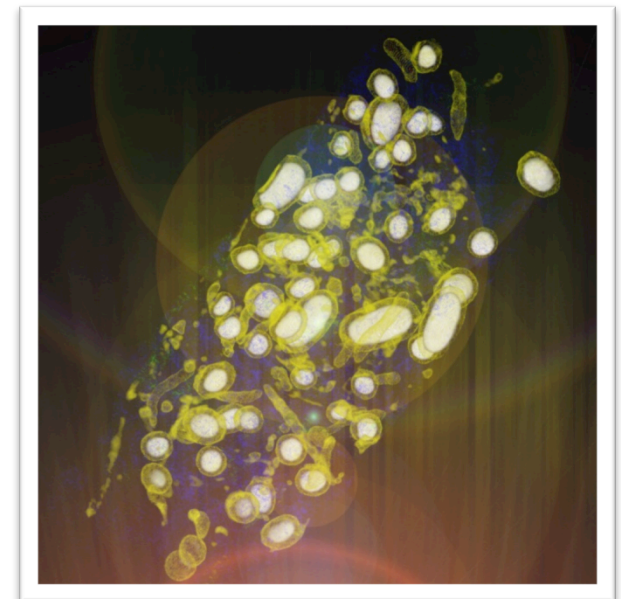
# Using Globus Online for Near-Real-Time Remote Analysis

CARINA LANSING & ZOE GUILLEN

Pacific Northwest National Laboratory

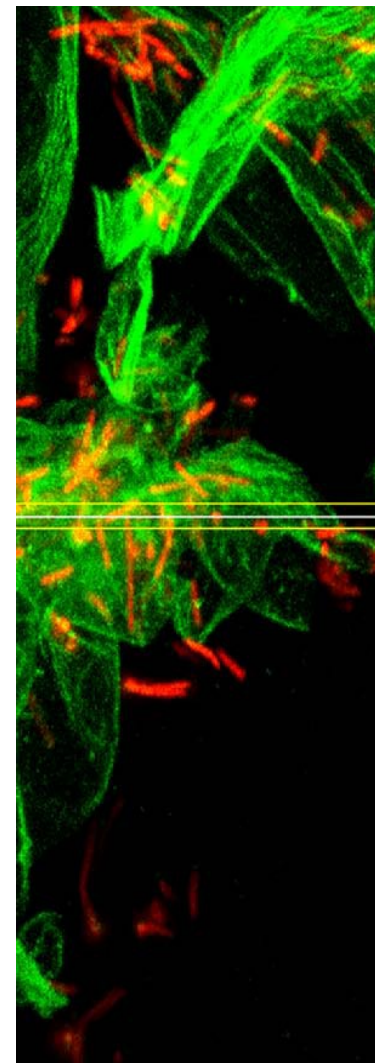
Globus World 2013

- ▶ The goal of PNNL's Chemical Imaging Initiative is to deliver new capabilities for in situ molecular-scale imaging
- ▶ The Chemical Imaging Initiative is developing a suite of tools with nanometer-scale resolution and element specificity which will allow scientists to go from observing to manipulating systems on a molecular level.
- ▶ Moving towards real time analysis
- ▶ Ability to deal with large data volumes (TBs)
- ▶ Enable multi-modal analysis
- ▶ Provide open framework



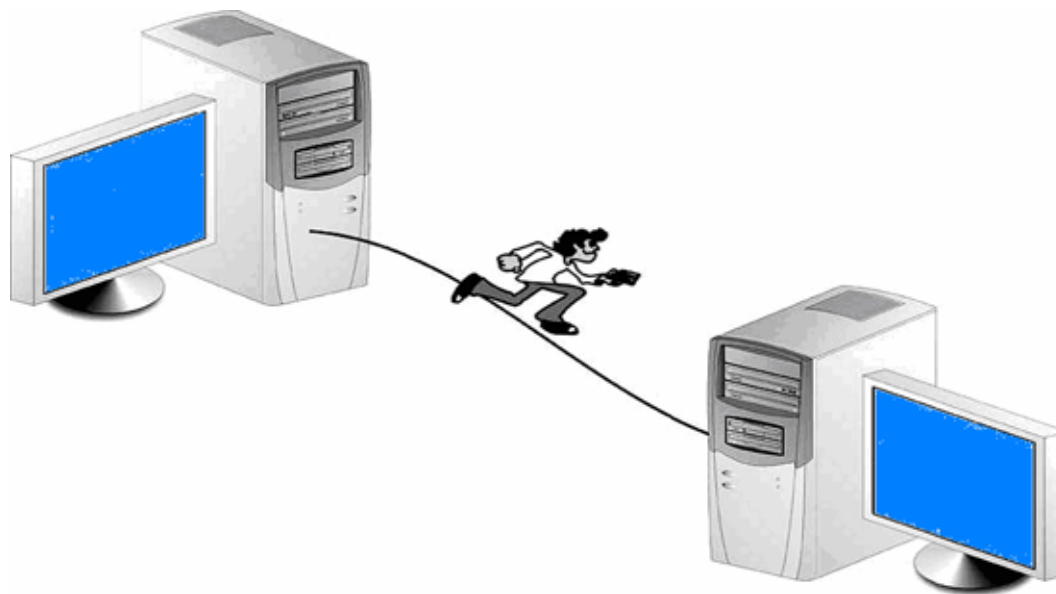
# CII Tomography Challenges

- ▶ Scientists use synchrotron microtomography at the Advanced Photon Source at ANL, which allows for high resolution imaging.
- ▶ Terabytes of data are generated remotely.
- ▶ Want near-real time processing on current sample in order to make decisions that affect same or subsequent samples.
- ▶ Would like to know what to change before sample quality degrades due to electron beam.
- ▶ PNNL's existing tools require an HPC Windows environment, which is difficult to configure remotely.



# Previous Solution

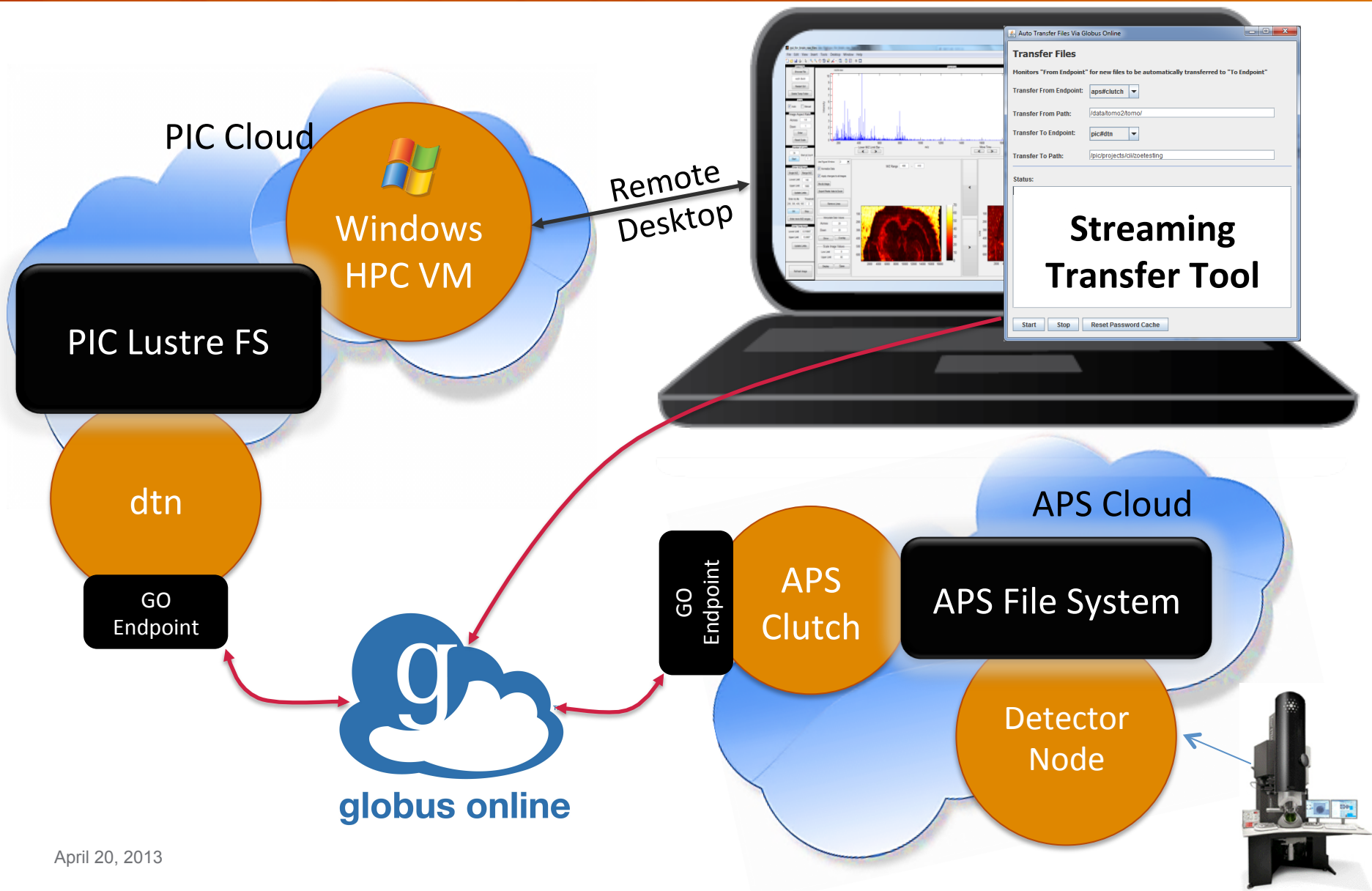
- ▶ Previously would only be able to analyze a handful of image slices because they had to move data around using a thumb drive and run code on their laptop.
- ▶ Would need to bring a hard drive with them to transfer all the data after the experiment was done and then ship home.
- ▶ Would take weeks to analyze all the samples.



- ▶ Developed a streaming transfer utility using Globus Online Transfer API.
- ▶ Utilized the “recursive” option of the transfer command to effectively sync folders from the source endpoint to the destination endpoint.
- ▶ Cached passwords.
- ▶ Streamed data to PNNL’s high performance institutional computing cluster (PIC).
- ▶ Provided high powered virtual machines on PIC with which to run the current tools (many Windows-based).
- ▶ Tested initial prototype with Erin Miller’s APS beamtime (March 13-16, 2013)

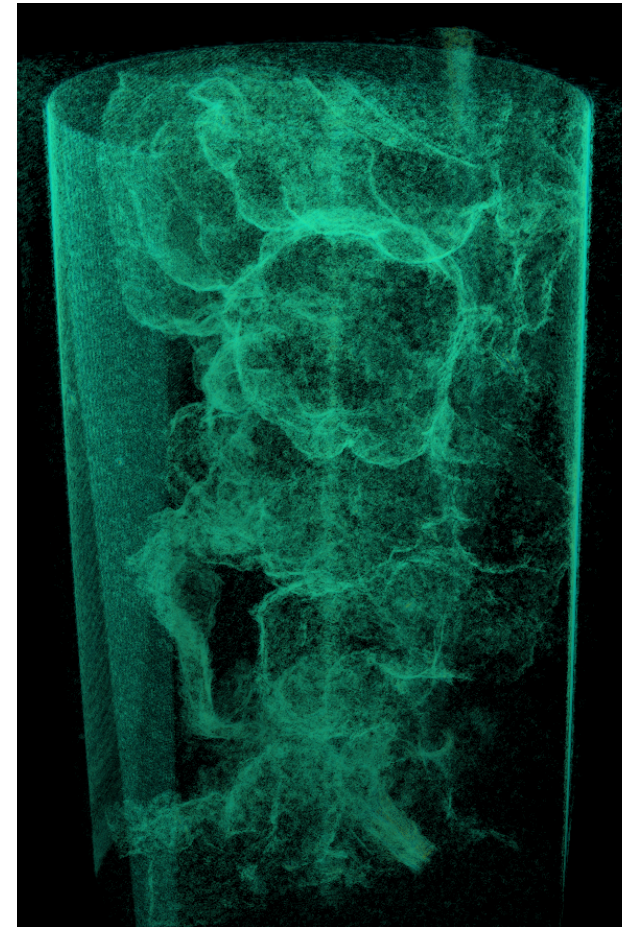


# Initial Architecture



- ▶ It took us a little time to get familiar with the Globus online API, however Globus Online developers were extremely helpful and quick to answer questions or debug problems.
- ▶ Authentication - users complained at having to remember too many passwords. No time to set up an SSO solution, so we cached for them.
- ▶ We locked out Erin's GO account:
  - Transfers weren't completing because they kept retrying on access failure
  - We were making transfer requests without waiting for previous transfer to complete
- ▶ Wish List:
  - More shell-like commands for endpoints
  - Be able to turn off email notifications
- ▶ We could greatly accelerate the processing time if we could adapt some of the users' Windows tools for an HPC environment.

- ▶ Success – implementation pulled together in a few weeks!
- ▶ 1.6 TB of data (71 data sets) were transferred over 3 days
- ▶ 17 samples were reconstructed at APS (~15 minutes each) with results streamed back to PNNL
- ▶ Reconstructed data were ready for visualization at PNNL within minutes of reconstruction completing
- ▶ Visualization for each sample completed within 10 minutes.
- ▶ Seamless beamtime experience
- ▶ Erin did not have to cart a hard drive home to bring back her data 😊





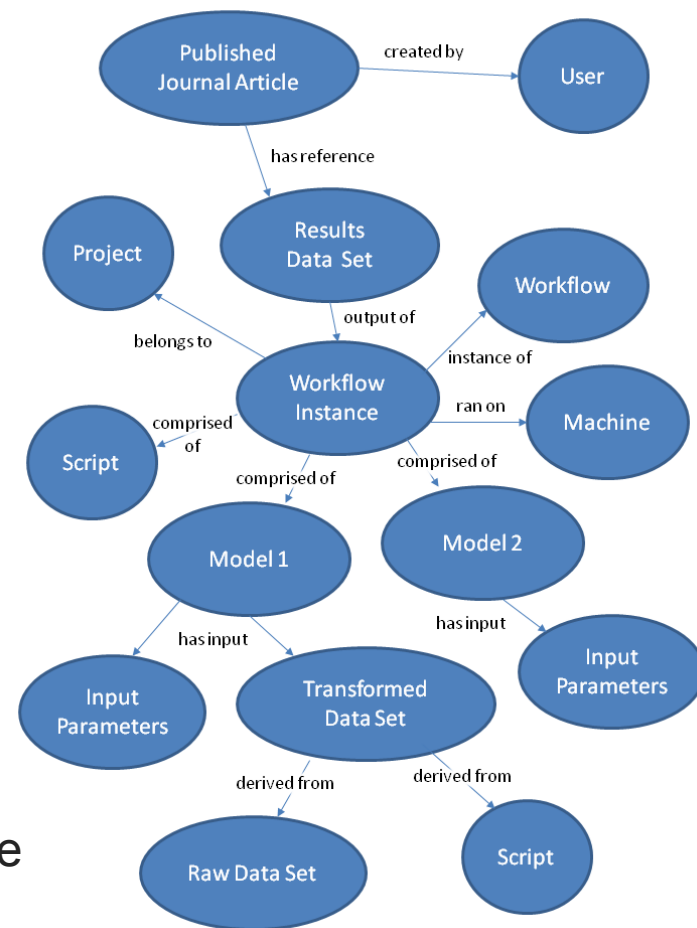
▶ Integrate Globus Online with our Velo scientific knowledge management system:

- Link Velo remote data nodes with Globus Online cloud (could we integrate via CMIS APIs?)
- Faster transfer of remote data sets
- Faster data staging in job launching infrastructure
- Faster sync with local desktop

▶ Integrate Globus Online directly with processing pipelines

▶ Improved imaging processing:

- Remove UI components from analysis code
- Parallelize code when possible



# Questions

