

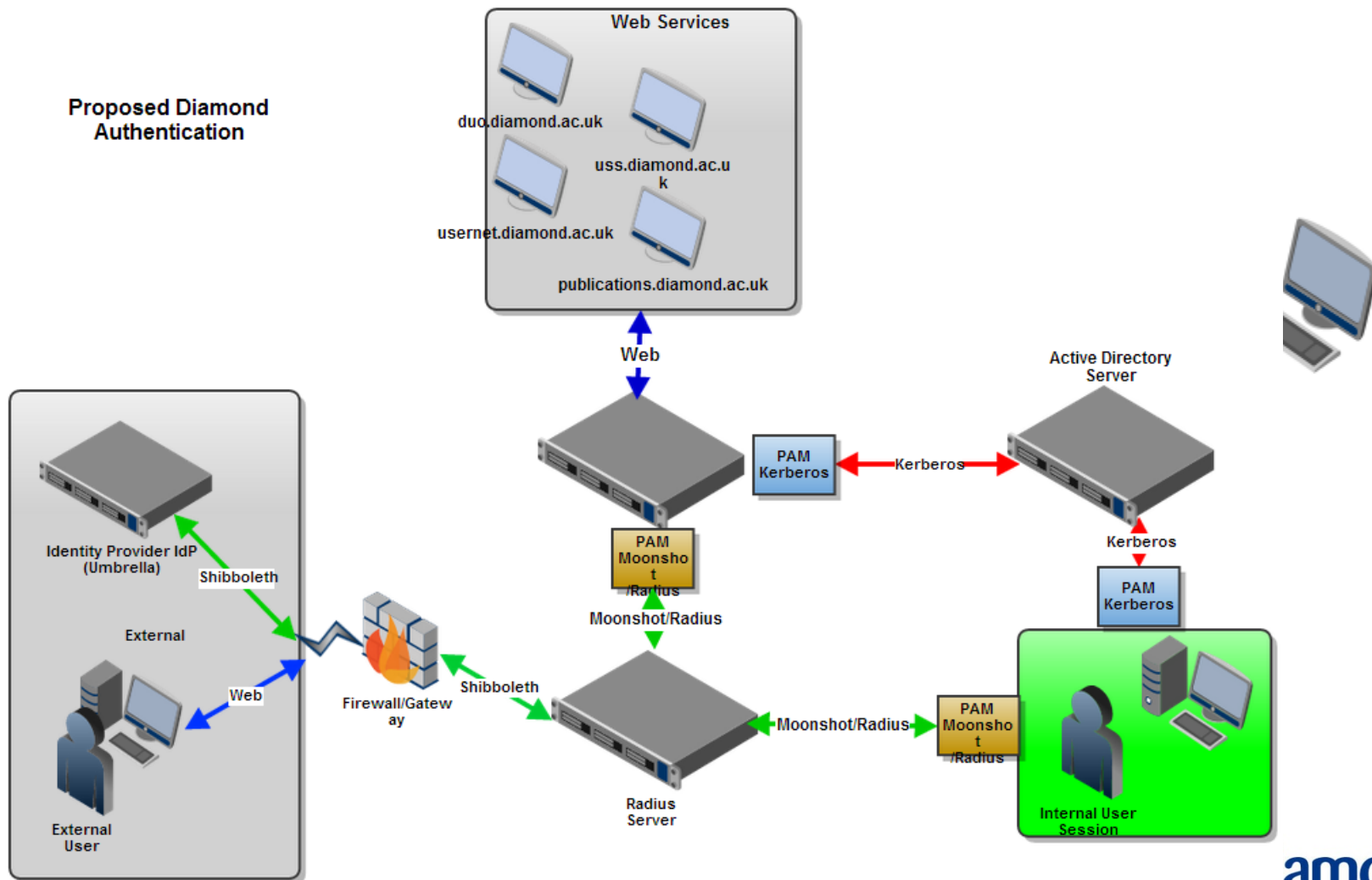
# PANData Lund 13-14 March 2013

Bill Pulford  
Head of Data Acquisition and  
Scientific Computing  
Diamond Light Source

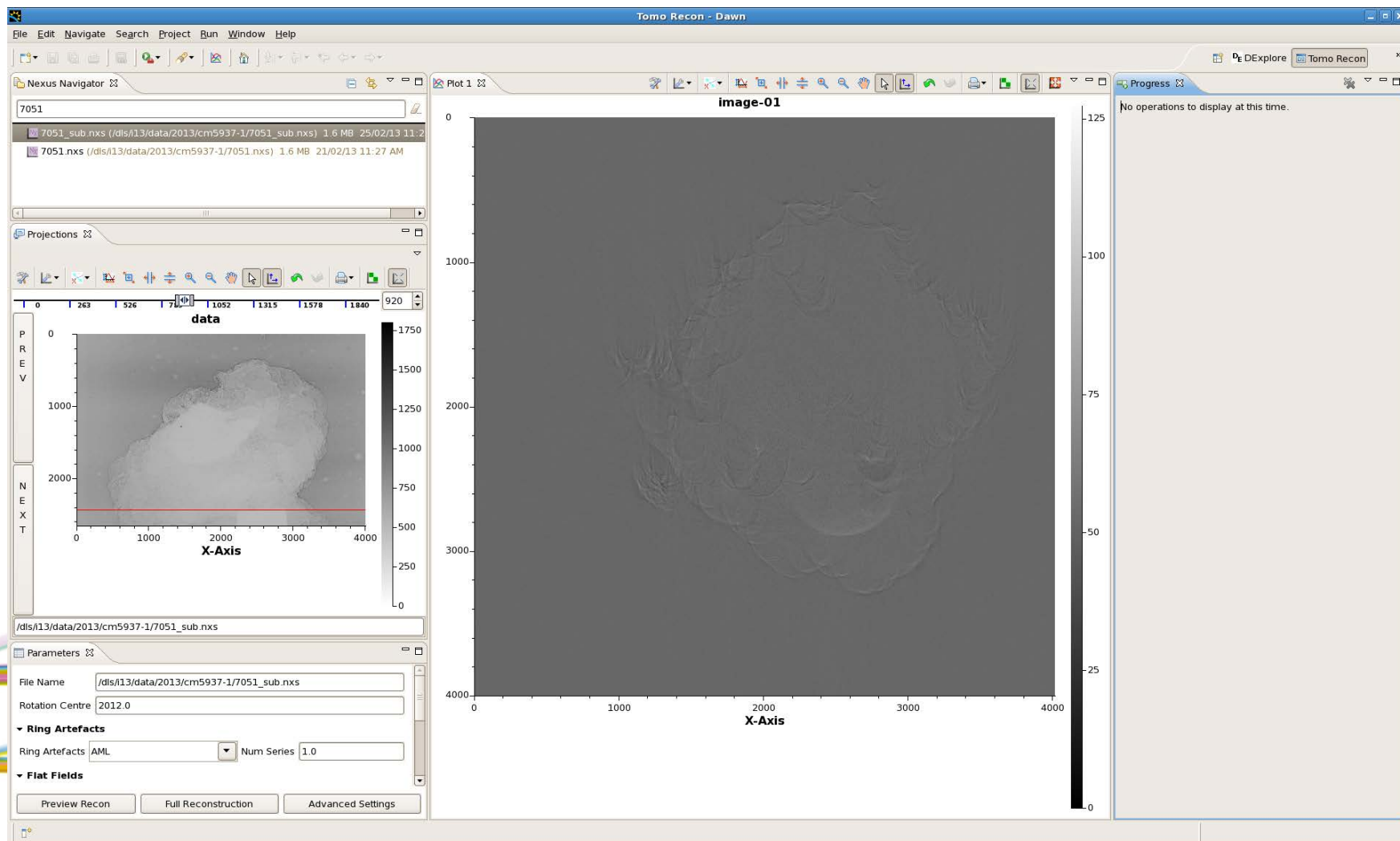
Jonathan

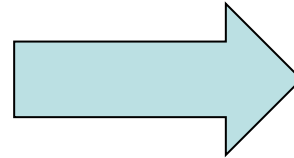
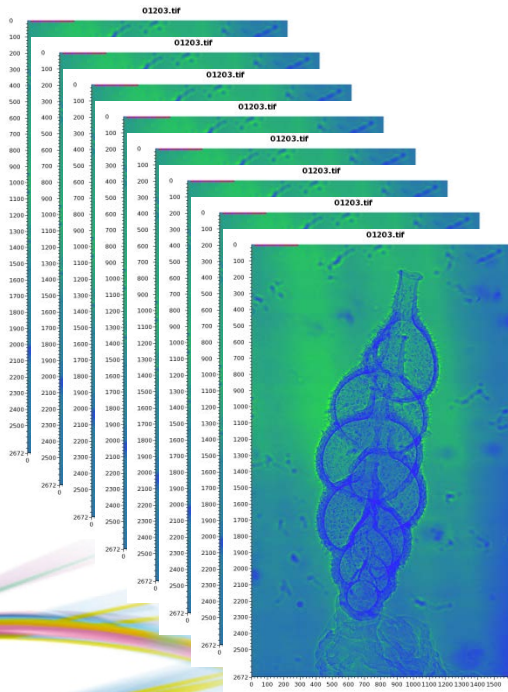
Stefan

## Proposed Diamond Authentication



# Tomography for I12 and I13





Reconstruction  
takes~20 minutes  
H5Py library used

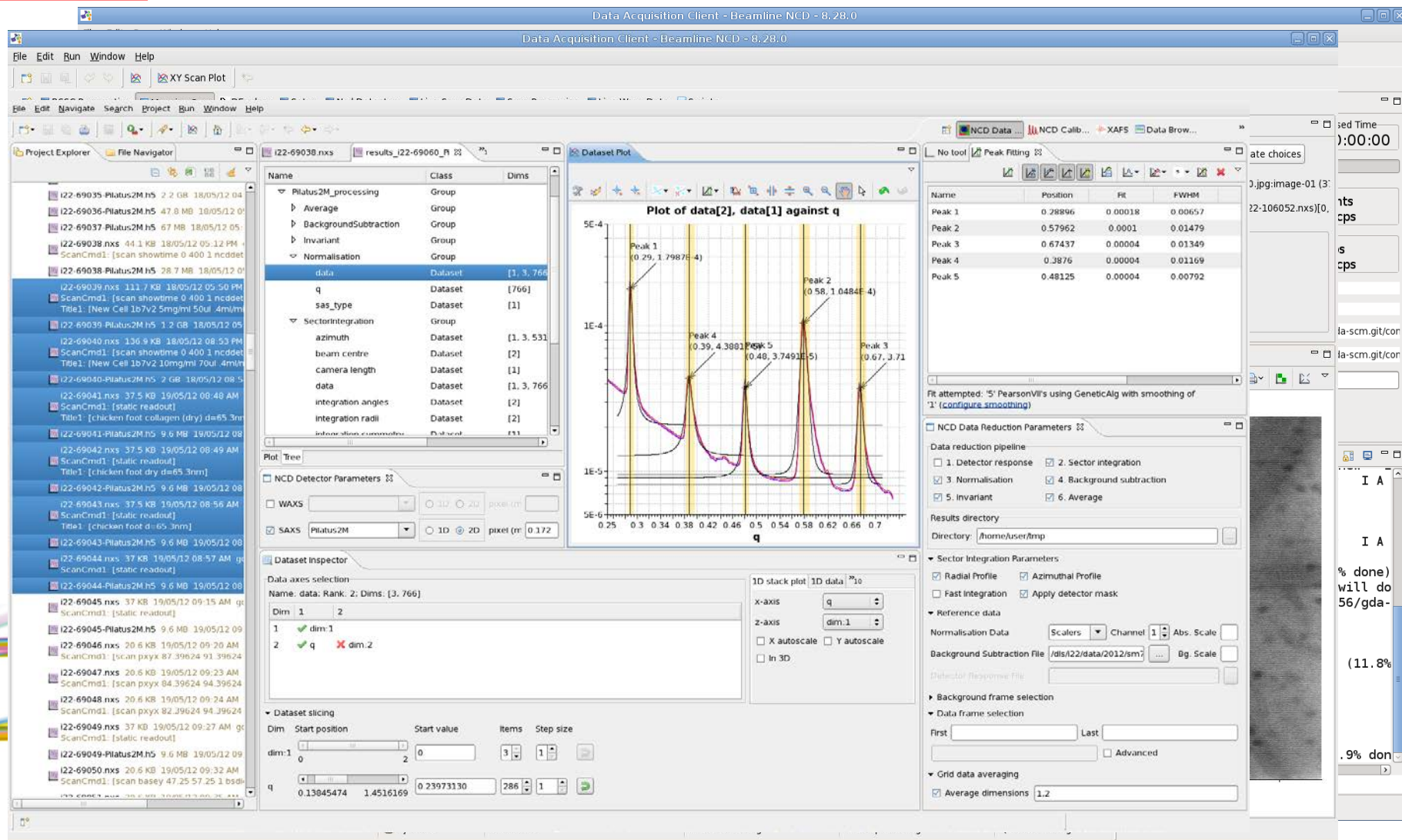


16GB

1 per 6 hours to  
2 per hour.



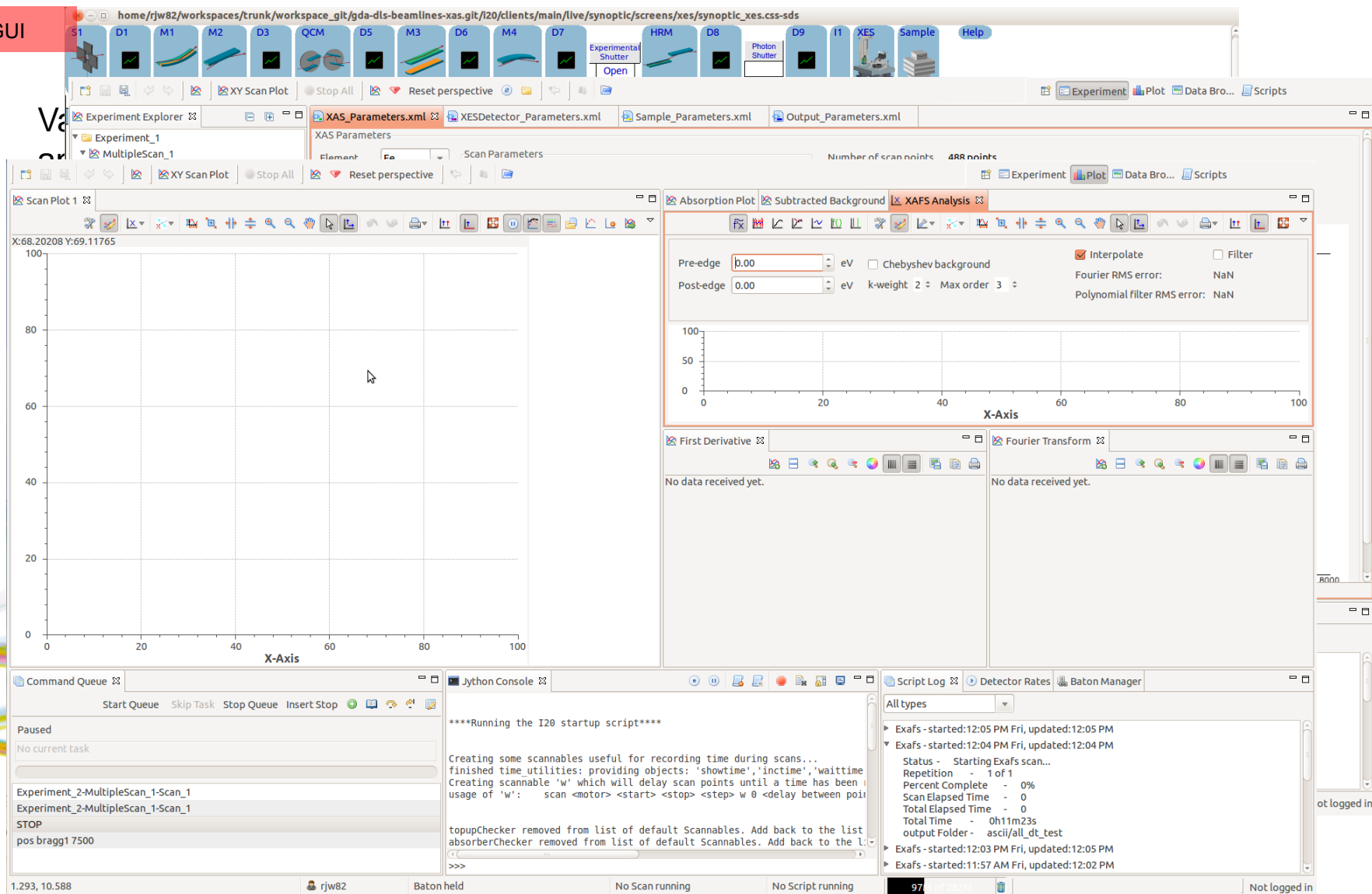
## I22 – Non Crystalline Diffraction



# Spectroscopy

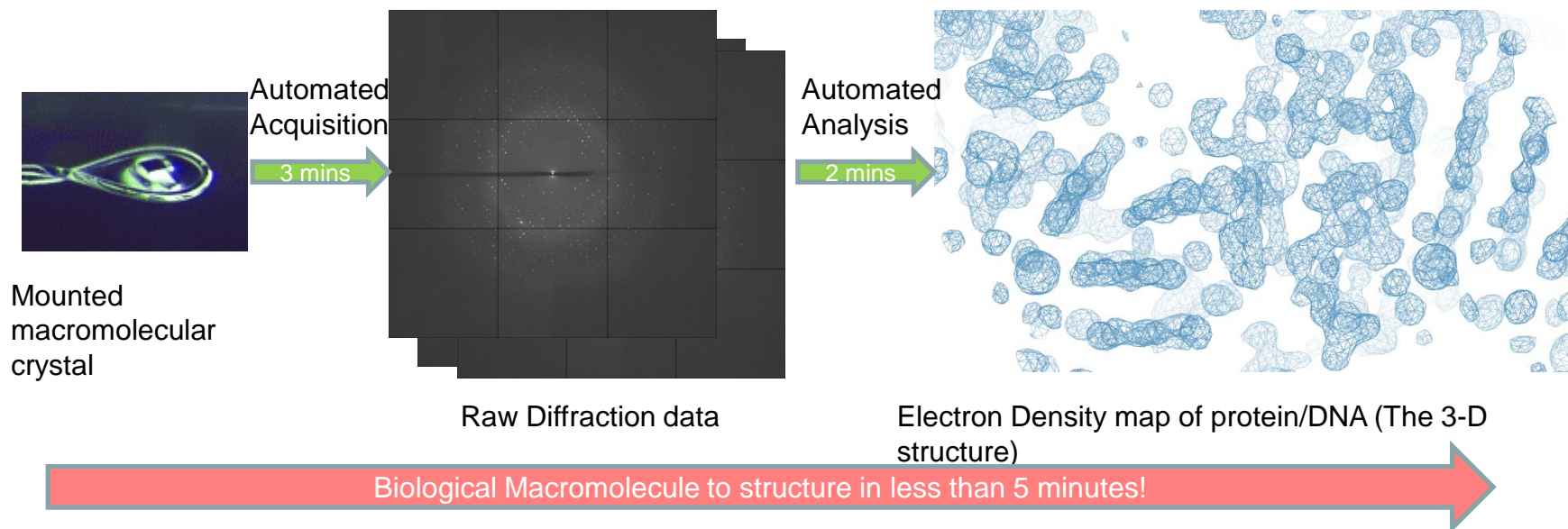
Expt. GUI

Plot GUI



This work involved an analysis of requirements for the NEXUS library as specified in report D5.1, and the current state of deployment of NEXUS at Diamond and DESY.

- a) A significant number of beamlines writing NEXUS files including NCD, Spectroscopy and Tomography . New beamlines by default write NEXUS.
- b) The Metadata derived from the experience gained from a) have been submitted to the NEXUS International Advisory Committee (NIAC) including NEXUS file links.
- c) Analysis tools now available from both Diamond and DESY that read and write NEXUS format files. Examples: DAWN science (Diamond/ESRF/EMBL), DPDAK (DESY)
- d) A coordination of metadata understanding has been reached between NEXUS metadata and imgCIF/CBF (IUCR standard) enabling possibility of scientific applications to be modified to use either format. An example Diamond has 17 beamlines that write NEXUS files and 7 that write CBF (primarily crystallography).
- e) There is a continuing dialogue between the collaborating facilities and the Nexus International Advisory Committee to add the Metadata definitions found to be necessary from experience in the scientific disciplines in a) above.



## In more detail

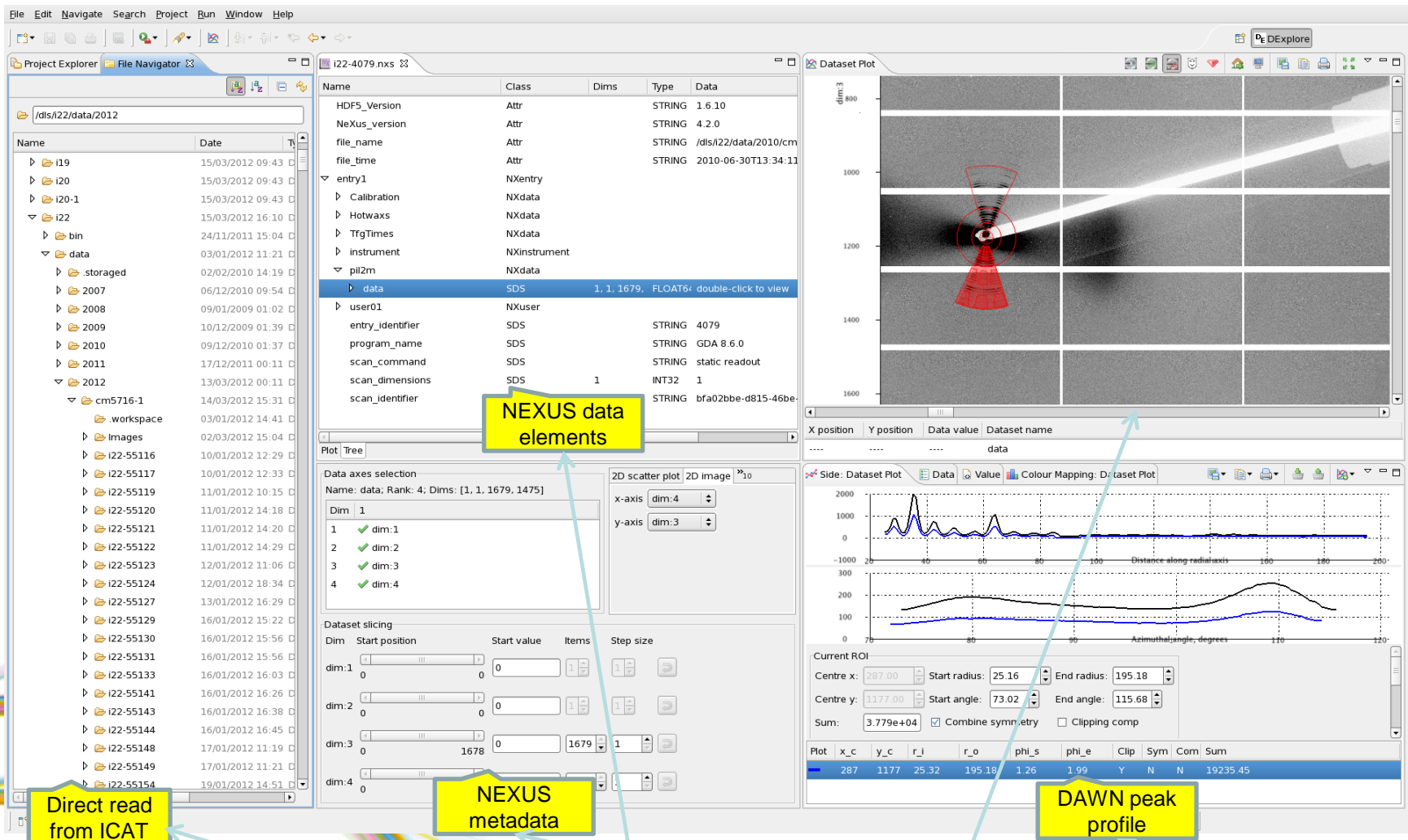
Raw diffraction data recorded in less than 3 minutes on Diamond MX beamline I02, as part of ongoing research into DNA / ligand structures, joint Ph.D. place between Diamond and Reading University (James Hall).

Automatically calculated map derived purely from experimental information, requiring a search over a number of parameters. Results were obtained less than two minutes after the data collection was completed, a step which would once have taken hours!

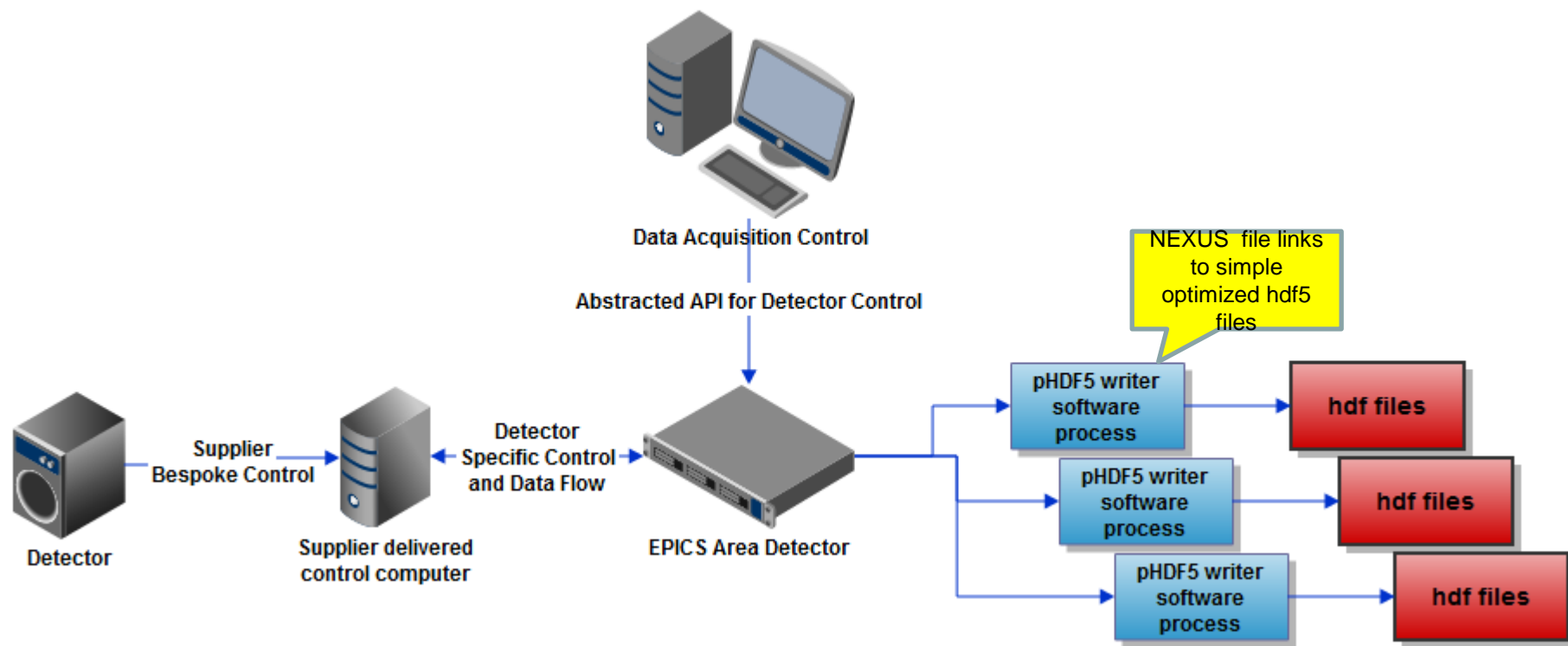
In both the processing and phase calculation extensive use was made of parallel processing, making use of one to two hours processing time in under two minutes.

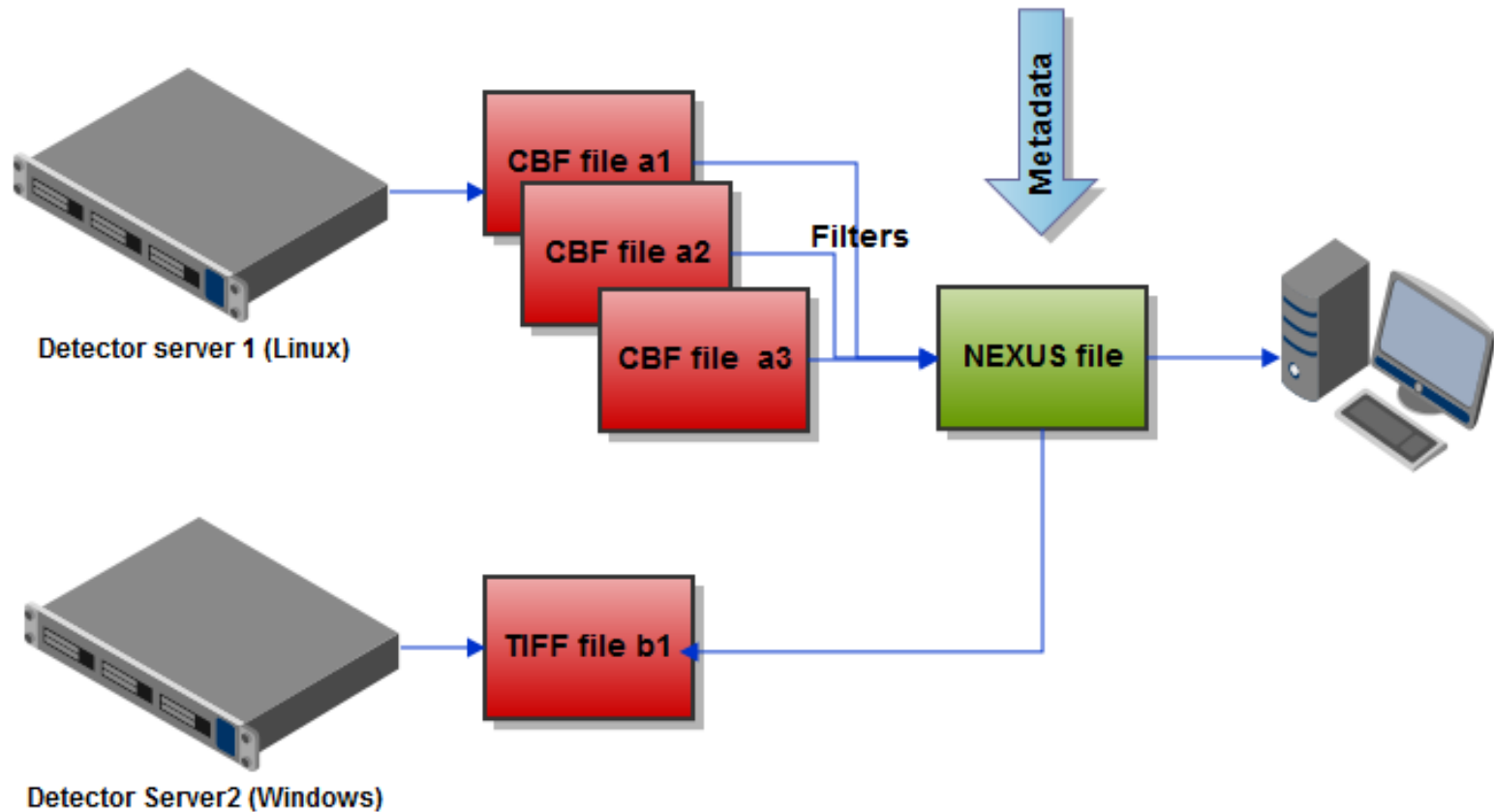


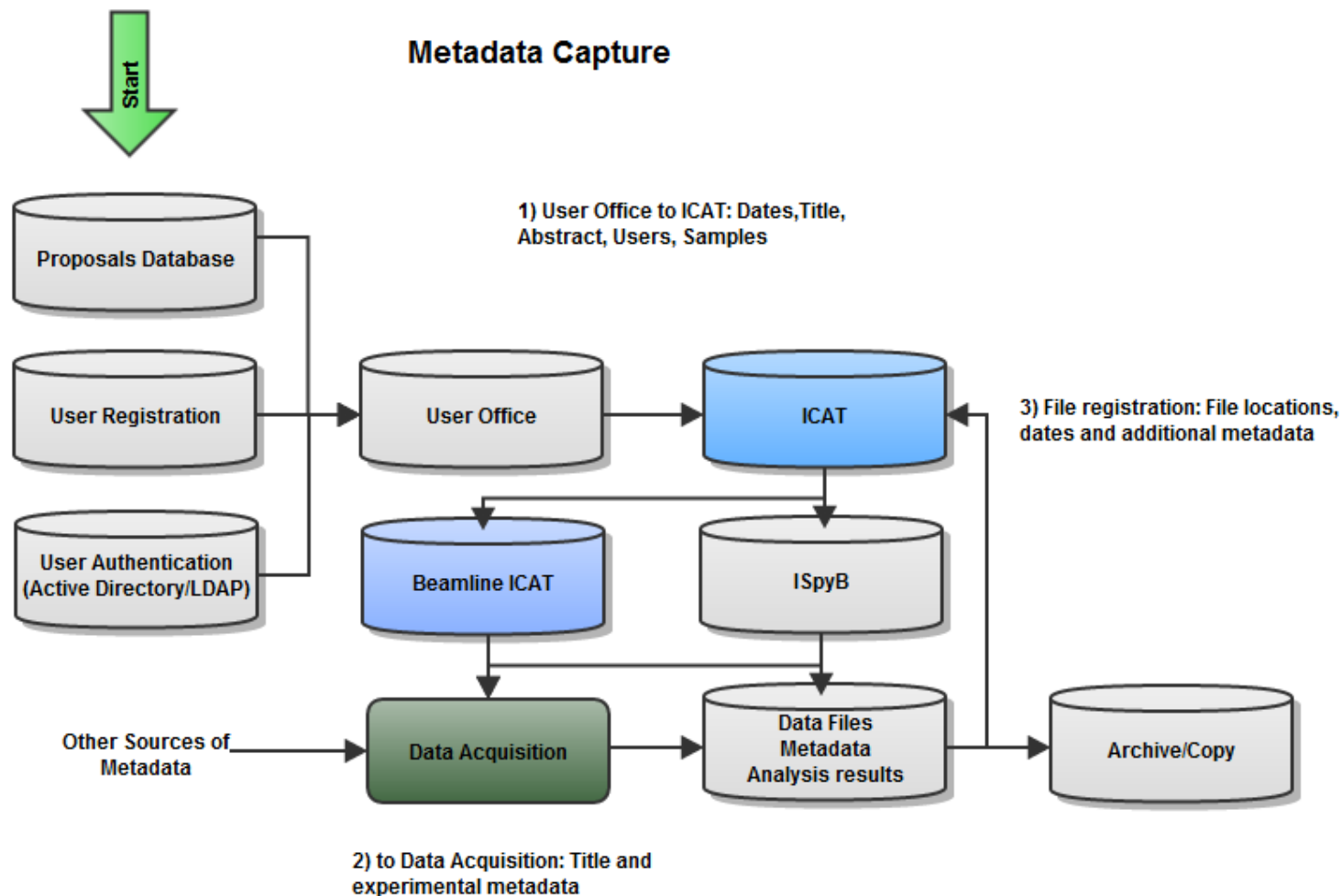
# parNEXUS enabled software DAWN Science Data Analysis suite.janet



DAWN Integrating archived data, descriptive metadata, image views and analysis in a consistent and coordinated view for Non Crystalline Diffraction



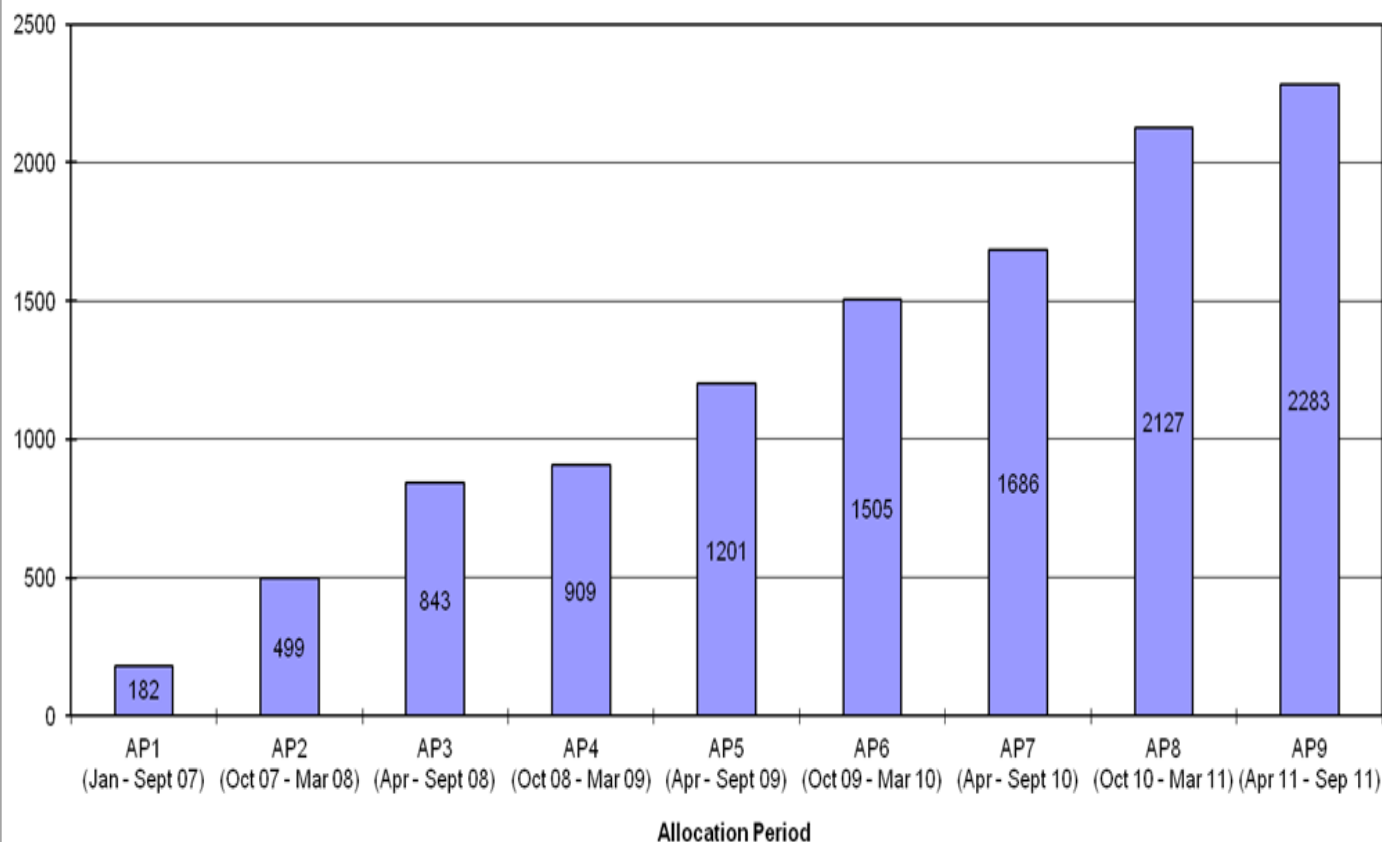






Most automated processes require Single Sign On and increasingly transparent access to remote resources.

In the year between 1-Jan-2011 and 1-Jan-2012 we had 1726 experimental visits and 4976 external experimenters of whom 1976 where unique. In Sep 2012 we had Data Volumes and Numbers of files (ICAT) ~ 300Tb/135,000,000



- Implement Single Writer Multiple Reader (SWMR) is being developed at the moment with the requirement that readers can see and act on a data file being written somewhere else without any other message path between processes. This is important for sites that need to read and write in parallel
- Scalable to support future detectors with increasing data rates
  - pHDF5 run as MPI jobs so scale easily to run as multiple processes
- Focus on optimizing performance of the parallel file system
  - Optimize for underlying parallel file system performance
  - Use best practices in the use of HDF5 libraries
  - Optimize file writing software with respect to the file system (e.g. automatic or user configurable data chunk sizes and boundaries)
  - Store N-dimensional datasets and 2-3 Dimensions per frame + M scan dimensions
- Implement and use extendible datasets
  - Grow the dataset as new data arrives
  - Handle dropped data frames by filling in a blank frame
- Fix HDF5 bugs

## Current Status

- Single process HDF5 writer already available
  - Part of the EPICS areaDetector distribution since version 1-7
  - Runs natively on Linux and Windows as an areaDetector plug-in
  - Performance is reasonable for a number of detectors in use at DLS
- Parallel HDF5 file writer application is continuing development
  - Initial design and some development done
  - Prototype is running but still lacking some features
- Integration of Windows based detector system
  - TCP transfer of N dimensional arrays with attached metadata
  - Currently using a prototype implementation of the new EPICS CA4 protocol

- Each of the three partners active in WP8 continues to provide significant resources working on the implementation of the prototypes that will form D8.3 and then contribute to future deliverables.
- In addition industrial concerns such as Dectris, who supply very high data rate detectors, have been included in discussions to gain any effort that they can provide and ensure as smooth a rollout of the technology as possible.
- A Single Write Multiple Reader (SWMR) is now at the advanced planning state with the HDF5 group.



# WP8 Timeline

D8.1: Definition of pHDF5 capable Nexus implementation (Month 9) – Software  
– **Report Delivered Aug 2012**

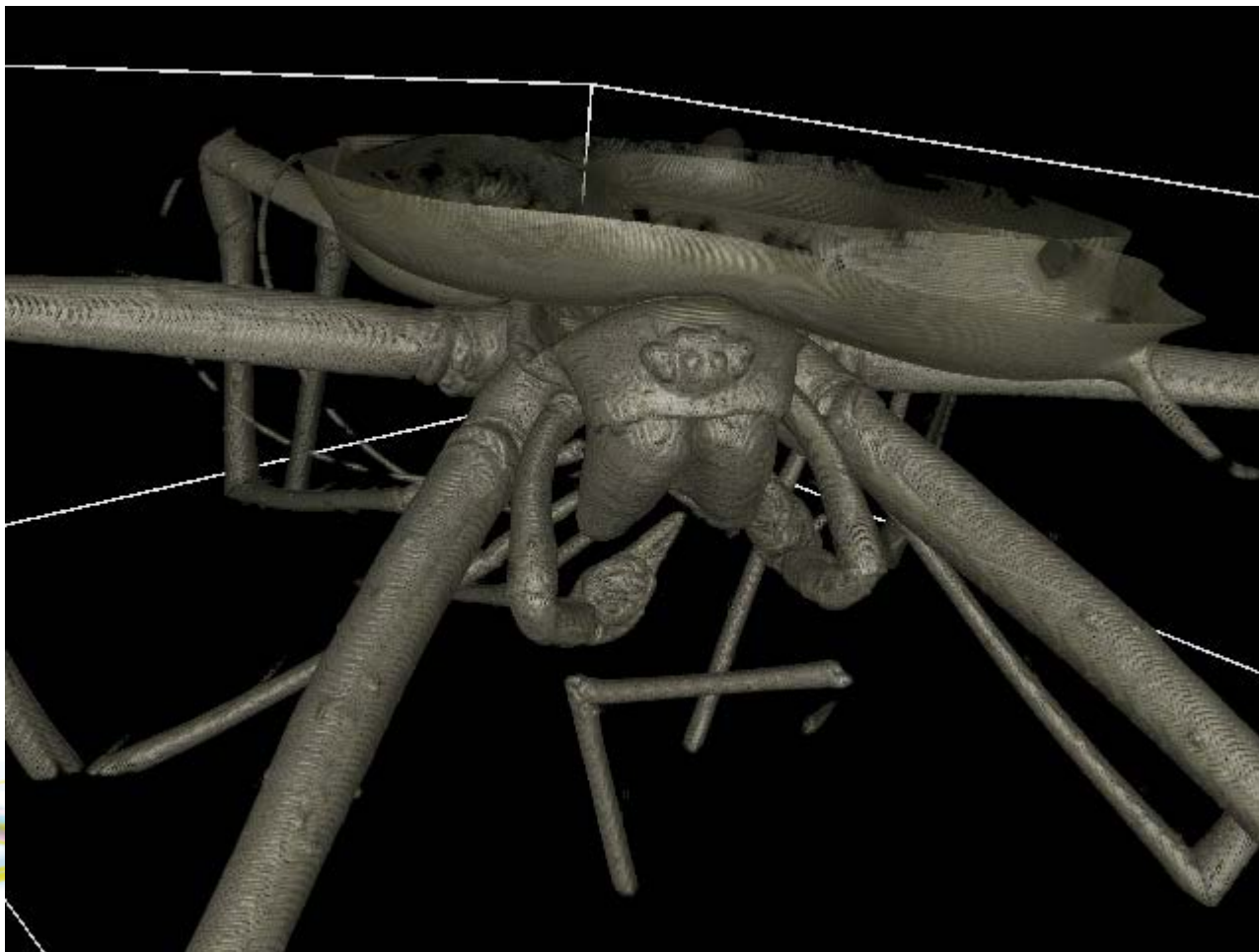
D8.2: Evaluation of Parallel file systems and MPI I/O implementations (Month 9)  
– **Report Delivered Aug 2012**

D8.3: Implementation of pNexus and MPI I/O on parallel file systems (Month 21 June 2013) - Prototype - **Currently in use in March 2013 but needs to be packaged.**

D8.5: Examination of Distributed parallel file system (Month 21 June 2013) - Report - **Constant evolution and a snapshot could be produced but actual value is in question.**

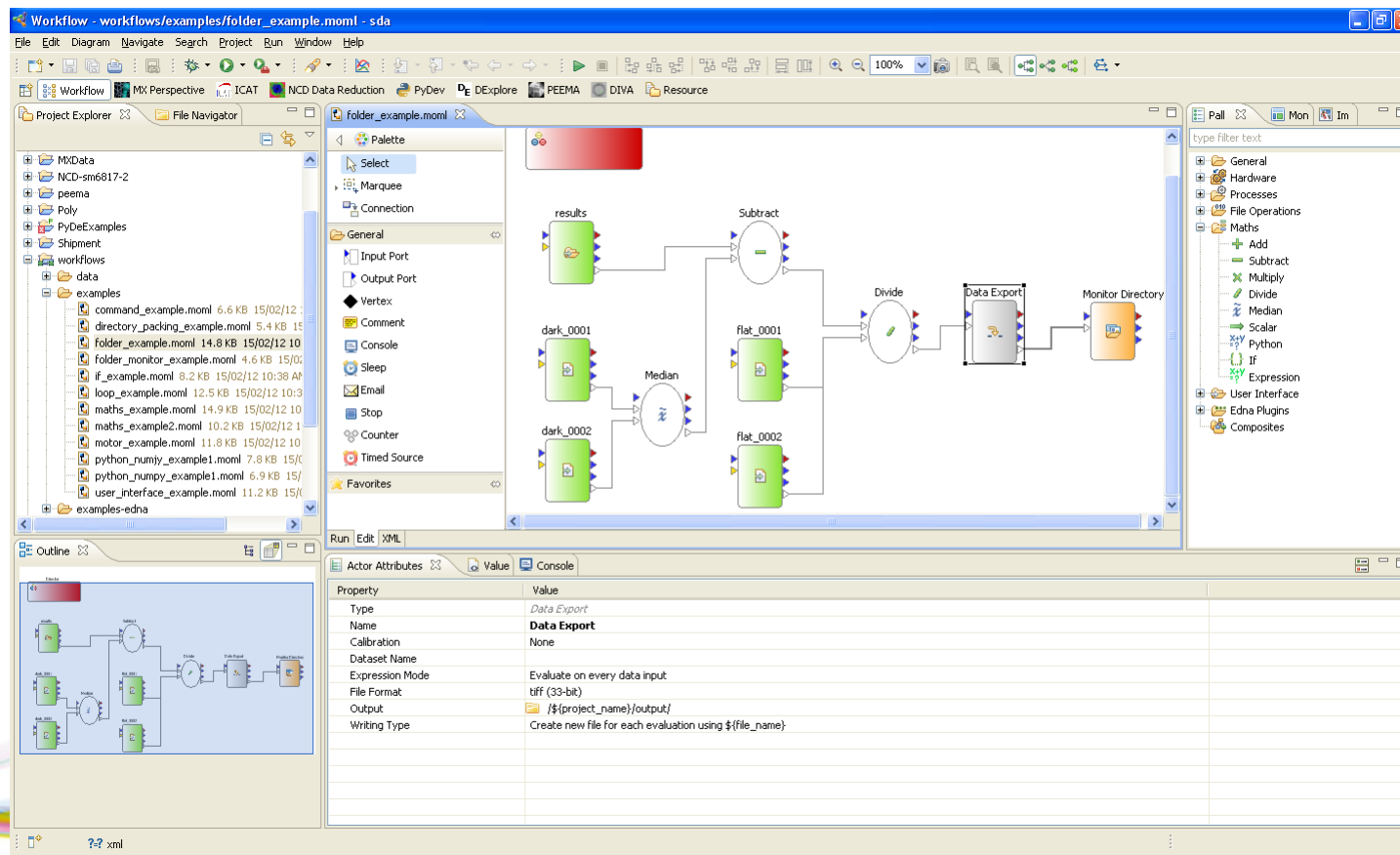
D8.6: Demonstrate capabilities on selected applications (Month 21 June 2013) - Demonstrator - **Already done in March 2013.**

D8.7: Evaluation of coupling of prototype to multi-core architectures (Month 30 March 2014) - Report - **Much work already done by community**



- Thank You





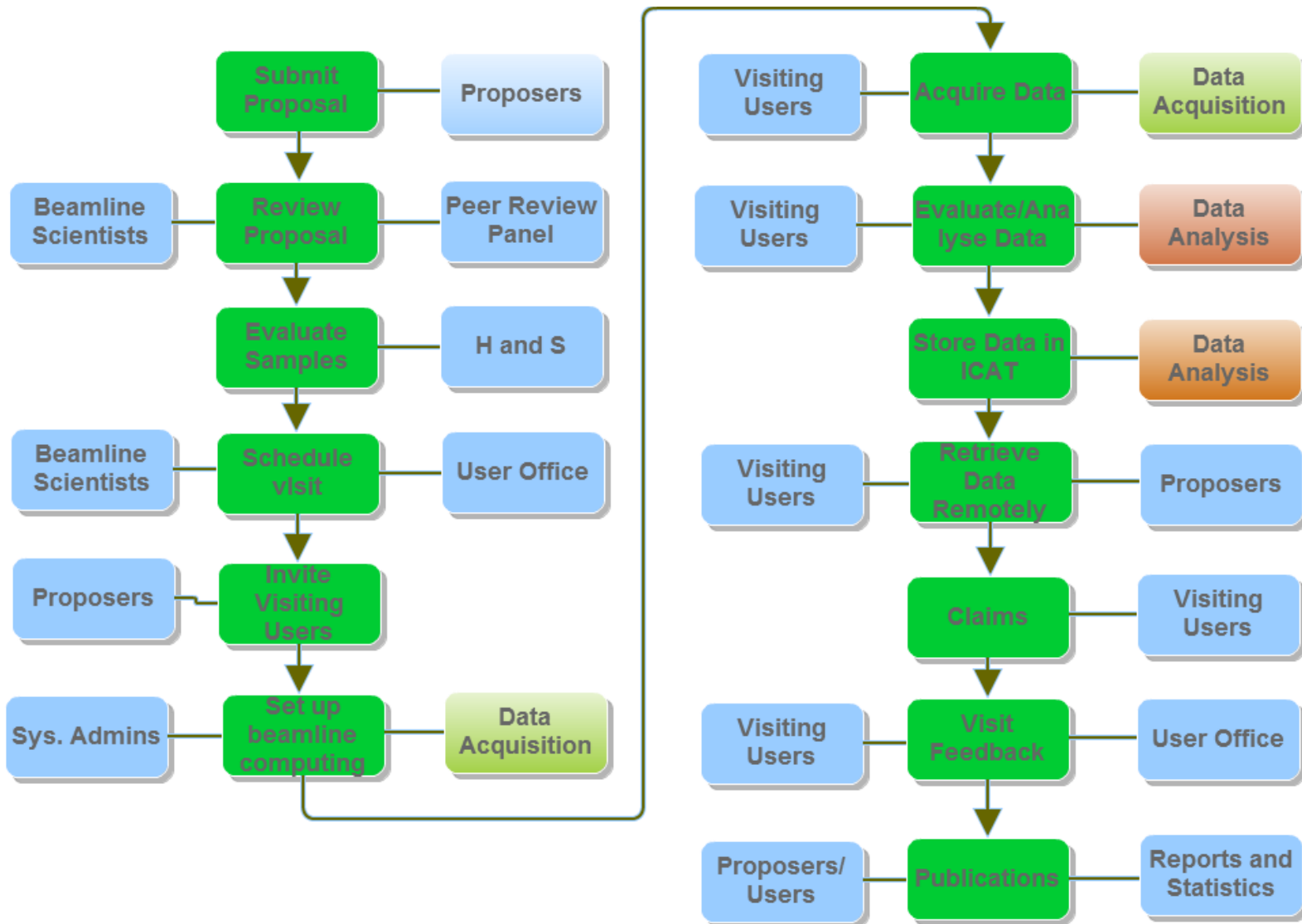
This enables data processing sequences to be run automatically and transparently either locally or remotely (or both) and repetitively; the processes can be moderated according to the operation of each step.

Each processing step can include almost any application (CCP4, FEFF etc) that allows intelligible input and output.

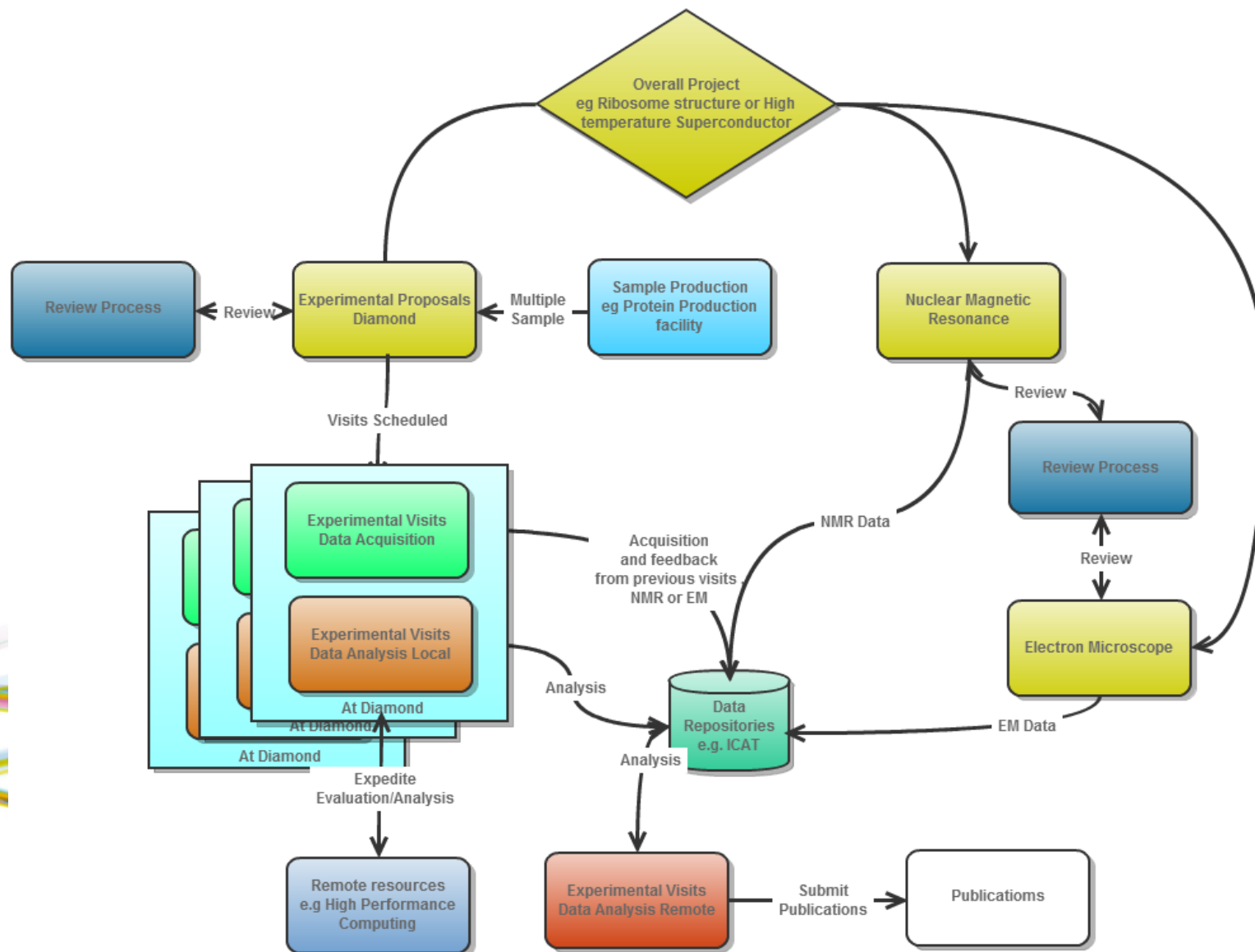


- Thank You





# An example overall future project hierarchy



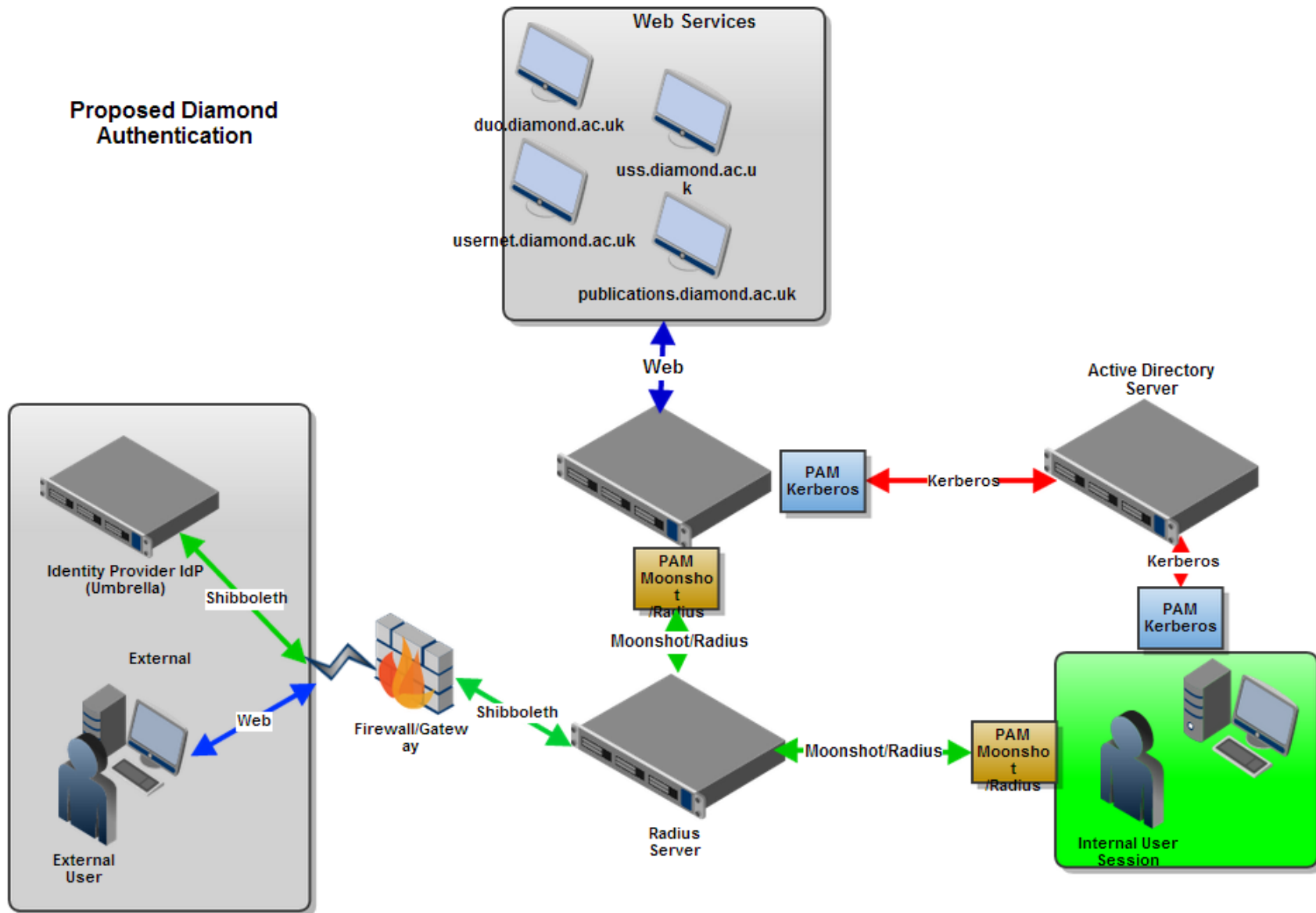
1. An authentication mechanism based on Umbrella should be implemented as widely as possible across participating facilities.
2. A technical solution needs to be provided to allow interactive sessions on computing resources at the facilities' sites. Initial coverage must include Linux and Windows (optionally Macintosh) systems in order for any complete adoption of the authentication system.

Project MOONShot provides this functionality

1. An important architectural advantage is offered by using a system such as the Jasig Central Authentication Service (CAS) where most of the internal Authentication and Authorization issues are covered by this one system thereby obviating the need to modify dependent systems.



## Proposed Diamond Authentication



- Current systems can be difficult to define the duration for which users can be authenticated.
  - a. X509 certificates may be useful to help with this problem
- A mechanism of persistence or caching is necessary to avoid loss of service due to network unavailability.
- PAM stacks. User X509 certificates
- Must support Windows, Macintosh and Linux
- Should be very simple to install and configure (consider rpm or msi)
- The overall network infrastructure should provide the necessary support.

Interesting:

- Compatible with any Microsoft initiative if possible.
- EduRoam
- Mobile devices

