PaNdata Open Workshop & Photon and Neutron Science Interest Group at the RDA 3rd plenary in Dublin 26-28.03.2014

PaNdata ODI 1st Open Workshop

Dublin 24-25/28th of March 2014

Co-located with the RDA 3rd plenary at Croke Park

https://indico.desy.de/event/1stow

Booklet of collected presentations

PART 3: PaNSIG

PaNdata Open Workshop & Photon and Neutron Science Interest Group at the RDA 3rd plenary in Dublin 26-28.03.2014

Photon and Neutron Science Interest Group meeting

26th and 28th of March at Croke Park / RDA

Thomas Proffen/ Amber Boehnlein	Photon and Neutron Science and facilities - the computing challenges
Tom Griffin:	Data Management solutions – ICAT
Brian Matthews:	Data Management solutions – Tardis
Brian Matthews/ Steve Androulakis	Potential areas for future collaboration and development of common standards and services within RDA
Adam Farquhar:	DataCite
Alun Ashton:	Data analysis issues and frameworks
Eugen Wintersberger:	HDF5
Brian Matthews:	PaNdata
PaNSIG chairs:	Developing a plan of activities for PaNSig

pandata ODI

Computing Challenges for Photon and Neutron Facilities

Driving Factors: Computing

•Meeting the current science goals for major initiatives such as energy research and materials discovery require improvements in computational tools and techniques.

•Science is driving source upgrades

- Brighter and more precise sources drive detector development
- Detector development drives computing needs—volume, and complexity

•Science also directly drives the computing needs

- Simulations
- Data Analysis and analytics

Facility Challenges

- O(10000) users
  - Diversity of science, needs, skills, longevity
- O(100) beam lines
- O(10) different imaging techniques
- Operational constraints
- Can Do/Make Do Culture
- Source and detector upgrades
  - Challenging in themselves;
  - Environment is changing

Context

It is relatively easy to make a list of common areas of interest for data and computing

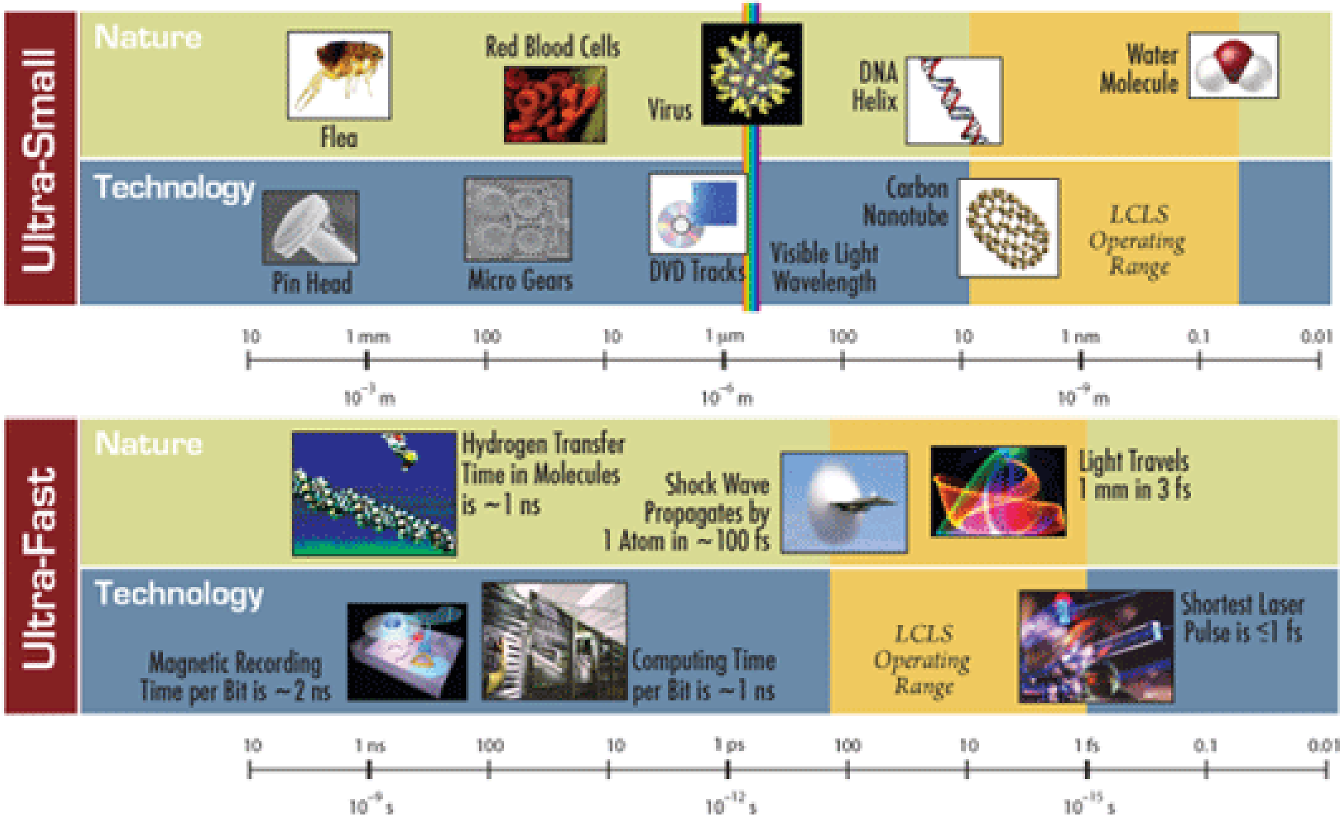
Can we collaborate across facilities on those areas productively?

With limited resources, what are good investments?

How do we plan and make the case?

Who, what, where, when and how?

A New Kind of Laser - The LCLS creates X-ray pulses that can capture images of atoms and molecules in motion



LCLS Data Rates & Volumes

.

Vary for different instruments/experiments

Three operation modes (X-Ray pulse rate):

30 Hz (first runs on LCLS, AMO/CAMP, Oct-Dec 2009)
60 HZ (second series of runs, May 2010 -> on)
120 Hz (Fall 2010)

Instruments:

Fall 2009: AMO/CAMP
Summer 2010: AMO/CAMP, XPP, SXR
Fall 2012: all 6+1 instruments (station #2 for CXI)

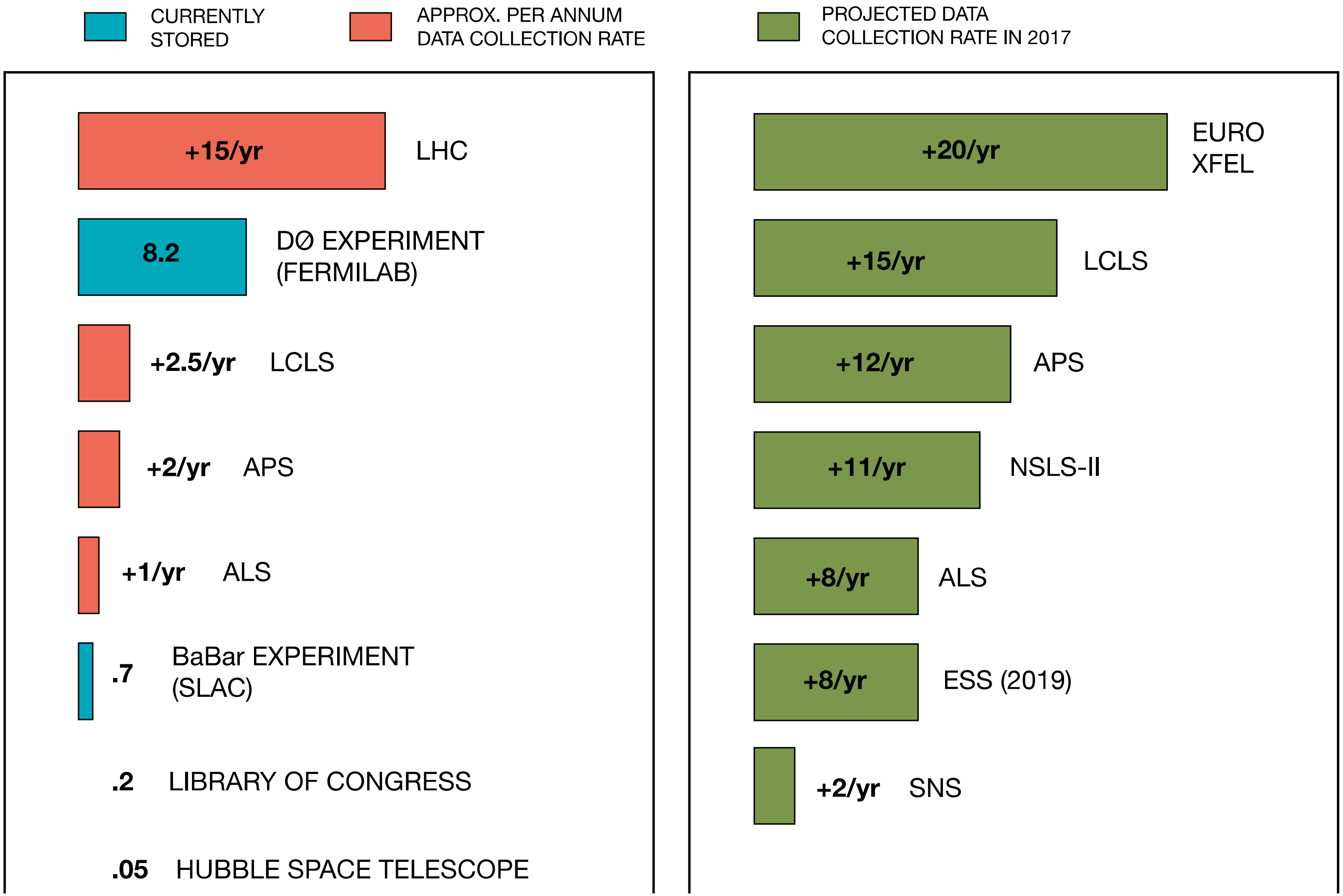
First runs at AMO/CAMP (30 Hz):

up to **180 MB/s**, **3 TB/day**, **~100 TB** of raw data recorded

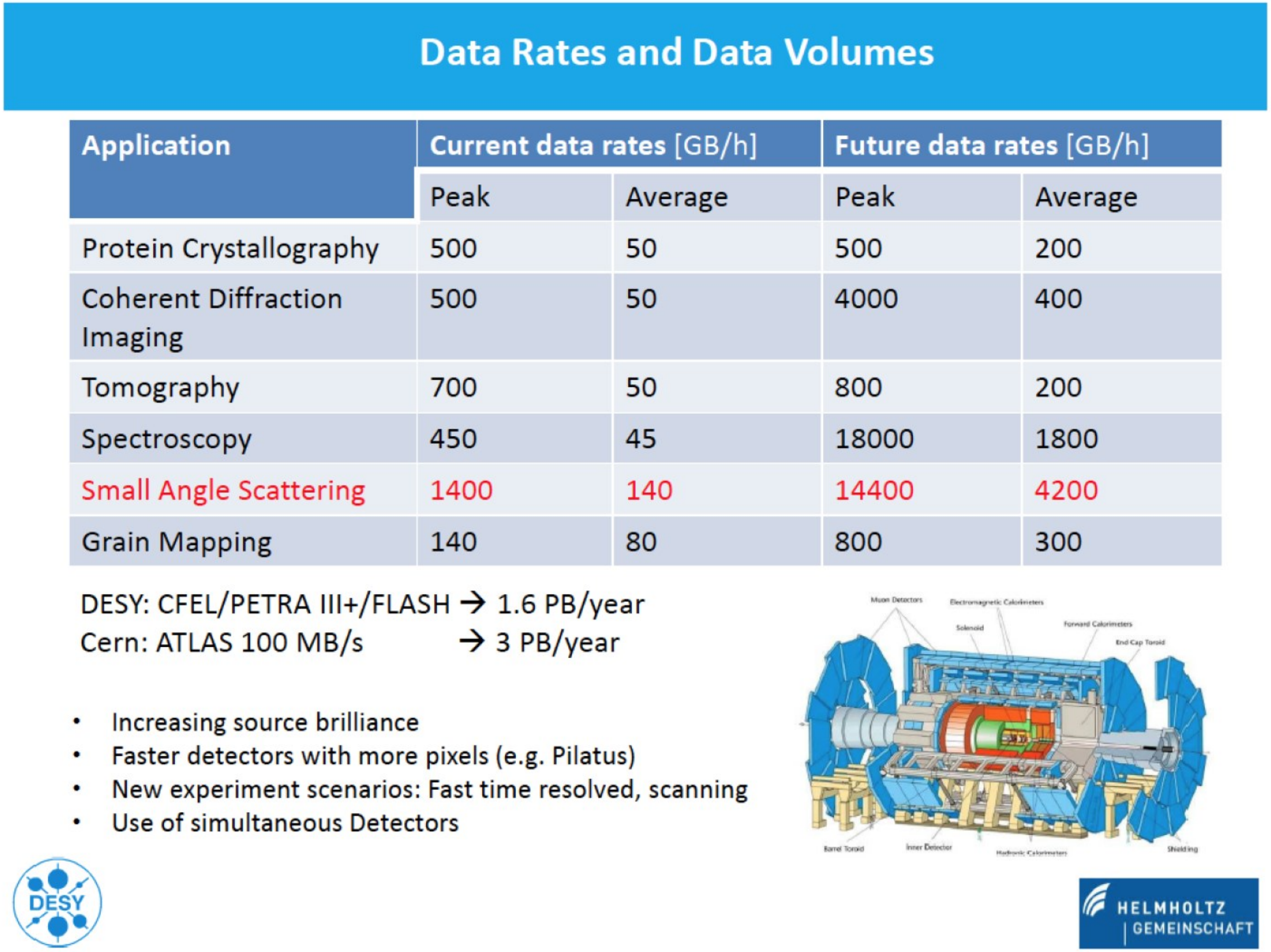
At full capacity (120 Hz):

Up to 1.5 GB/s for CXI (event size exceeds 10 MB)
30 TB/shift
Up to 3 instruments can take data in parallel (approaching this condition**)**
1-2 PB of raw data per year

The Data Challenge view from 2012



Data Volumes Growing everywhere



Data Rates and Data Volumes

Application	Current data rates [GB/h]		Future data rates [GB/h]	
	Peak	Average	Peak	Average
Protein Crystallography	500	50	500	200
Coherent Diffraction Imaging	500	50	4000	400
Tomography	700	50	800	200
Spectroscopy	450	45	18000	1800
Small Angle Scattering	1400	140	14400	4200
Grain Mapping	140	80	800	300

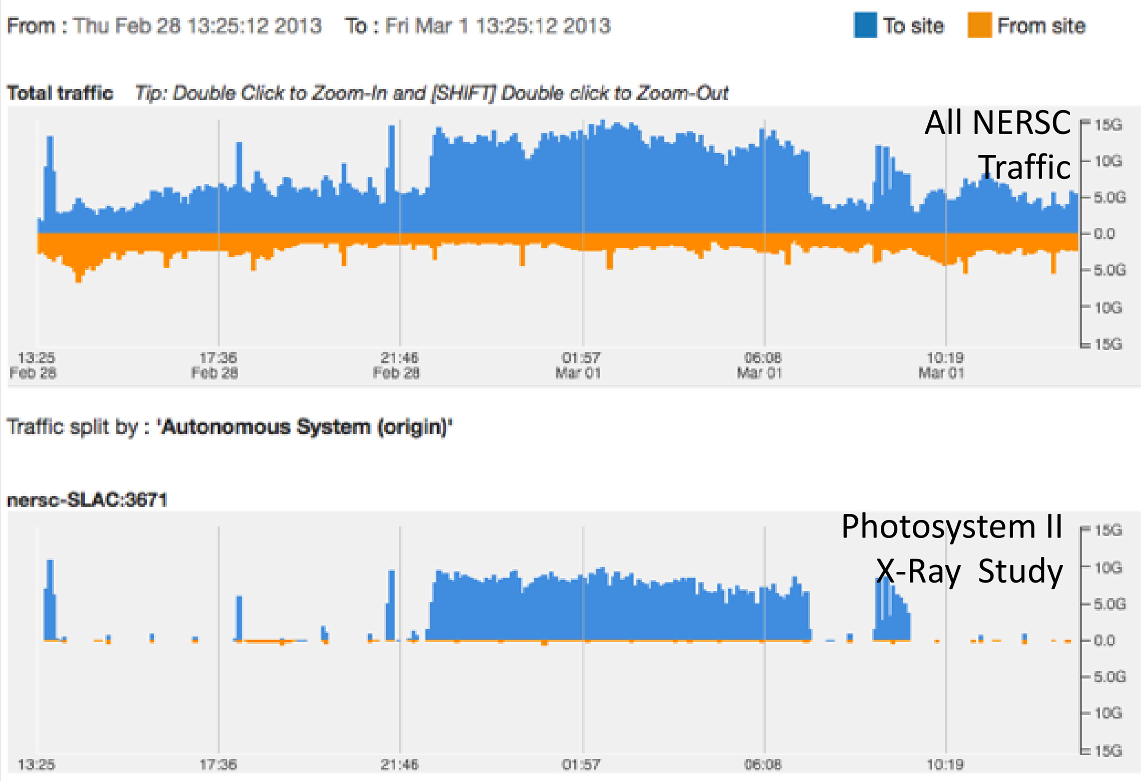
DESY: CFEL/PETRA III+/FLASH → 1.6 PB/year
Cern: ATLAS 100 MB/s → 3 PB/year

- Increasing source brilliance
- Faster detectors with more pixels (e.g. Pilatus)
- New experiment scenarios: Fast time resolved, scanning
- Use of simultaneous Detectors

LCLS Diffract & Destroy

Photo-system II, Nick Sauter (LBL): A single experiment in 2013 generated **120TB** over four days. Data was relayed to NERSC via Esnet, analysis required **135K CPU hours** of computing.

LCLS-II will require > **100M CPU hours** per experiment. Higher resolution and advanced image analysis could grow computational complexity. Some algorithms scale an M*NlogN for M images of N pixels.



From : Thu Feb 28 13:25:12 2013 To : Fri Mar 1 13:25:12 2013

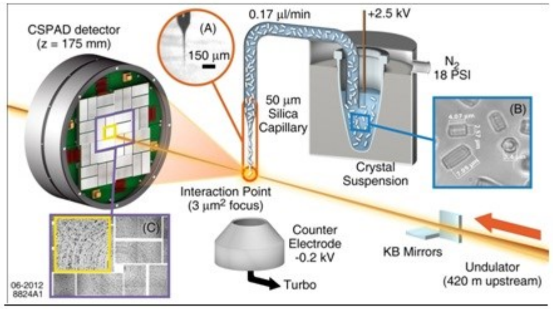
Total traffic *Tip: Double Click to Zoom-In and [SHIFT] Double click to Zoom-Out*

Traffic split by : **'Autonomous System (origin)'**

nersc-SLAC:3671

All NERSC Traffic

Photosystem II X-Ray Study



A starter wish list…

Easy to use data management and processing frameworks that scale with data rate

Conditions data storage and management; meta data

Data validation and fast feedback at collection

Data volume reduction/compression techniques

Development and availability of algorithm and algorithm tools

Community developed simulations and simulation tool kits for beamlines and detectors

Computational science

Compute and storage hardware platforms appropriate for the task**S**

Data management

- Life is easier if the data is managed from point of origin
- Detector readouts can have proprietary readouts

Is standardization possible?
Data containers/formats
Metadata
Workflows/tool kits
Visualization tools
Curation/cataloging

Naively, some standardization
would make good use of resources and expertise,
simplify life, open up many possibilities

Rapid feedback

Time is money

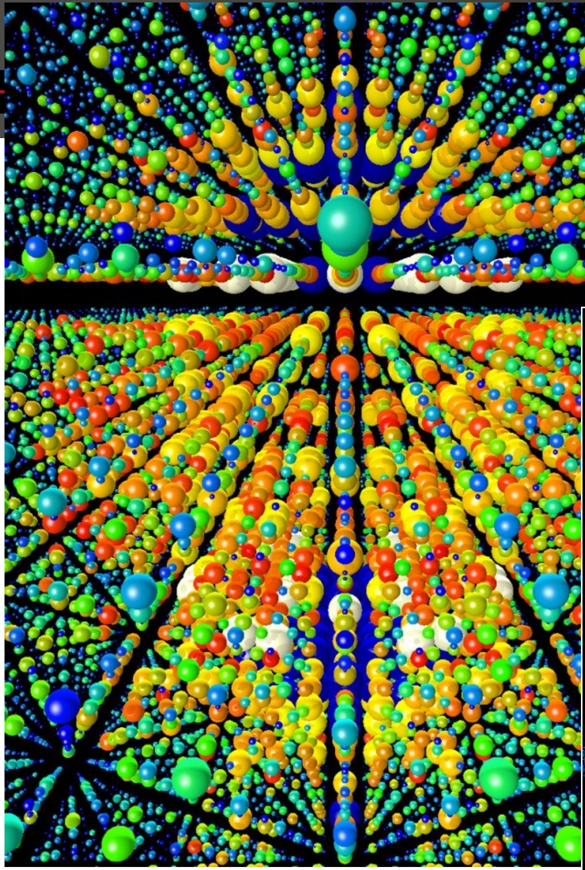
    Beam time, experimenter time; instrument scientist time

Experiment simulations

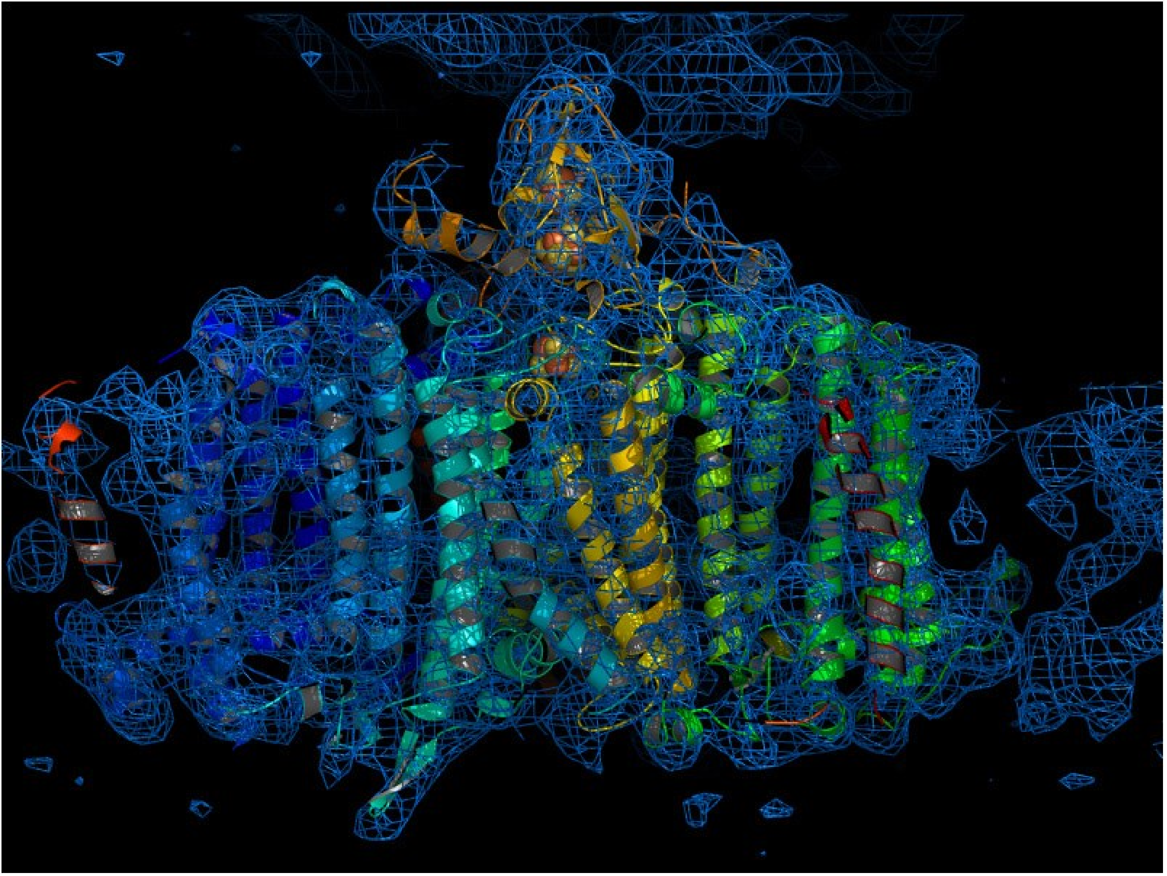
Advance preparation of analysis tools

Could provide better use of beam time, shorten time to publication and refine where to spend resources on computing improvements http://home.slac.stanford.edu/pressreleases/2011/20110202.htm

A reconstructed image of t[illegible]osystem I complex. (Image courtesy Raimund Fromme, Arizona State University.)



Three-dimensional rendering of the X-ray diffraction pattern for the Photosystem I protein, reconstructed from more than 15,000 single nanocrystal snapshots taken at the LCLS. (Image courtesy Thomas White, DESY.)



High Throughput XRD for Material Discovery

Strategy:

- Ternary and Quaternary Combinatorial Libraries

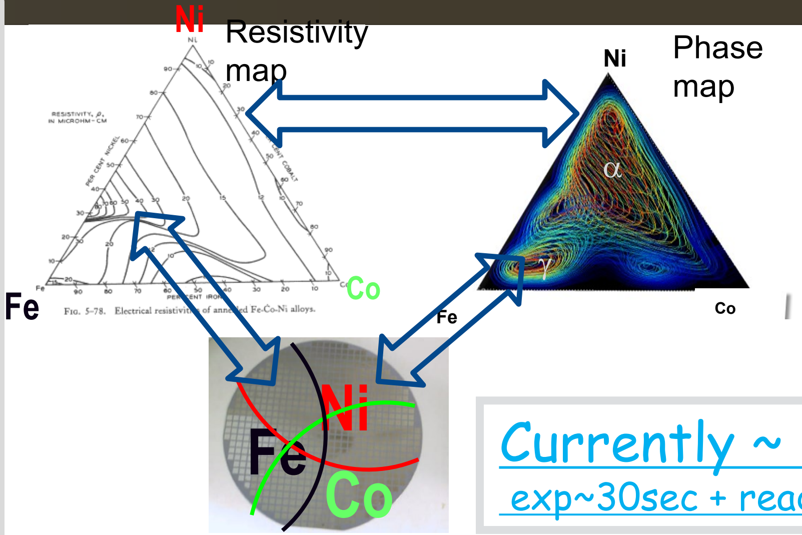
- Screen for Properties
  - HT Spectroscopy
  - New Materials

- Determine Structural Phase Diagram
  - HT XRD, HT EXAFS Sam Webb
  - New Phase Diagrams

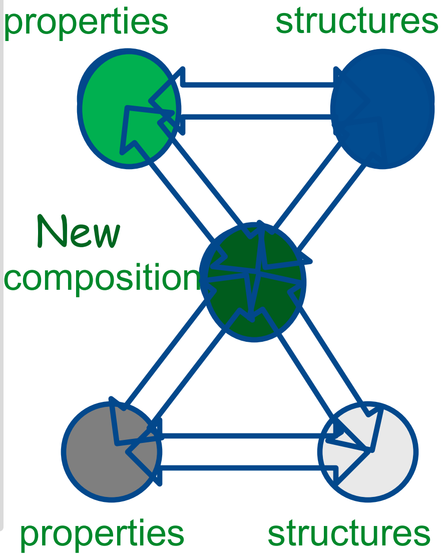
- Composition-Structure-Property Relations

- Phase Transitions
  - Metastable & Near Equilibrium

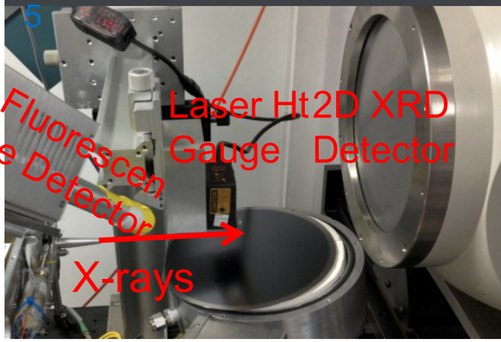
- Input for Atomistic Material Modeling
  - New Theory for (Metastable) Materials



Ni
Resistivity map
Fe
Co
Fig. 5-78. Electrical resistivity of annealed Fe-Co-Ni alloys.
RESISTIVITY, ρ, IN MICROHM-CM
PER CENT NICKEL
PER CENT COBALT
PER CENT IRON
Ni
Phase map
α
γ
Fe
Co
Ni
Fe
Co



Prototype setup @ SSRL BL 1-5



Currently ~ 2000xrd+mca/day

exp~30sec + readout ~10sec

Moving Forward with HT-XRD :

Enhance Capabilities:

- Auto Alignment and Calibrations
- Low Z Detection Sensitivity
- Rapid Annealing

IncreaseThroughput:

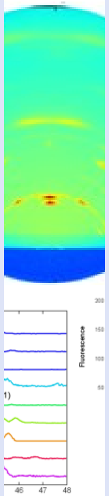
- New X-ray Focusing Optics: exp ~1-5sec
- New 2D detector: readout <1 sec
  - → 20-50,000 xrd+mca/day
  - Need Automation and Robotics

Data Management:

High-Throughput Pipeline:

Robotics, Automation & Machine Learning

John Gregoire
Ichiro Takeuchi
Matt Kramer
Apurva Mehta



Hubs

Hubs

ction

SSRL

New Composition-Structure-Property Relations

$_{40}$Ga$_{60}$

New Materials

Scientific Communities

New Theory of (metastable) Materials

tabase

CD

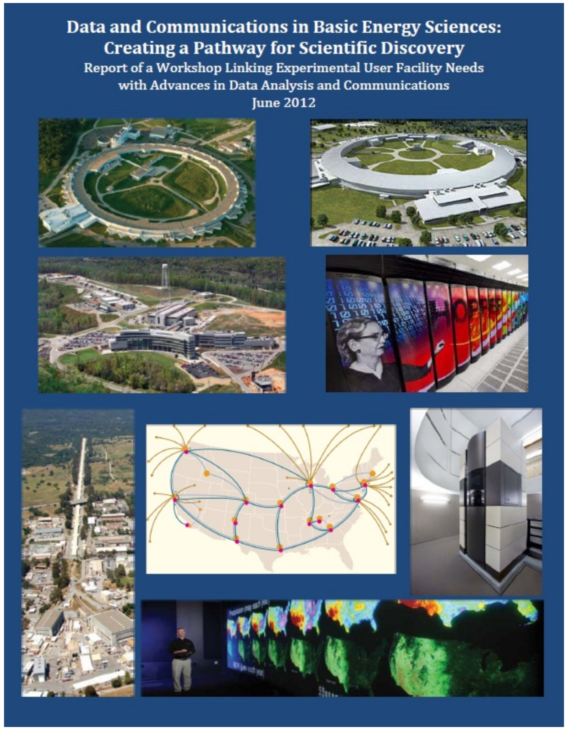
COMPUTING CHALLENGES FOR PHOTON AND NEUTRON FACILITIES

Thomas Proffen, ORNL

Amber Boehnlein, SLAC

Creating a Pathway for Scientific Discovery

- Accelerating discovery in materials science
- Enhancing predictive capabilities



Data and Communications in Basic Energy Sciences: Creating a Pathway for Scientific Discovery

Report of a Workshop Linking Experimental User Facility Needs with Advances in Data Analysis and Communications

June 2012

- Theory and analysis components should be integrated seamlessly within experimental workflow.
- Move analysis closer to experiment – future possibility of experiment steering.
- Match data management access and capabilities with advancements in detectors and sources.

Driving Factors: Computing

- Meeting the current science goals for major initiatives such as energy research and materials discovery require improvements in computational tools and techniques.

- Science is driving source upgrades
  - Brighter and more precise sources drive detector development
  - Detector development drives computing needs—volume, and complexity

- Science also directly drives the computing needs
  - Simulations
  - Data Analysis and analytics

Facility Challenges

- O(10000) users
  - Diversity of science, needs, skills, longevity
- O(100) beam lines
- O(10) different imaging techniques
- Operational constraints
- Can Do/Make Do Culture
- Source and detector upgrades
  - Challenging in themselves;
  - Environment is changing

Context

- It is relatively easy to make a list of common areas of interest for data and computing
- Can we collaborate across facilities on those areas productively?
- With limited resources, what are good investments?
- How do we plan and make the case?
- Who, what, where, when and how?
-

Neutron Data Life Cycle

Feedback

DAS

- Neutron events
- Events from sample environment
- Other triggers

Reduction

- Corrected reduced data (histograms, S(Q,E), ..)
- Merging, reconstruction of data
- Instrument/technique dependent
- Need for ’real’ time reduction

Analysis

- Multi dimensional fitting
- Advanced visualization
- Comparison to simulation / feedback
- Field dependent, large variety of approaches

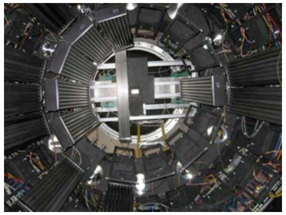
Simulation Modeling

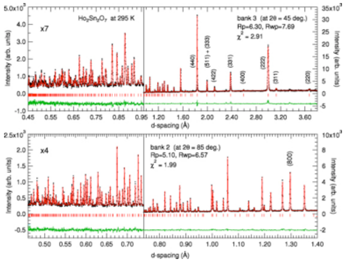
- Multitude of techniques (DFT, MD, ..)
- Advanced simulation of experiments
- ‘Refinement’ using experimental data
- Multiple experiments / probes

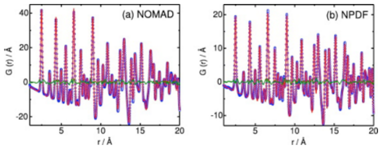
User Facility

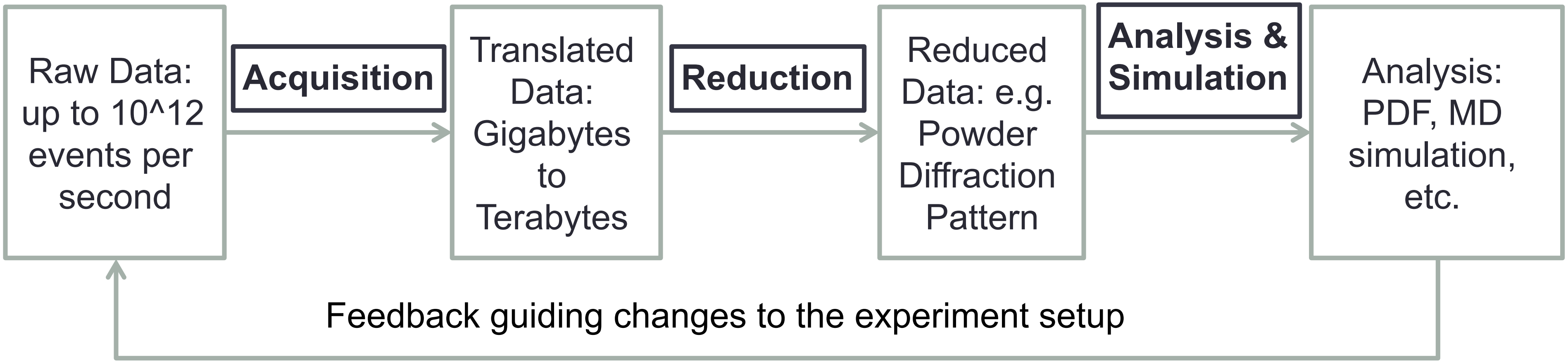
- Variety of experiments, topics, methods and ‘computer literacy’ of users are significant challenge.

Example: NOMAD Diffractometer









Raw Data: up to 10^12 events per second

Acquisition

Translated Data: Gigabytes to Terabytes

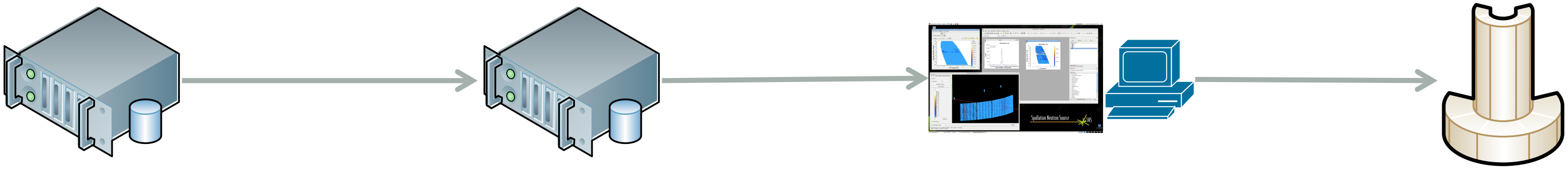
Reduction

Reduced Data: e.g. Powder Diffraction Pattern

Analysis & Simulation

Analysis: PDF, MD simulation, etc.

Feedback guiding changes to the experiment setup



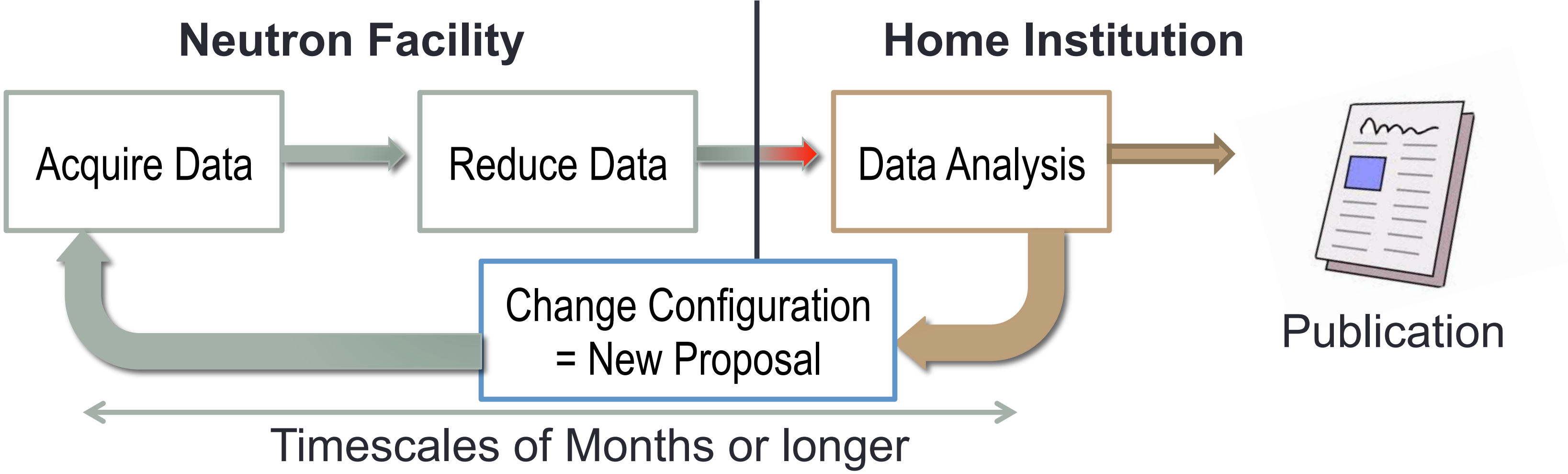
Data captured and stored on multiple systems at the beamline

After completion of a "run" data is aggregated on a single system, translation begins

Once data is aggregated reduction begins using a workstation

Analysis and Simulation using mid-scale compute

Improving Productivity = Changing the Workflow



Neutron Facility

Home Institution

Acquire Data

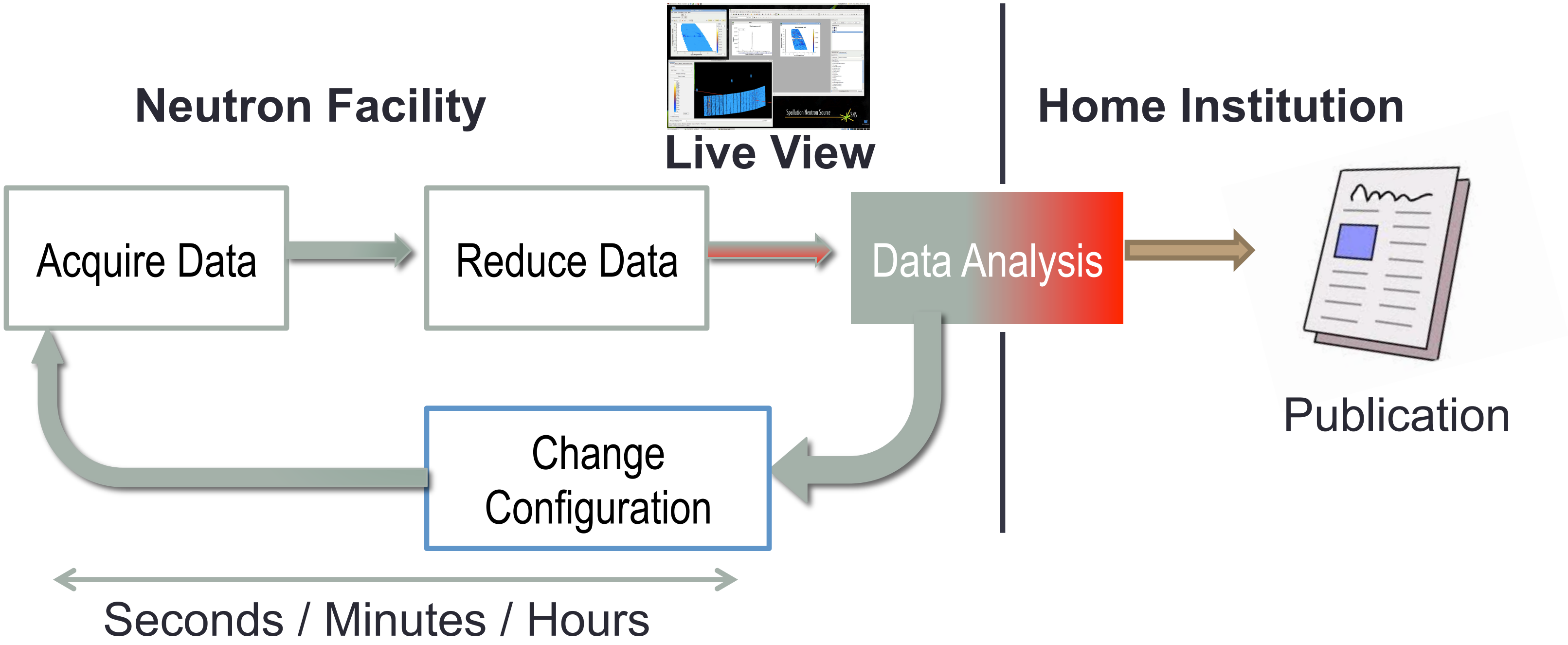
Reduce Data

Data Analysis

Publication

Change Configuration
= New Proposal

Timescales of Months or longer



Neutron Facility

Live View

Home Institution

Acquire Data

Reduce Data

Data Analysis

Publication

Change
Configuration

Seconds / Minutes / Hours

ADARA is enabling real-time feedback from experiment, analysis and computational steering

ADARA

- Leverages our multi-disciplinary capabilities at ORNL coupling Neutron Sciences Directorate with Computing and Computational Sciences Directorate.

- The ADARA Project lets us stream data to computational resources and provide live feedback from experiment in real-time S(Q,E).

- Provides a high performance data backplane for reduction, analysis, and coupling with simulation forming the basis for future work to integrate experiment and simulation.

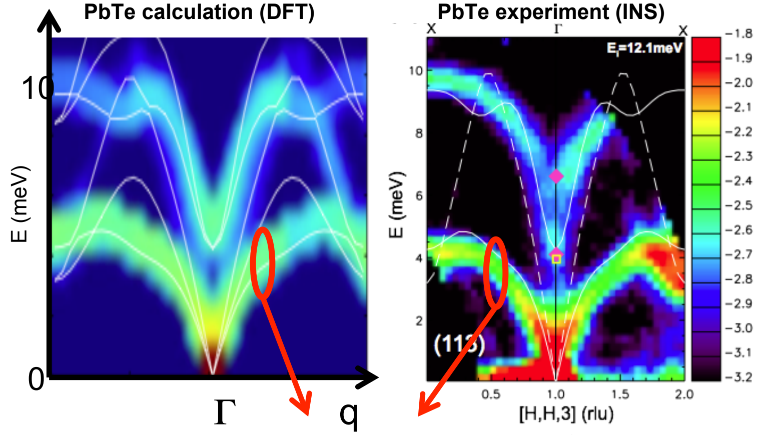
- Prototype running on HYSPEC instrument. Deployment to other beamlines in 2013/2014.

ORNL has launched the Center for Accelerating Materials Modeling (CAMM)

- The CAMM will integrate materials modeling/simulation (MD/DFT) directly into the chain for neutron scattering data analysis, **offline** and **online** (in near real time)

- Developing workflows for refinement, integration of MD codes, **neutron scattering corrections** ..

- The CAMM is working with ORNL's Materials Science and Technology Division to study coarse grained MD simulations of polymers PEO-AA (CNMS), *ab-initio* MD simulations for ferroelectrics/thermoelectrics

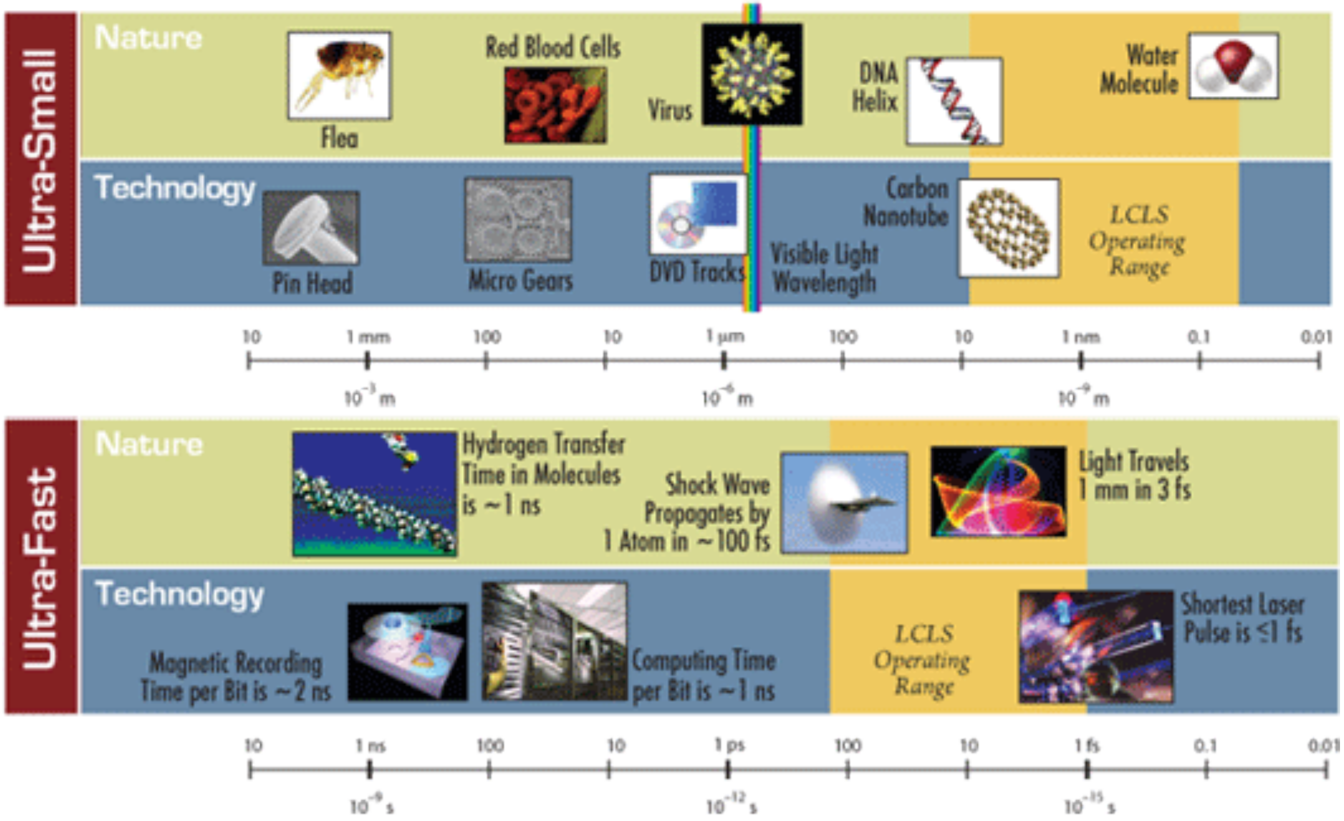


Example: *ab-initio* MD simulations for ferroelectrics/thermoelectrics. Focus on *width* of dispersions

The Center for Accelerating Materials Modeling (CAMM)

- *Partnership between ORNL's Neutron Sciences, Physical Sciences and Computing and Computational Sciences Directorates*
- *ORNL SEED money and DOE funds provided to study force field refinement from quasi-elastic and inelastic neutron scattering data*
- *CAMM formed in response to BES proposal call for* Predictive Theory and Modeling

A New Kind of Laser - The LCLS creates X-ray pulses that can capture images of atoms and molecules in motion

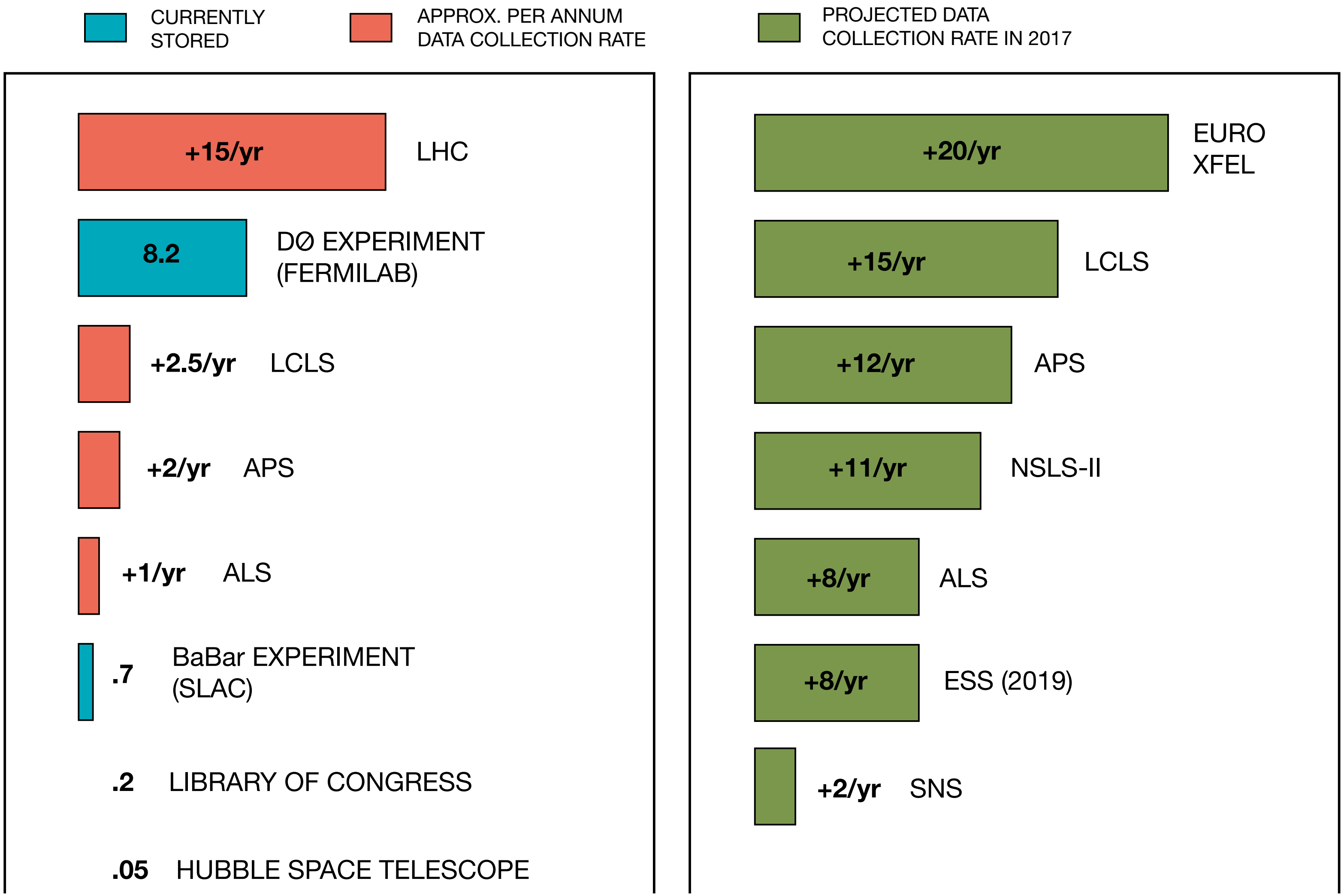


LCLS Data Rates & Volumes

- Three operation modes (X-Ray pulse rate):
  - **30 Hz** (first runs on LCLS, AMO/CAMP, Oct-Dec 2009)
  - **60 HZ** (second series of runs, May 2010 -> on)
  - **120 Hz** (Fall 2010)
- Instruments:
  - **Fall 2009**: AMO/CAMP
  - **Summer 2010**: AMO/CAMP, XPP, SXR
  - **Fall 2012:** all 6+1 instruments (station #2 for CXI)
- First runs at AMO/CAMP (30 Hz):
  - up to **180 MB/s**, **3 TB/day**, **~100 TB** of raw data recorded
- At full capacity (120 Hz):
  - **Up to 1.5 GB/s** for CXI (event size exceeds 10 MB)
  - **30 TB/shift**
  - **Up to 3 instruments can take data in parallel (**approaching this condition**)**
  - **1-2 PB** of raw data per year

Vary for different instruments/experiments

The Data Challenge view from 2012



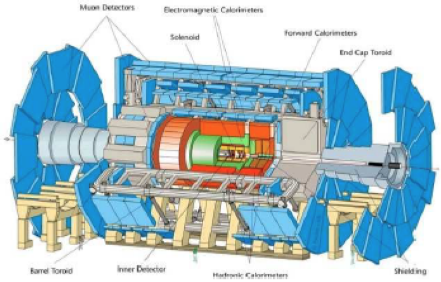
Data Volumes Growing everywhere

Data Rates and Data Volumes

Application	**Current data rates** [GB/h]		**Future data rates** [GB/h]	
	Peak	Average	Peak	Average
Protein Crystallography	500	50	500	200
Coherent Diffraction Imaging	500	50	4000	400
Tomography	700	50	800	200
Spectroscopy	450	45	18000	1800
Small Angle Scattering	1400	140	14400	4200
Grain Mapping	140	80	800	300

DESY: CFEL/PETRA III+/FLASH → 1.6 PB/year
Cern: ATLAS 100 MB/s → 3 PB/year

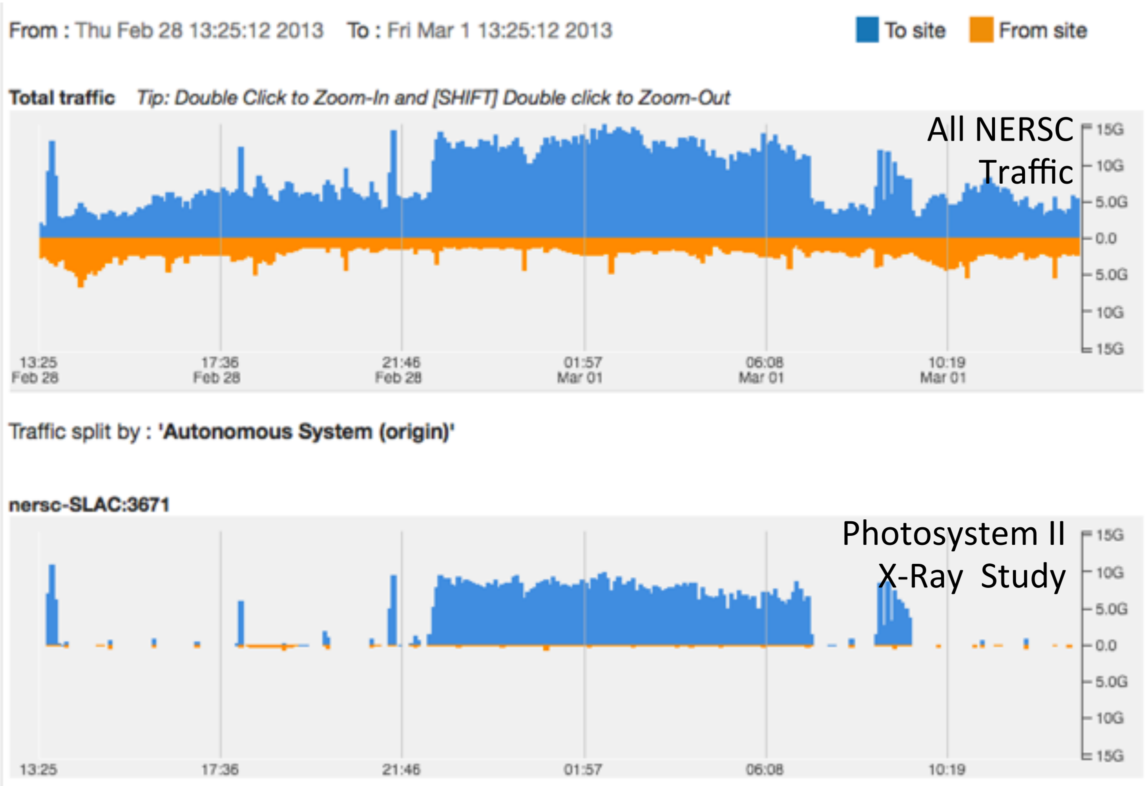
- Increasing source brilliance
- Faster detectors with more pixels (e.g. Pilatus)
- New experiment scenarios: Fast time resolved, scanning
- Use of simultaneous Detectors



LCLS Diffract & Destroy

Photo-system II, Nick Sauter (LBL): A single experiment in 2013 generated **120TB** over four days. Data was relayed to NERSC via Esnet, analysis required **135K CPU hours** of computing.

LCLS-II will require > **100M CPU hours** per experiment. Higher resolution and advanced image analysis could grow computational complexity. Some algorithms scale an M*NlogN for M images of N pixels.

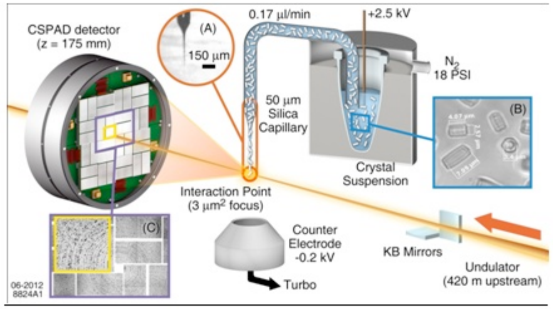


From : Thu Feb 28 13:25:12 2013 To : Fri Mar 1 13:25:12 2013

Total traffic Tip: Double Click to Zoom-In and [SHIFT] Double click to Zoom-Out

Traffic split by : 'Autonomous System (origin)'

nersc-SLAC:3671



A starter wish list…

- Easy to use data management and processing frameworks that scale with data rate
- Conditions data storage and management; meta data
- Data validation and fast feedback at collection
- Data volume reduction/compression techniques
- Development and availability of algorithm and algorithm tools
- Community developed simulations and simulation tool kits for beamlines and detectors
- Computational science
- Compute and storage hardware platforms appropriate for the taskS

Data management

- Life is easier if the data is managed from point of origin
- Detector readouts can have proprietary readouts

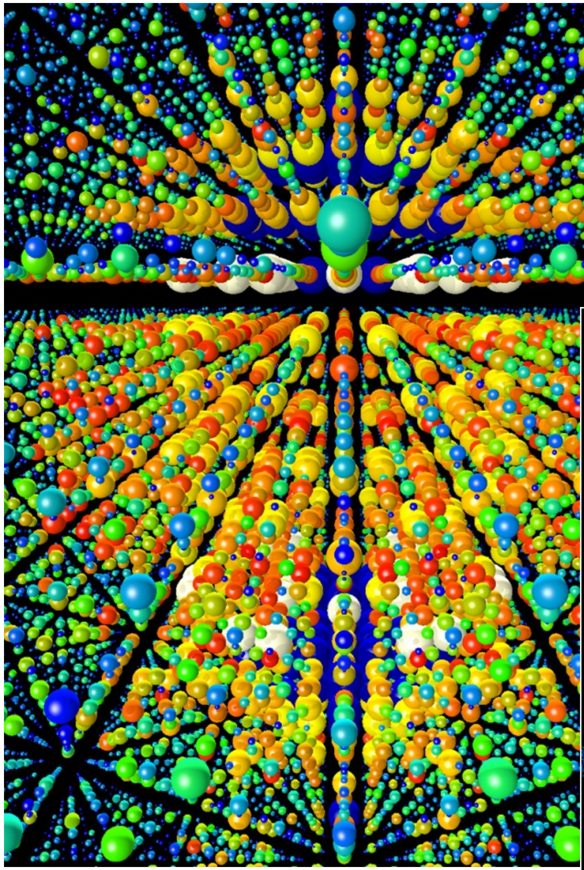
- Is standardization possible?
- Data containers/formats
- Metadata
- Workflows/tool kits
- Visualization tools
- Curation/cataloging

- Naively, some standardization
 would make good use of resources and expertise,
simplify life, open up many possibilities

Rapid feedback

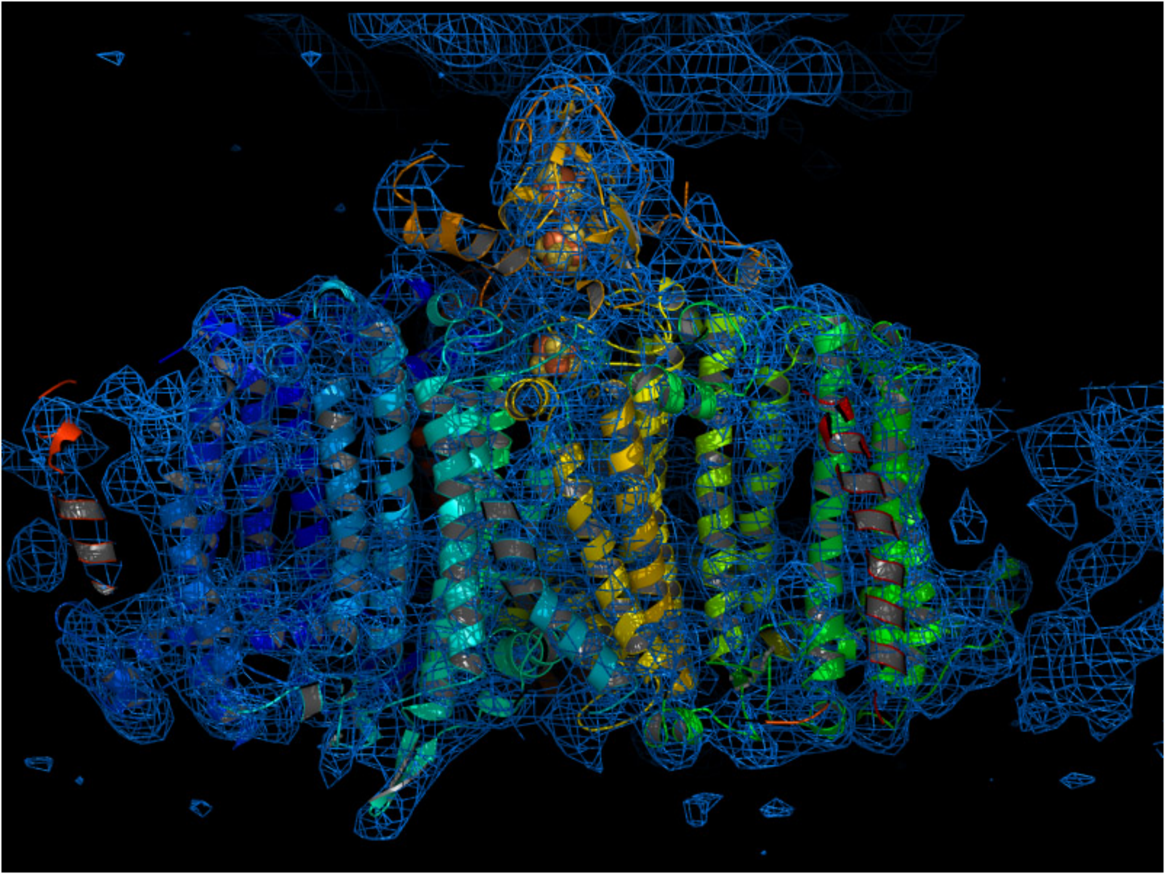
- Time is money
  - Beam time, experimenter time; instrument scientist time
- Experiment simulations
- Advance preparation of analysis tools

- Could provide better use of beam time, shorten time to publication and refine where to spend resources on computing improvements http://home.slac.stanford.edu/pressreleases/2011/20110202.htm



Three-dimensional rendering of the X-ray diffraction pattern for the Photosystem I protein, reconstructed from more than 15,000 single nanocrystal snapshots taken at the LCLS. (Image courtesy Thomas White, DESY.)

A reconstructed image of the Photosystem I complex. (Image courtesy Raimund Fromme, Arizona State University.)

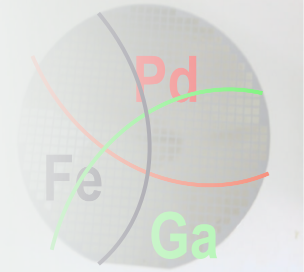


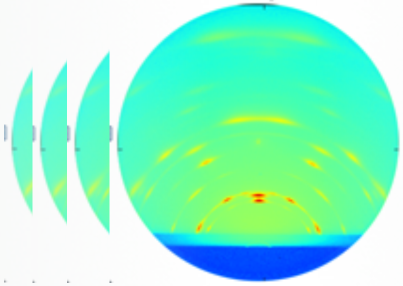
High-Throughput Pipeline:

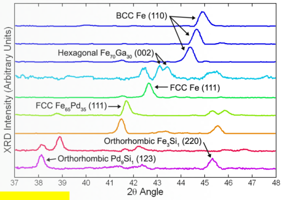
Robotics, Automation & Machine Learning

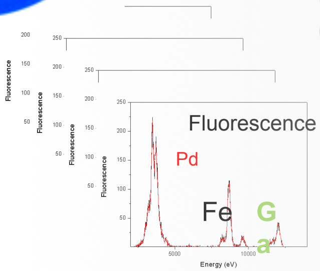
JCAP
JOINT CENTER FOR ARTIFICIAL PHOTOSYNTHESIS

John Gregoire
Ichiro Takeuchi
Matt Kramer
Apurva Mehta

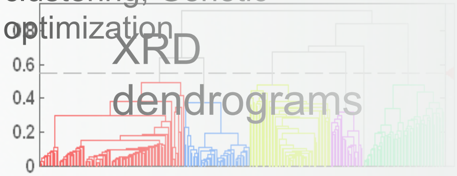


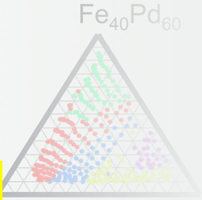






PCA, Density based clustering, Genetic optimization





Hubs

Hubs

Library Production

Data Collection

Data Archiving

Property Screening

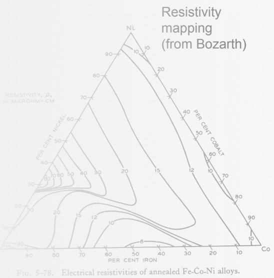
SSRL

New Composition-Structure-Property Relations

New Materials

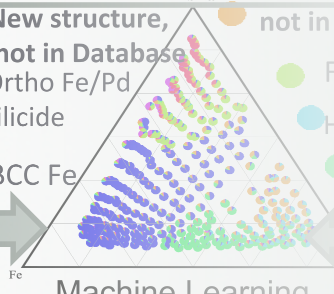
Scientific Communities

New Theory of (metastable) Materials



- Auto-alignment and Calibration
- Robotic Library Changer
- Library Tracking

- Multiple Data Stream Archiving
  - Hdf5?



● New structure, not in Database
● Ortho Fe/Pd silicide
● BCC Fe

New structure, not in Database
FCC FePd
Hex FeGa
FCC Fe

Theory

Machine Learning

Database, ICSD

Science & Technology
Facilities Council

ICAT

Facility Metadata Catalogue

Tom Griffin, STFC ISIS Facility

Overview

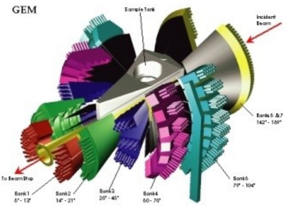
- What is ICAT?
- Uses and Benefits
- Installation
- Data Integration
- Interfaces
- About the ICAT project
- Future developments

What is ICAT?

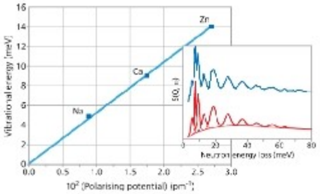
ICAT is a database, with a well defined API that provides a uniform interface to experimental data and a mechanism to link all aspects of research from proposal through to publication.

What is ICAT?

Example ISIS Proposal

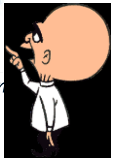


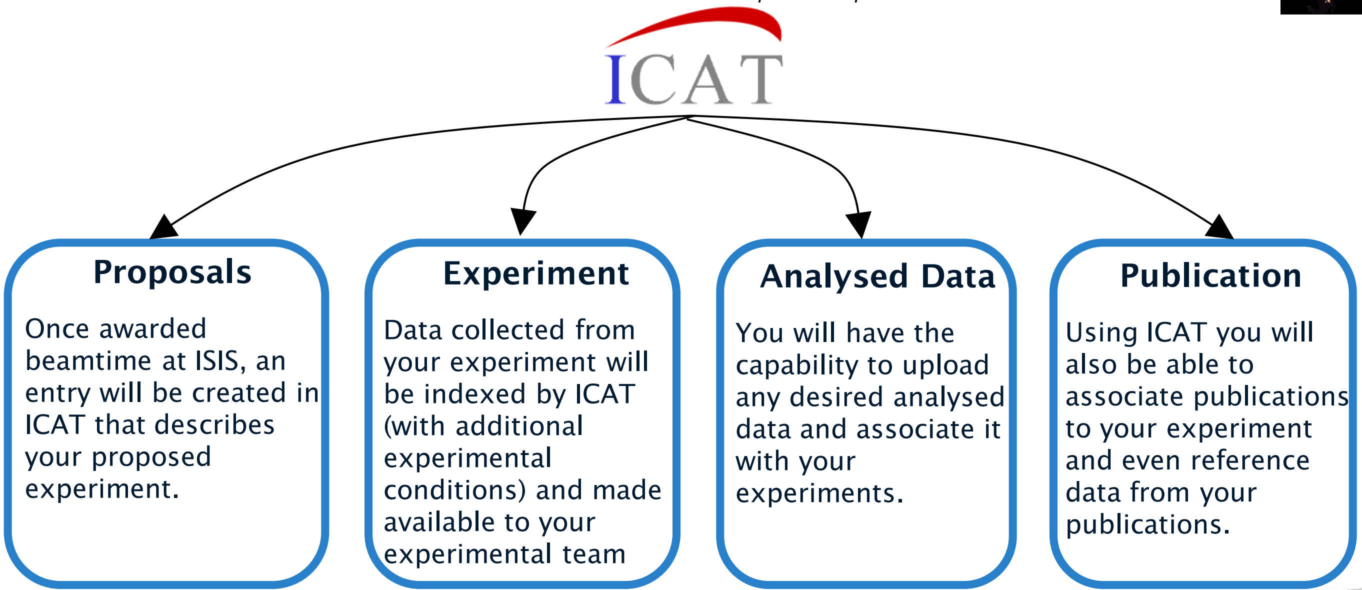
GEM – High intensity, high resolution neutron diffractometer



H2-(zeolite) vibrational frequencies vs polarising potential of cations

B-lactoglobulin protein interfacial structure





Proposals

Once awarded beamtime at ISIS, an entry will be created in ICAT that describes your proposed experiment.

Experiment

Data collected from your experiment will be indexed by ICAT (with additional experimental conditions) and made available to your experimental team

Analysed Data

You will have the capability to upload any desired analysed data and associate it with your experiments.

Publication

Using ICAT you will also be able to associate publications to your experiment and even reference data from your publications.

Science & Technology Facilities Council

Benefits

- Links proposal to data to analysis to publication
- Makes data searchable
- Remote data access
- Assign DOIs to data
- Increased opportunities for data sharing and re-use
- Implements a data policy
- Provenance (Creation, ownership, history)
- Integration with applications

Where is it used?

- STFC (main developer)
  - ISIS pulsed neutron and muon facility
  - CLF (UK national laser facility)
- Diamond
  - UK national synchrotron
- ILL
  - European reactor neutron source
- Oak Ridge National Laboratory (USA)
  - Spallation Neutron Source
- Currently being rolled out at: ESRF, ALBA, SOLEIL, other PaNdata partners

Technical

Installation

- Relational database - Oracle, MySQL
- Java application server - Glassfish
- Support available

Components

- ICAT is a modular system
  - Authentication
    - Plain text simple database – testing and development
    - LDAP/Active Directory
    - Custom user office connections
  - Data Server
    - Defines an interface – can be implemented as appropriate
    - ‘Disk only’ implementation (ISIS)
    - Tape/disk cache (Diamond)

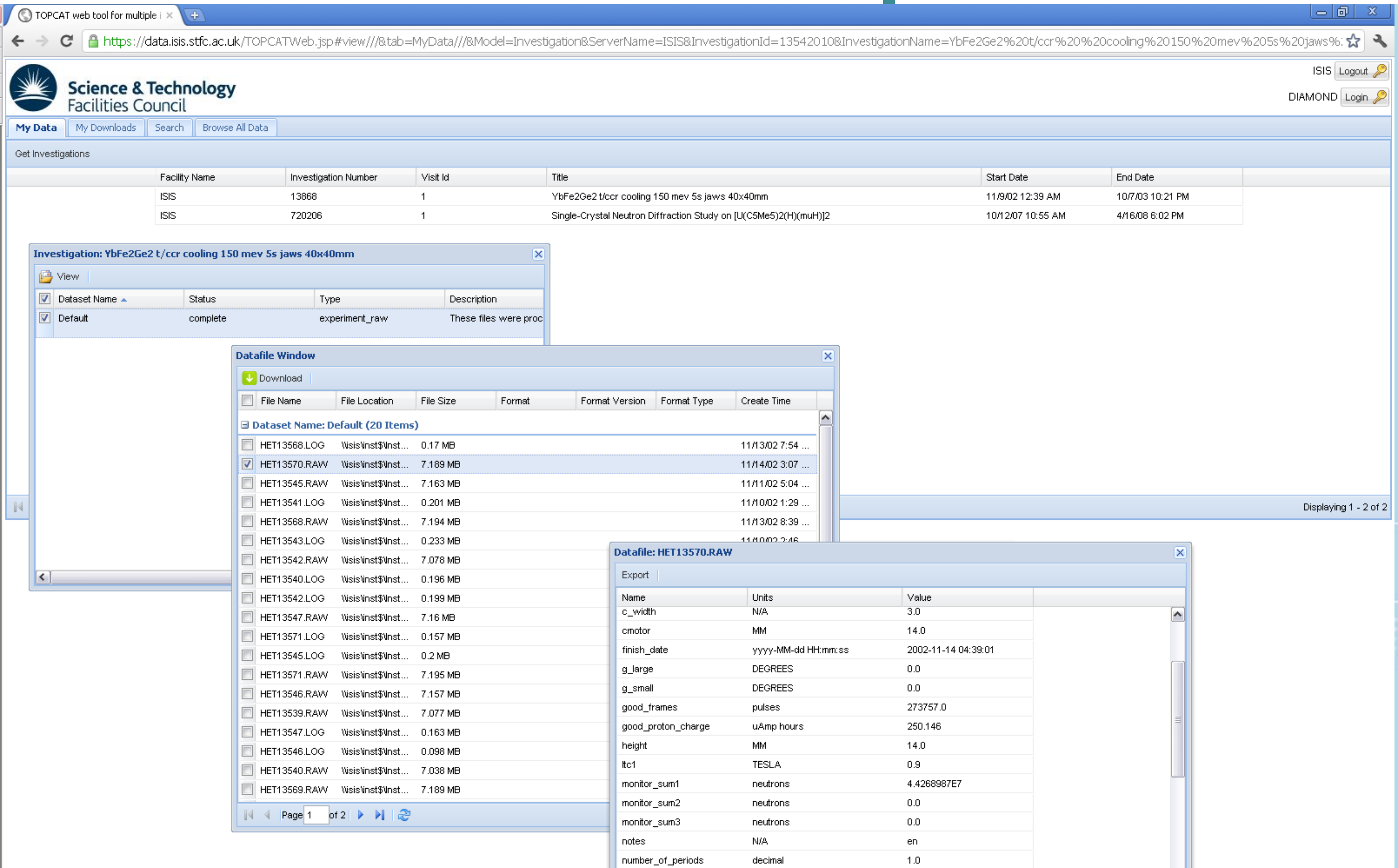
Flexibility

- Access to data is defined through a ‘rules’ system
  - E.g. Grant read access to data when a user is a Co-Investigator
  - Grant full control to data when a user is the Principal Investigator
  - Grant full control to all data on an instrument when the user is the instrument scientist for that instrument.
- Facility specific: Instruments (beamlines) , Instrument Scientists, Datafile types, parameters
  - E.g. wavelength, total_proton_count, etc

Interfaces

- ICAT exposes a web service (SOAP) API
  - Available for client applications
  - Enables integration with data analysis applications such as DAWN and Mantid.
  - RESTful interface planned
- Default web interface – ‘TopCAT’
  - Allows basic data browsing
  - Able to search many ICAT’s in parallel

Interfaces - TopCAT



https://data.isis.stfc.ac.uk/TOPCATWeb.jsp#view///&tab=MyData///&Model=Investigation&ServerName=ISIS&InvestigationId=13542010&InvestigationName=YbFe2Ge2%20t/ccr%20%20cooling%20150%20mev%205s%20jaws%:

Facility Name	Investigation Number	Visit Id	Title	Start Date	End Date
ISIS	13868	1	YbFe2Ge2 t/ccr cooling 150 mev 5s jaws 40x40mm	11/9/02 12:39 AM	10/7/03 10:21 PM
ISIS	720206	1	Single-Crystal Neutron Diffraction Study on [U(C5Me5)2(H)(muH)]2	10/12/07 10:55 AM	4/16/08 6:02 PM

Investigation: YbFe2Ge2 t/ccr cooling 150 mev 5s jaws 40x40mm

☑	Dataset Name	Status	Type	Description
☑	Default	complete	experiment_raw	These files were proc

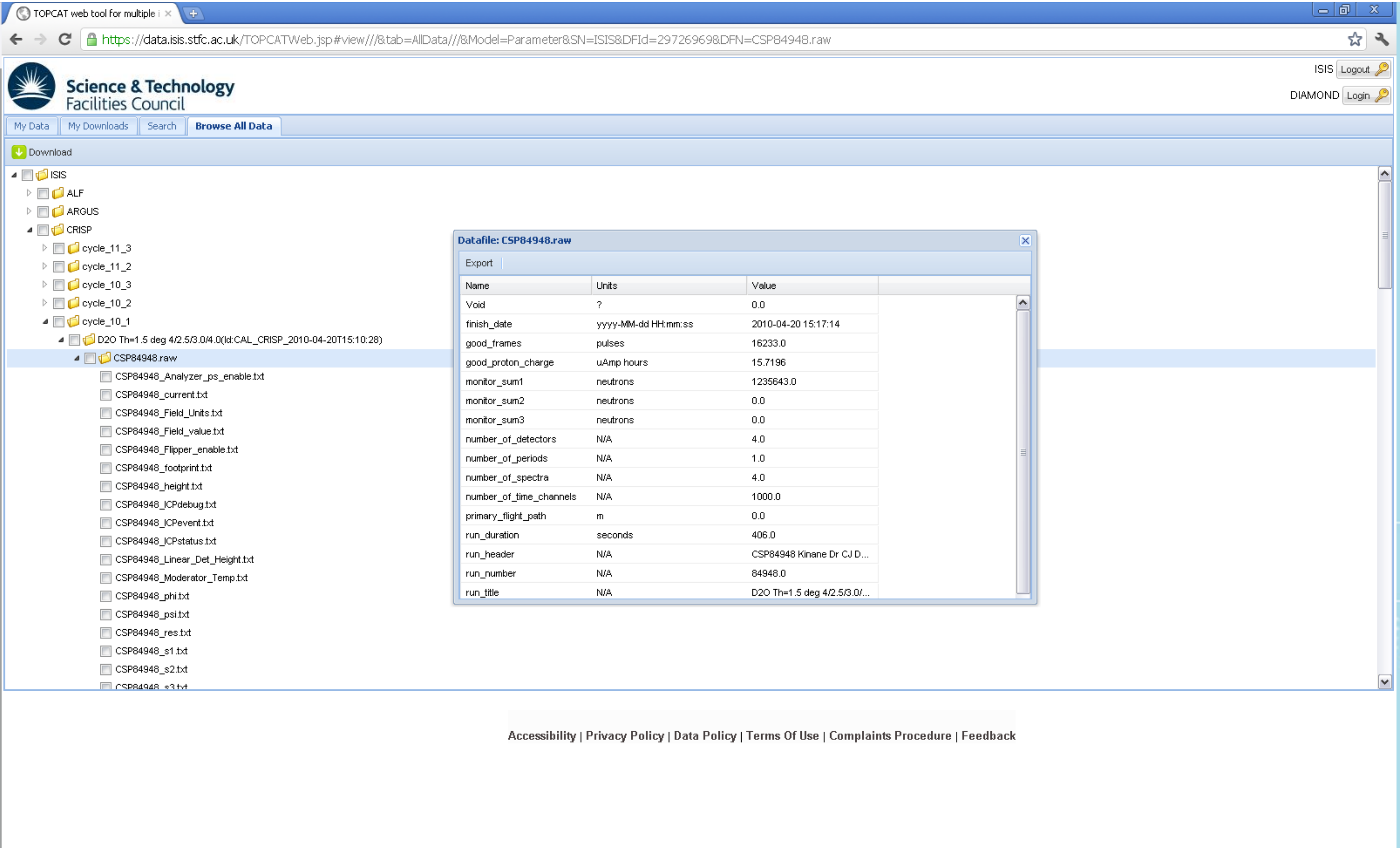
Datafile Window

□	File Name	File Location	File Size	Format	Format Version	Format Type	Create Time
	Dataset Name: Default (20 Items)						
□	HET13568.LOG	\\isis\inst$\Inst...	0.17 MB				11/13/02 7:54 ...
☑	HET13570.RAW	\\isis\inst$\Inst...	7.189 MB				11/14/02 3:07 ...
□	HET13545.RAW	\\isis\inst$\Inst...	7.163 MB				11/11/02 5:04 ...
□	HET13541.LOG	\\isis\inst$\Inst...	0.201 MB				11/10/02 1:29 ...
□	HET13568.RAW	\\isis\inst$\Inst...	7.194 MB				11/13/02 8:39 ...
□	HET13543.LOG	\\isis\inst$\Inst...	0.233 MB				11/10/02 2:46
□	HET13542.RAW	\\isis\inst$\Inst...	7.078 MB				
□	HET13540.LOG	\\isis\inst$\Inst...	0.196 MB				
□	HET13542.LOG	\\isis\inst$\Inst...	0.199 MB				
□	HET13547.RAW	\\isis\inst$\Inst...	7.16 MB				
□	HET13571.LOG	\\isis\inst$\Inst...	0.157 MB				
□	HET13545.LOG	\\isis\inst$\Inst...	0.2 MB				
□	HET13571.RAW	\\isis\inst$\Inst...	7.195 MB				
□	HET13546.RAW	\\isis\inst$\Inst...	7.157 MB				
□	HET13539.RAW	\\isis\inst$\Inst...	7.077 MB				
□	HET13547.LOG	\\isis\inst$\Inst...	0.163 MB				
□	HET13546.LOG	\\isis\inst$\Inst...	0.098 MB				
□	HET13540.RAW	\\isis\inst$\Inst...	7.038 MB				
□	HET13569.RAW	\\isis\inst$\Inst...	7.189 MB				

Datafile: HET13570.RAW

Name	Units	Value
c_width	N/A	3.0
cmotor	MM	14.0
finish_date	yyyy-MM-dd HH:mm:ss	2002-11-14 04:39:01
g_large	DEGREES	0.0
g_small	DEGREES	0.0
good_frames	pulses	273757.0
good_proton_charge	uAmp hours	250.146
height	MM	14.0
ltc1	TESLA	0.9
monitor_sum1	neutrons	4.4268987E7
monitor_sum2	neutrons	0.0
monitor_sum3	neutrons	0.0
notes	N/A	en
number_of_periods	decimal	1.0

Interfaces - TopCAT



TOPCAT web tool for multiple i
https://data.isis.stfc.ac.uk/TOPCATWeb.jsp#view///&tab=AllData//&Model=Parameter&SN=ISIS&DFId=29726969&DFN=CSP84948.raw
Science & Technology Facilities Council
ISIS Logout
DIAMOND Login
My Data
My Downloads
Search
Browse All Data
Download

- ISIS
  - ALF
  - ARGUS
  - CRISP
    - cycle_11_3
    - cycle_11_2
    - cycle_10_3
    - cycle_10_2
    - cycle_10_1
      - D2O Th=1.5 deg 4/2.5/3.0/4.0(Id:CAL_CRISP_2010-04-20T15:10:28)
        - CSP84948.raw
          - CSP84948_Analyzer_ps_enable.txt
          - CSP84948_current.txt
          - CSP84948_Field_Units.txt
          - CSP84948_Field_value.txt
          - CSP84948_Flipper_enable.txt
          - CSP84948_footprint.txt
          - CSP84948_height.txt
          - CSP84948_ICPdebug.txt
          - CSP84948_ICPevent.txt
          - CSP84948_ICPstatus.txt
          - CSP84948_Linear_Det_Height.txt
          - CSP84948_Moderator_Temp.txt
          - CSP84948_phi.txt
          - CSP84948_psi.txt
          - CSP84948_res.txt
          - CSP84948_s1.txt
          - CSP84948_s2.txt
          - CSP84948_s3.txt

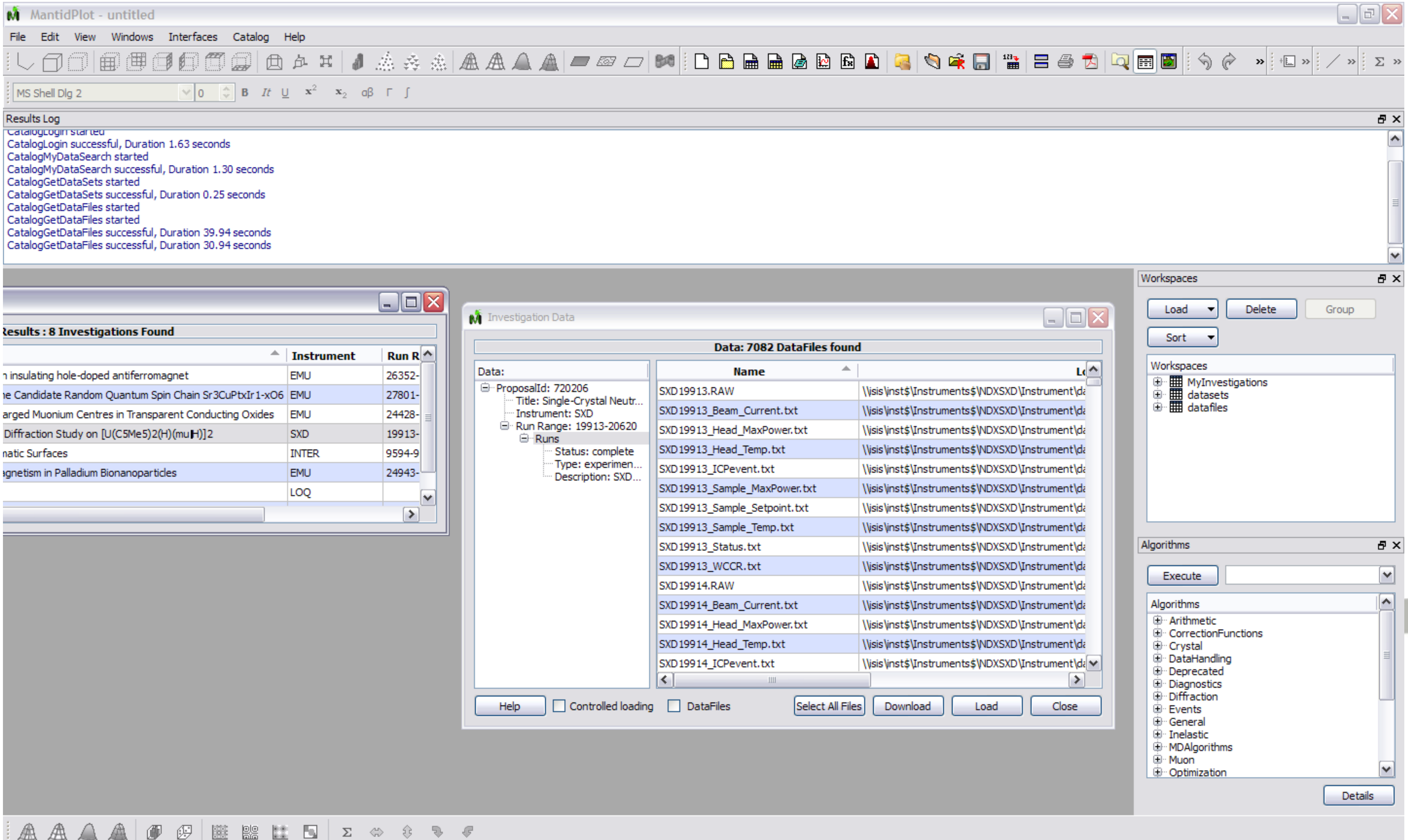
Datafile: CSP84948.raw

Export

Name	Units	Value
Void	?	0.0
finish_date	yyyy-MM-dd HH:mm:ss	2010-04-20 15:17:14
good_frames	pulses	16233.0
good_proton_charge	uAmp hours	15.7196
monitor_sum1	neutrons	1235643.0
monitor_sum2	neutrons	0.0
monitor_sum3	neutrons	0.0
number_of_detectors	N/A	4.0
number_of_periods	N/A	1.0
number_of_spectra	N/A	4.0
number_of_time_channels	N/A	1000.0
primary_flight_path	m	0.0
run_duration	seconds	406.0
run_header	N/A	CSP84948 Kinane Dr CJ D...
run_number	N/A	84948.0
run_title	N/A	D2O Th=1.5 deg 4/2.5/3.0/...

Accessibility | Privacy Policy | Data Policy | Terms Of Use | Complaints Procedure | Feedback

Interfaces - Mantid



```
CatalogLogin started
CatalogLogin successful, Duration 1.63 seconds
CatalogMyDataSearch started
CatalogMyDataSearch successful, Duration 1.30 seconds
CatalogGetDataSets started
CatalogGetDataSets successful, Duration 0.25 seconds
CatalogGetDataFiles started
CatalogGetDataFiles started
CatalogGetDataFiles successful, Duration 39.94 seconds
CatalogGetDataFiles successful, Duration 30.94 seconds
```

Results : 8 Investigations Found

	Instrument	Run R
n insulating hole-doped antiferromagnet	EMU	26352-
ne Candidate Random Quantum Spin Chain Sr3CuPtxIr1-xO6	EMU	27801-
arged Muonium Centres in Transparent Conducting Oxides	EMU	24428-
Diffraction Study on [U(C5Me5)2(H)(muH)]2	SXD	19913-
natic Surfaces	INTER	9594-9
agnetism in Palladium Bionanoparticles	EMU	24943-
	LOQ	

Investigation Data

Data: 7082 DataFiles found

Data:
- ProposalId: 720206
  - Title: Single-Crystal Neutr...
  - Instrument: SXD
  - Run Range: 19913-20620
    - Runs
      - Status: complete
      - Type: experimen...
      - Description: SXD...

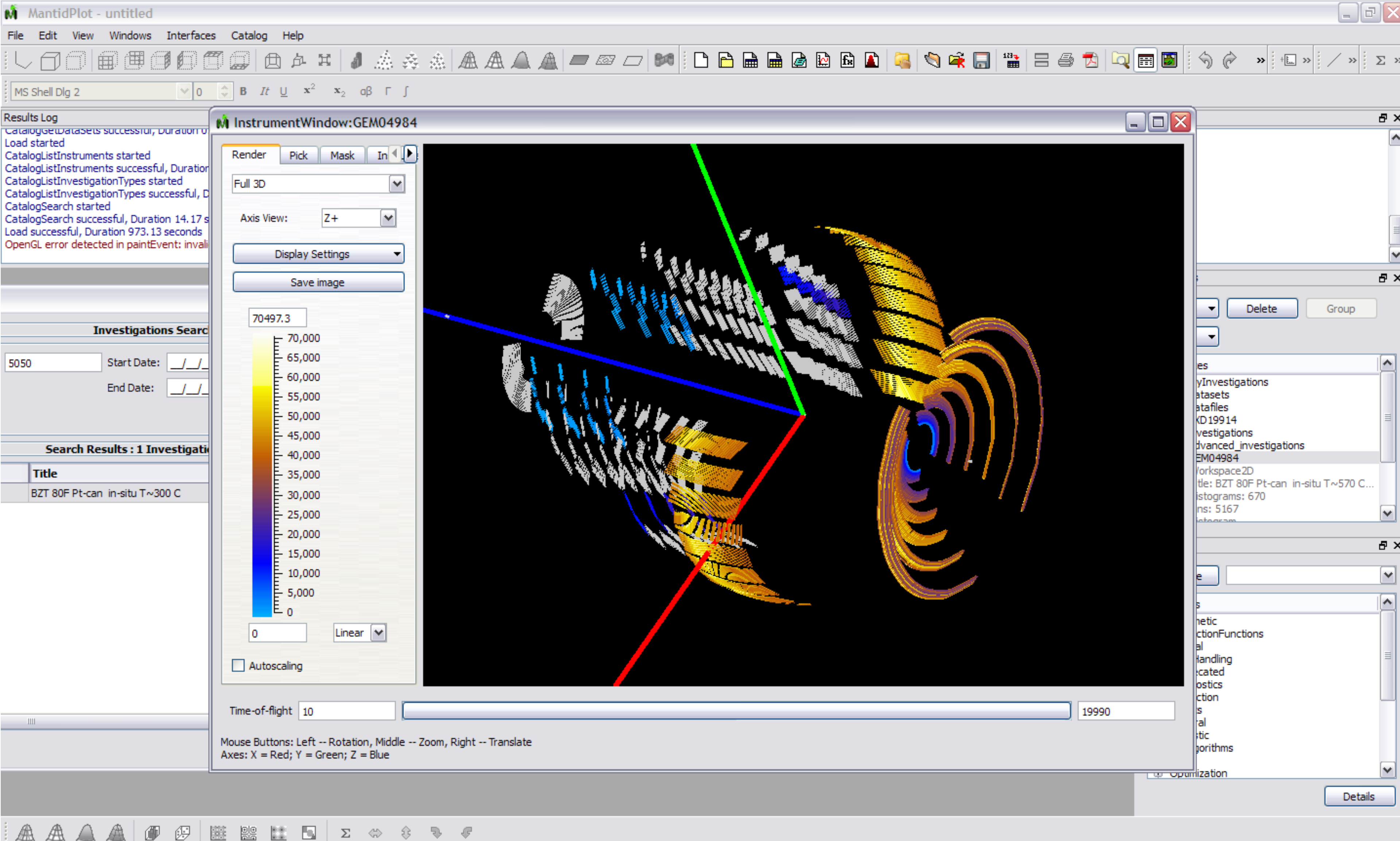
Name	Lo
SXD19913.RAW	\\isis\inst$\Instruments$\NDXSXD\Instrument\da
SXD19913_Beam_Current.txt	\\isis\inst$\Instruments$\NDXSXD\Instrument\da
SXD19913_Head_MaxPower.txt	\\isis\inst$\Instruments$\NDXSXD\Instrument\da
SXD19913_Head_Temp.txt	\\isis\inst$\Instruments$\NDXSXD\Instrument\da
SXD19913_ICPevent.txt	\\isis\inst$\Instruments$\NDXSXD\Instrument\da
SXD19913_Sample_MaxPower.txt	\\isis\inst$\Instruments$\NDXSXD\Instrument\da
SXD19913_Sample_Setpoint.txt	\\isis\inst$\Instruments$\NDXSXD\Instrument\da
SXD19913_Sample_Temp.txt	\\isis\inst$\Instruments$\NDXSXD\Instrument\da
SXD19913_Status.txt	\\isis\inst$\Instruments$\NDXSXD\Instrument\da
SXD19913_WCCR.txt	\\isis\inst$\Instruments$\NDXSXD\Instrument\da
SXD19914.RAW	\\isis\inst$\Instruments$\NDXSXD\Instrument\da
SXD19914_Beam_Current.txt	\\isis\inst$\Instruments$\NDXSXD\Instrument\da
SXD19914_Head_MaxPower.txt	\\isis\inst$\Instruments$\NDXSXD\Instrument\da
SXD19914_Head_Temp.txt	\\isis\inst$\Instruments$\NDXSXD\Instrument\da
SXD19914_ICPevent.txt	\\isis\inst$\Instruments$\NDXSXD\Instrument\da

Help □ Controlled loading □ DataFiles Select All Files Download Load Close

Workspaces
- MyInvestigations
- datasets
- datafiles

Algorithms
- Arithmetic
- CorrectionFunctions
- Crystal
- DataHandling
- Deprecated
- Diagnostics
- Diffraction
- Events
- General
- Inelastic
- MDAlgorithms
- Muon
- Optimization

Interfaces - Mantid



```
CatalogGetDataSets successful, Duration 0
Load started
CatalogListInstruments started
CatalogListInstruments successful, Duration
CatalogListInvestigationTypes started
CatalogListInvestigationTypes successful, D
CatalogSearch started
CatalogSearch successful, Duration 14.17 s
Load successful, Duration 973.13 seconds
OpenGL error detected in paintEvent: invali
```

5050 Start Date: __/__/__

End Date: __/__/__

Title
BZT 80F Pt-can in-situ T~300 C

Full 3D

Axis View: Z+

70497.3

0

Linear

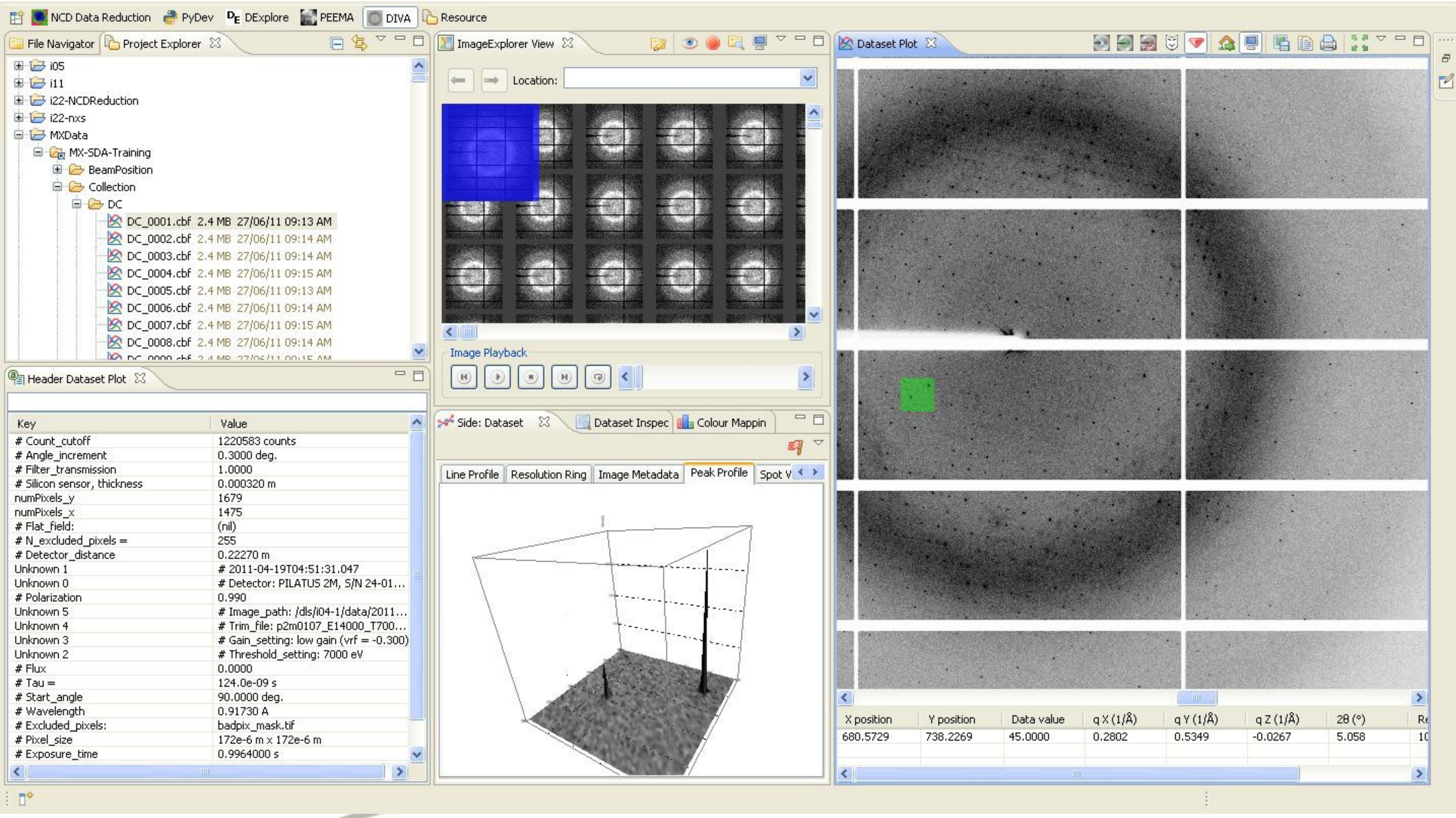
☐ Autoscaling

Time-of-flight 10 19990

Mouse Buttons: Left -- Rotation, Middle -- Zoom, Right -- Translate
Axes: X = Red; Y = Green; Z = Blue

tle: BZT 80F Pt-can in-situ T~570 C...
istograms: 670
ns: 5167

Interfaces - DAWN

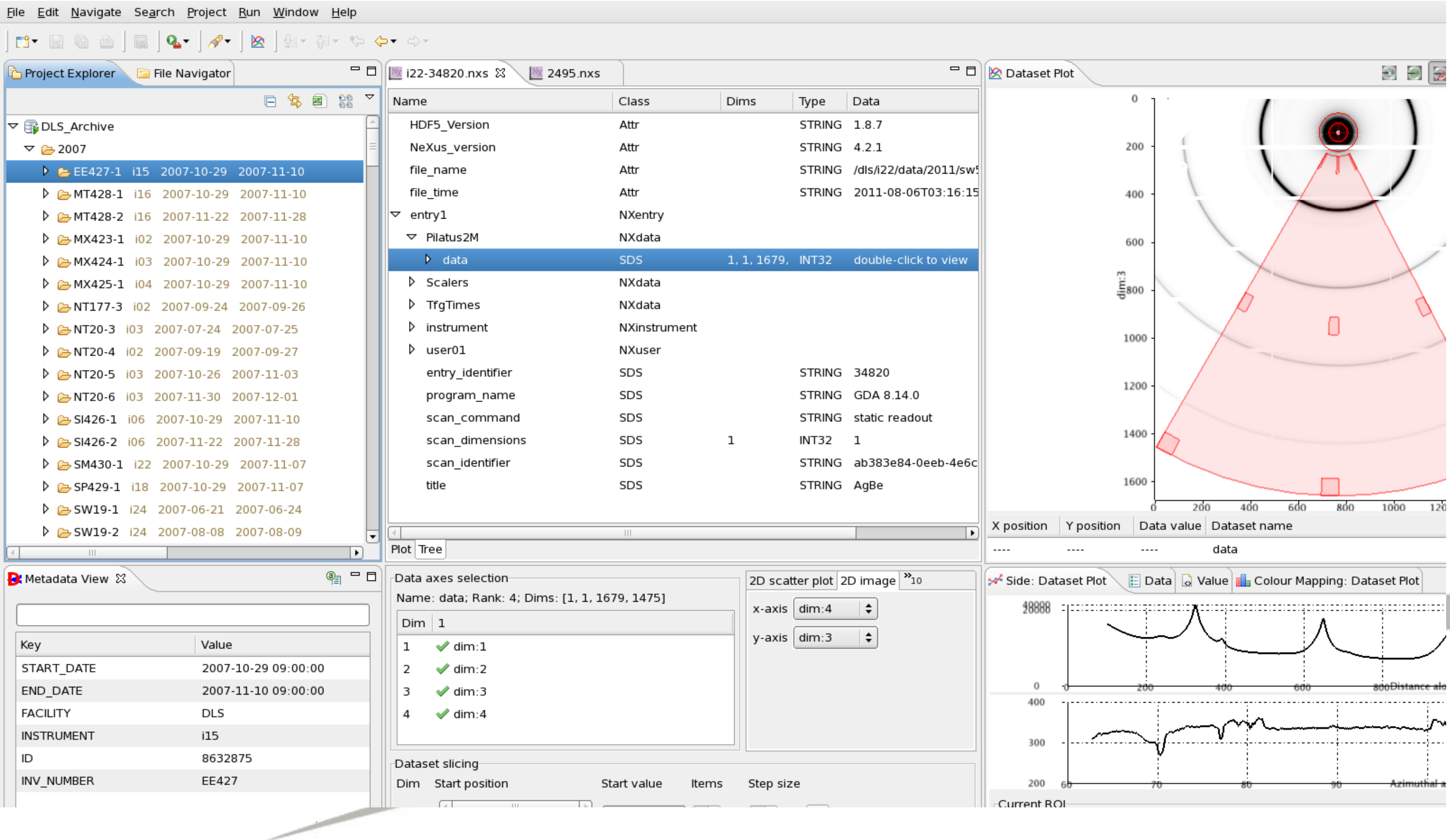


i05
i11
i22-NCDReduction
i22-nxs
MXData
MX-SDA-Training
BeamPosition
Collection
DC
DC_0001.cbf 2.4 MB 27/06/11 09:13 AM
DC_0002.cbf 2.4 MB 27/06/11 09:14 AM
DC_0003.cbf 2.4 MB 27/06/11 09:14 AM
DC_0004.cbf 2.4 MB 27/06/11 09:15 AM
DC_0005.cbf 2.4 MB 27/06/11 09:13 AM
DC_0006.cbf 2.4 MB 27/06/11 09:14 AM
DC_0007.cbf 2.4 MB 27/06/11 09:15 AM
DC_0008.cbf 2.4 MB 27/06/11 09:14 AM
DC_0009.cbf 2.4 MB 27/06/11 09:15 AM

Key	Value
# Count_cutoff	1220583 counts
# Angle_increment	0.3000 deg.
# Filter_transmission	1.0000
# Silicon sensor, thickness	0.000320 m
numPixels_y	1679
numPixels_x	1475
# Flat_field:	(nil)
# N_excluded_pixels =	255
# Detector_distance	0.22270 m
Unknown 1	# 2011-04-19T04:51:31.047
Unknown 0	# Detector: PILATUS 2M, S/N 24-01…
# Polarization	0.990
Unknown 5	# Image_path: /dls/i04-1/data/2011…
Unknown 4	# Trim_file: p2m0107_E14000_T700…
Unknown 3	# Gain_setting: low gain (vrf = -0.300)
Unknown 2	# Threshold_setting: 7000 eV
# Flux	0.0000
# Tau =	124.0e-09 s
# Start_angle	90.0000 deg.
# Wavelength	0.91730 A
# Excluded_pixels:	badpix_mask.tif
# Pixel_size	172e-6 m x 172e-6 m
# Exposure_time	0.9964000 s

X position	Y position	Data value	q X (1/Å)	q Y (1/Å)	q Z (1/Å)	2θ (°)	Re
680.5729	738.2269	45.0000	0.2802	0.5349	-0.0267	5.058	10

Interfaces - DAWN



DLS_Archive			
2007			
EE427-1	i15	2007-10-29	2007-11-10
MT428-1	i16	2007-10-29	2007-11-10
MT428-2	i16	2007-11-22	2007-11-28
MX423-1	i02	2007-10-29	2007-11-10
MX424-1	i03	2007-10-29	2007-11-10
MX425-1	i04	2007-10-29	2007-11-10
NT177-3	i02	2007-09-24	2007-09-26
NT20-3	i03	2007-07-24	2007-07-25
NT20-4	i02	2007-09-19	2007-09-27
NT20-5	i03	2007-10-26	2007-11-03
NT20-6	i03	2007-11-30	2007-12-01
SI426-1	i06	2007-10-29	2007-11-10
SI426-2	i06	2007-11-22	2007-11-28
SM430-1	i22	2007-10-29	2007-11-07
SP429-1	i18	2007-10-29	2007-11-07
SW19-1	i24	2007-06-21	2007-06-24
SW19-2	i24	2007-08-08	2007-08-09

Name	Class	Dims	Type	Data
HDF5_Version	Attr		STRING	1.8.7
NeXus_version	Attr		STRING	4.2.1
file_name	Attr		STRING	/dls/i22/data/2011/sw5
file_time	Attr		STRING	2011-08-06T03:16:15
entry1	NXentry			
Pilatus2M	NXdata			
data	SDS	1, 1, 1679,	INT32	double-click to view
Scalers	NXdata			
TfgTimes	NXdata			
instrument	NXinstrument			
user01	NXuser			
entry_identifier	SDS		STRING	34820
program_name	SDS		STRING	GDA 8.14.0
scan_command	SDS		STRING	static readout
scan_dimensions	SDS	1	INT32	1
scan_identifier	SDS		STRING	ab383e84-0eeb-4e6c
title	SDS		STRING	AgBe

X position	Y position	Data value	Dataset name
----	----	----	data

Key	Value
START_DATE	2007-10-29 09:00:00
END_DATE	2007-11-10 09:00:00
FACILITY	DLS
INSTRUMENT	i15
ID	8632875
INV_NUMBER	EE427

Data axes selection

Name: data; Rank: 4; Dims: [1, 1, 1679, 1475]

Dim	1
1	✓ dim:1
2	✓ dim:2
3	✓ dim:3
4	✓ dim:4

x-axis dim:4

y-axis dim:3

Dataset slicing

Dim	Start position	Start value	Items	Step size

Populating ICAT with Data

- ICAT has a SOAP API so data can be pushed in from most languages
- Typically metadata is imported from two sources
  - User office
  - Experiment data files
- User office link code tends to be bespoke
- Can be simple (experiment title, people) or more complex (abstracts, samples, links between related work etc)

Populating ICAT with Data

- Data files
  - Tools exist to assist ingesting nexus files
  - Ingest custom/specific data formats will require code to extract the metadata from them
- Can be done using the API or ‘XML Ingest’ layer
- XML Ingest defines a simple schema that describes a set of datafiles and inserts them into an ICAT
- Prototype available, but still work-in-progress
- Will simplify data ingestion.

The ICAT Project

The ICAT Project

- ICAT is an open source project
- Currently managed by STFC
- Released under BSD and Apache licenses (permissive)
- www.icatproject.org
- http://code.google.com/p/icatproject
- Monthly collaboration meetings (telephone)
- Annual developer meetings (face-to-face)
- Steering committee

Future Developments

- Under active development
- Releases every 6 months
- Current roadmap
  - Clustered deployment
  - Improved documentation
  - Data import/export/migration
  - Non-relational databases (hybrid)
  - Interface improvements to TopCAT

  http://icatproject.org/releases/road-map/
- Tell us what you want to see…..

Summary

- ICAT is a mature solution for facility metadata management
- Enables remote data access and linking between proposals, data, analysis and publications
- Flexible architecture enables integration with differing facility systems and requirements
- API enables integration with other software (analysis)
- Open source project with an active collaboration

Questions?

Thanks to: Steve Fisher, Brian Matthews, Alistair Mills, Alun Ashton, Antony Wilson

MONASH University

Store.Synchrotron.org.au

Steve Androulakis

Store.Synchrotron.org.au

- Store.Synchrotron is a system that captures all macromolecular beamline data, available online to all non-commercial Australian Synchrotron users. It was developed by Monash University in a strategic, ongoing partnership.

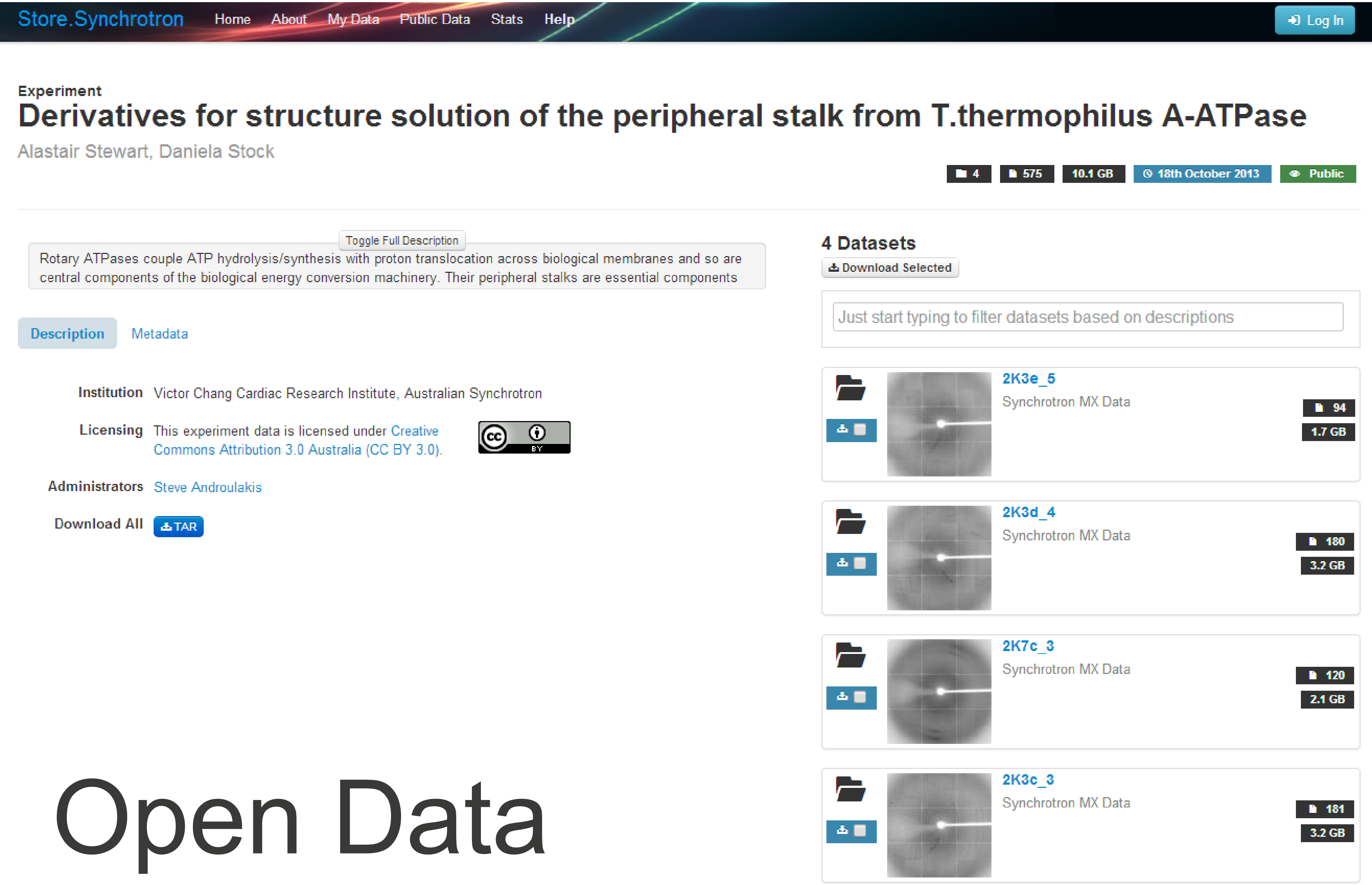
- Data is **immediately shareable** by the researcher on the web and able to be **published**.

- The service operates on the Australian NeCTAR Research compute cloud in a scalable setup able to withstand load. (http://nectar.org.au/)

- We're actively opening access to raw data behind high-impact research publications under CC BY licenses. Six institutions have opened data so far.

- Built on MyTardis – an open source, Australian made data management platform used all over Australia in proteomics, genomics, electron microscopy, medical imaging, astrophysics, quantum physics and more.

- Visit store.synchrotron.org.au for access / contact steve.Androulakis@monash.edu



Experiment

Derivatives for structure solution of the peripheral stalk from T.thermophilus A-ATPase

Alastair Stewart, Daniela Stock

Rotary ATPases couple ATP hydrolysis/synthesis with proton translocation across biological membranes and so are central components of the biological energy conversion machinery. Their peripheral stalks are essential components

Institution Victor Chang Cardiac Research Institute, Australian Synchrotron

Licensing This experiment data is licensed under Creative Commons Attribution 3.0 Australia (CC BY 3.0).

Administrators Steve Androulakis

Download All TAR

4 Datasets

2K3e_5
Synchrotron MX Data

2K3d_4
Synchrotron MX Data

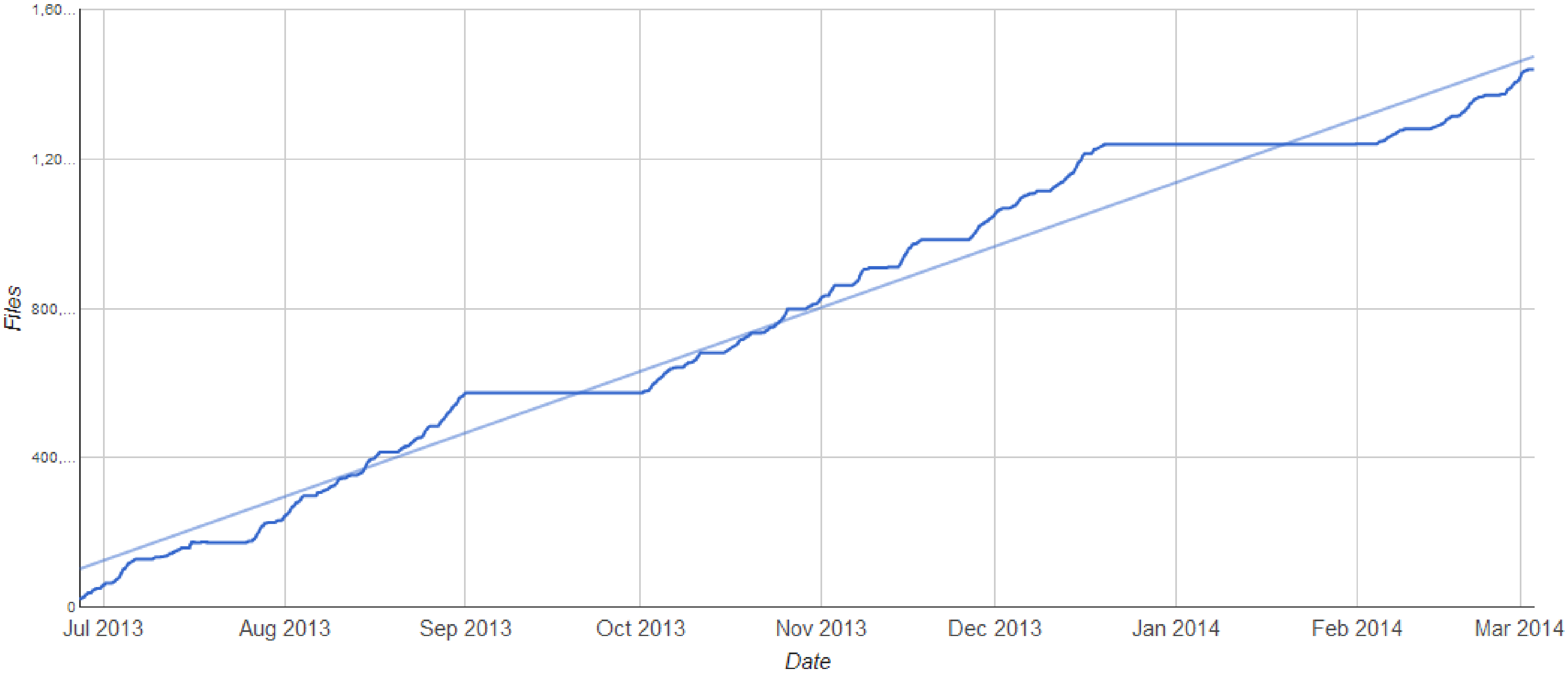
2K7c_3
Synchrotron MX Data

2K3c_3
Synchrotron MX Data

Open Data

MONASH University Australian Synchrotron nectar

Real-time instrument data capture



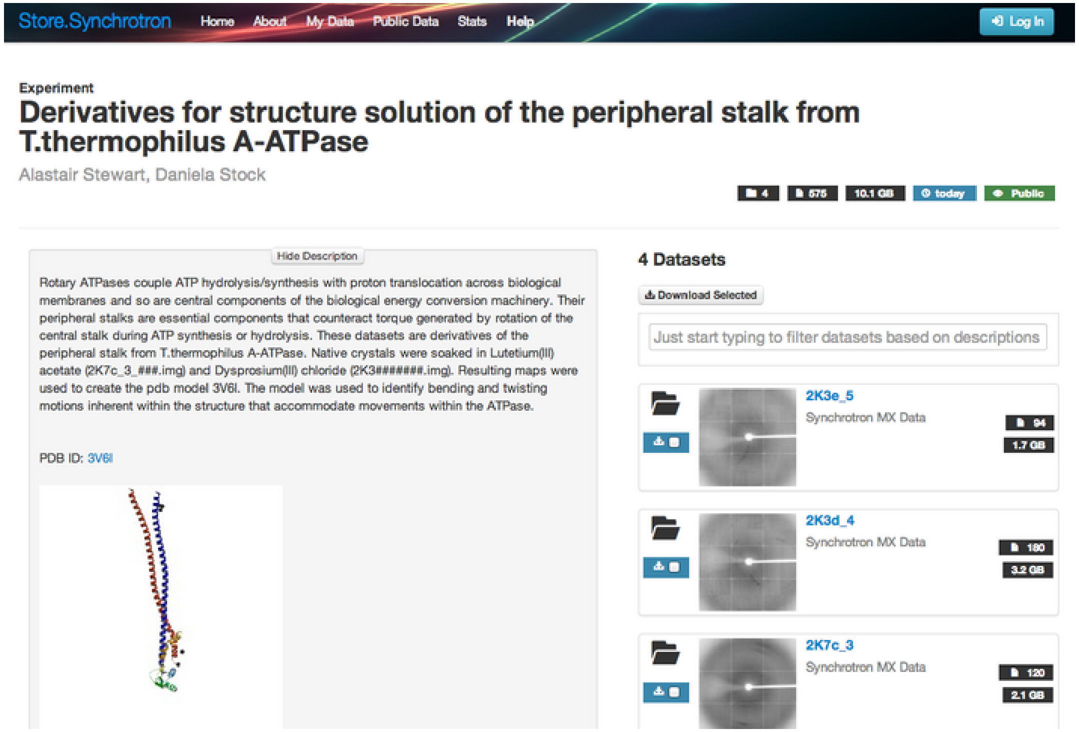
Capture began June 2013. As of March 2013, it has captured over 18 terabytes of data in over 1.5 million raw diffraction images.

MONASH University

Source: http://bdp-aaf-dev.dyndns.org/graphtime.html

Publish your research, then publish your raw data with us.

The Australian Synchrotron is working with researchers to open access to datasets associated with publications. This process is entirely opt-in for researchers and gives permission for the Synchrotron to host your data publicly.



Experiment

Derivatives for structure solution of the peripheral stalk from T.thermophilus A-ATPase

Alastair Stewart, Daniela Stock

Rotary ATPases couple ATP hydrolysis/synthesis with proton translocation across biological membranes and so are central components of the biological energy conversion machinery. Their peripheral stalks are essential components that counteract torque generated by rotation of the central stalk during ATP synthesis or hydrolysis. These datasets are derivatives of the peripheral stalk from T.thermophilus A-ATPase. Native crystals were soaked in Lutetium(III) acetate (2K7c_3_###.img) and Dysprosium(III) chloride (2K3#######.img). Resulting maps were used to create the pdb model 3V6I. The model was used to identify bending and twisting motions inherent within the structure that accommodate movements within the ATPase.

PDB ID: 3V6I

4 Datasets

2K3e_5
Synchrotron MX Data
94
1.7 GB

2K3d_4
Synchrotron MX Data
180
3.2 GB

2K7c_3
Synchrotron MX Data
120
2.1 GB

Data is made available via MyTardis at Store.Synchrotron.org.au

What do you need to do?

Contact
mxtickets@synchrotron.org.au

You'll be emailed a short form to fill in basic information about open access to your datasets associated with PDBs and publications.

Once data is made available via our service, you will receive a DOI to this data for citation.

Note: The data you're collecting today is also archived privately and securely via this system.

Log in to https://store.synchrotron.org.au with your Australian Synchrotron credentials to browse and download your data.

Presented on the Synchrotron's Control PCs in a slideshow.

How can we contribute to PaNSig?

- Store.Synchrotron.org.au is a scalable cloud service that was created to withstand network bandwidth, CPU, disk IO. We have considerable engineering experience in this area and can provide advice and guidance.

- We have worked with Synchrotron users in order to have their raw data released publicly (http://tardis.edu.au/syncpublish/) and have a basic model for this process.

- The Australian National Data Service (ANDS.org.au) is funding an 18 month project to refine the Store.Synchrotron.org.au process of collecting quality metadata at time of collection, and time of publication. Many of its outputs would be relevant to PaNSig.

- Monash University are in the process of increasing its 'instrument integration' of this kind using the same underlying system as Store.Synchrotron. This includes expansion to more beamlines at the Australian Synchrotron and likely the Australian neutron source ANSTO. Experiences here with different research communities, policies, practices, technology and culture should be broadly useful to this group.

Developing a plan of activities for RDA PaNSig

RDA Plenary, Dublin, March 2014

Co-chairs

Currently:

- – Amber Boehnlein , SLAC, US
- – Frank Schluenzen, DESY, DE
- – Brian Matthews, STFC, UK

Topics for Discussion

- Purpose
- Technical activities: RDA WGs etc
- Interaction within RDA
- Future Meetings
- Interactions between meetings
- Members

Purpose

- Data related issues of science applications associated with large scale source facilities
- Source facilities share a number of issues in their data handling.
- These could include:
  - Scalability of data volumes and data access rates
  - Standardization of (meta-)data and vocabularies
  - Data, cataloguing, publishing, discovery, sharing, transfer and access, policies
  - Data analysis tools and frameworks supporting workflows and provenance
  - Interaction with the data handling practices and standards within different communities.

Technical Activities

- Spinning out RDA technical working groups
  - Specific
  - Time-bound (c. 18 months)
  - Clear outcome (document, format, code etc)
  - Need commitment
  - **Topics** ?

- Any documents as a group ?
  - capturing community views.
- Future projects?

Interaction with other RDA groups

- We will be requested to comment on draft recommendations from other RDA working groups
  - Will notify group and request if people want to review documents
- Are there specific groups we want to interact with?
  - Domain specific groups: Structural Biology, Materials Data
  - Technical infrastructure: Persistent IDs, Big data Analytics, cloud computing, metadata, data publication, …..

Future meetings

- RDA Plenary
  - 4th RDA Plenary: Amsterdam, 22-24 September, 2014
  - 5th RDA Plenary: USA, March 2015
- NoBugs 2014 ?
- Other events?
- Opportunities to establish a PaN Computing event similar to CHEP ??

Interactions in between f2f

- Keeping up momentum is important
  - Website/Wiki (RDA Mediated)
  - mailing list (RDA Mediated)
  - phone/video meetings?
- Subgroups on particular topics?

Other members

- Who should we be aiming at ?
- PaNS outside US/EU/AU ?

Possible immediate activities?

- Online technical forum
- Counting users
- Collecting views data usage
- Project/software “market”
- Common terms
- Other RDA Groups: Structural Biology, Materials, Persistent ID …
- Meetings : RDA the Hague

Other discussion?

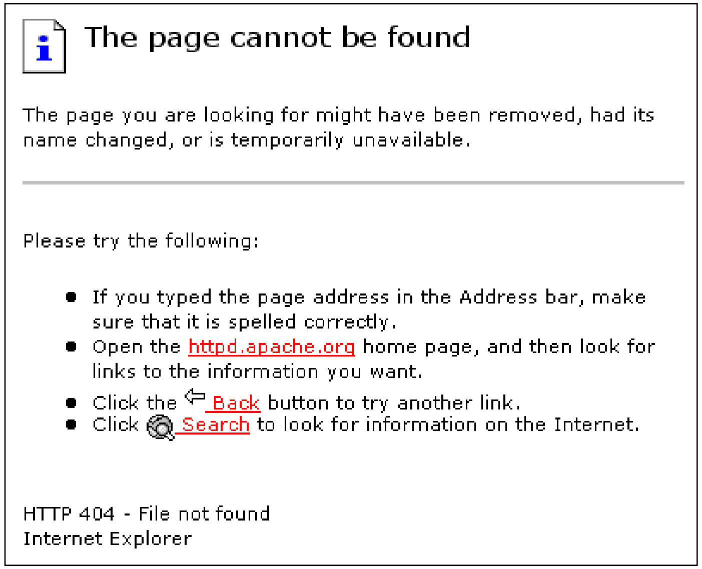
DataCite –
Persistent links to scientific data

DataCite

DOI names for citations

URLs are not persistent

- (e.g. Wren JD: **URL decay in MEDLINE- a 4-year follow-up study**. Bioinformatics. 2008, Jun 1;24(11):1381-5).



The page cannot be found

The page you are looking for might have been removed, had its name changed, or is temporarily unavailable.

Please try the following:

- If you typed the page address in the Address bar, make sure that it is spelled correctly.
- Open the httpd.apache.org home page, and then look for links to the information you want.
- Click the Back button to try another link.
- Click Search to look for information on the Internet.

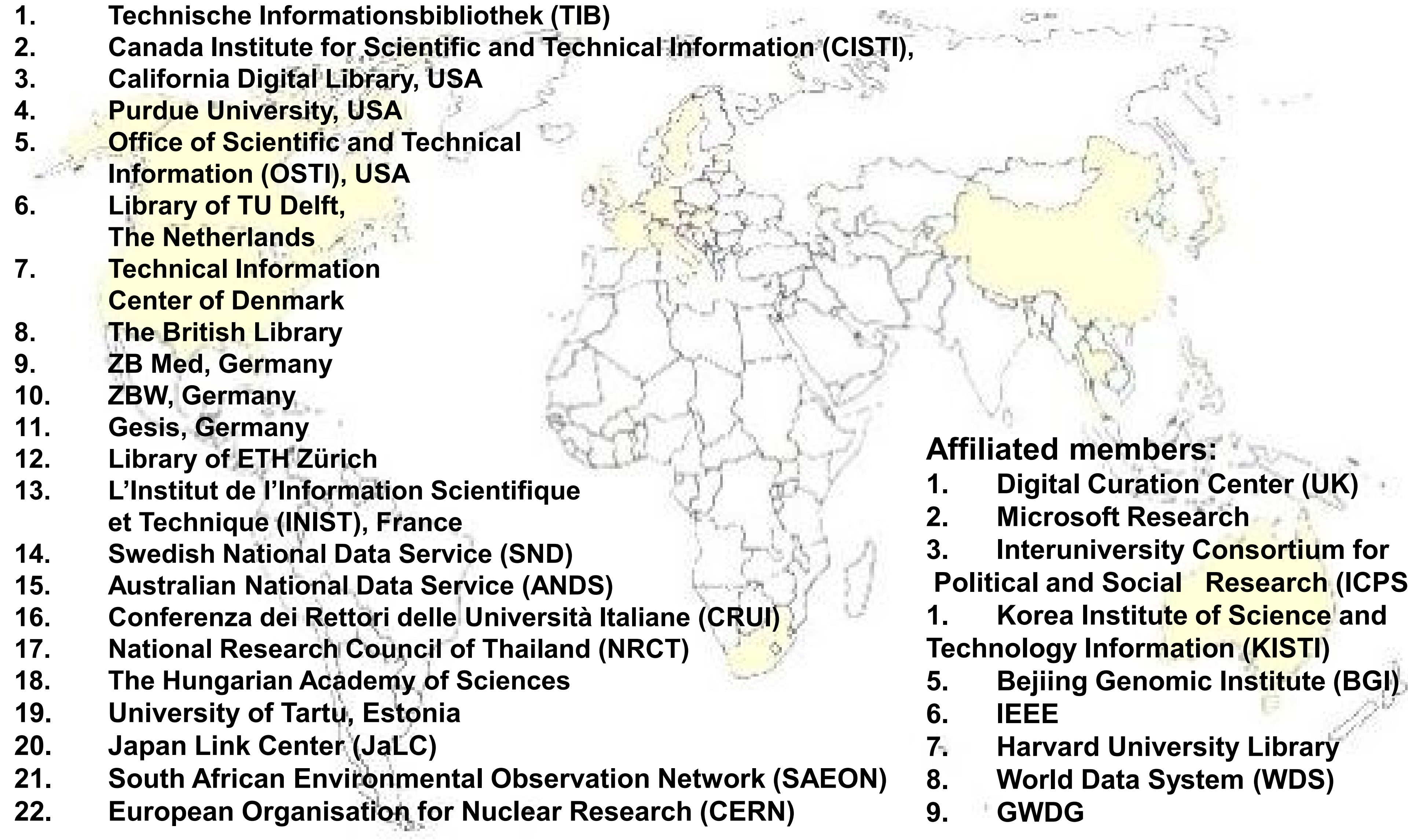
HTTP 404 - File not found
Internet Explorer

Digital Object Identifiers (DOI names) offer a solution

- Mostly widely used identifier for scientific articles
- Researchers, authors, publishers know how to use them
- Put datasets on the same playing field as articles

Dataset
Yancheva et al (2007). Analyses on sediment of Lake Maar. PANGAEA. doi:10.1594/PANGAEA.587840

DataCite members

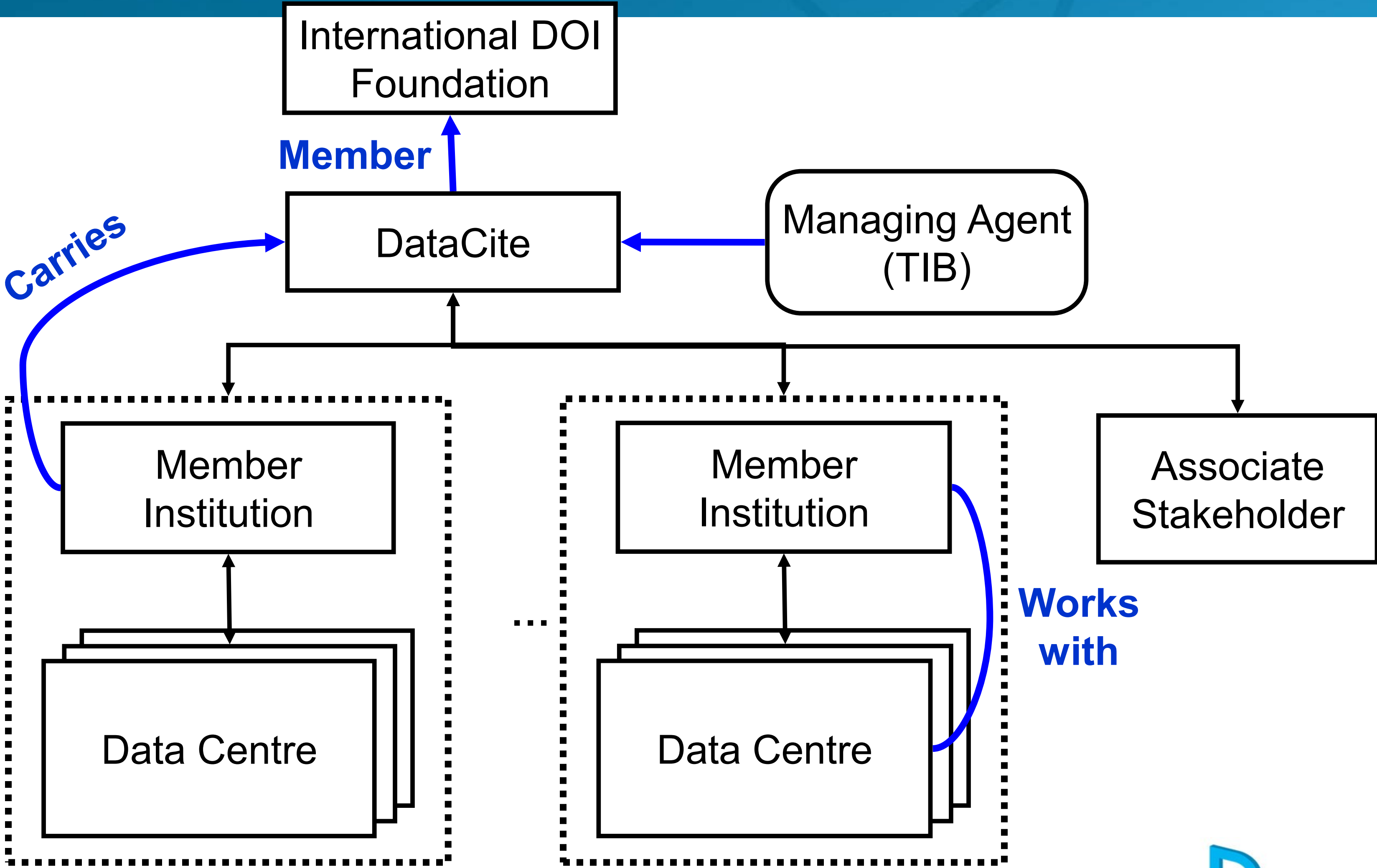


1. Technische Informationsbibliothek (TIB)
2. Canada Institute for Scientific and Technical Information (CISTI),
3. California Digital Library, USA
4. Purdue University, USA
5. Office of Scientific and Technical Information (OSTI), USA
6. Library of TU Delft, The Netherlands
7. Technical Information Center of Denmark
8. The British Library
9. ZB Med, Germany
10. ZBW, Germany
11. Gesis, Germany
12. Library of ETH Zürich
13. L'Institut de l'Information Scientifique et Technique (INIST), France
14. Swedish National Data Service (SND)
15. Australian National Data Service (ANDS)
16. Conferenza dei Rettori delle Università Italiane (CRUI)
17. National Research Council of Thailand (NRCT)
18. The Hungarian Academy of Sciences
19. University of Tartu, Estonia
20. Japan Link Center (JaLC)
21. South African Environmental Observation Network (SAEON)
22. European Organisation for Nuclear Research (CERN)

Affiliated members:

1. Digital Curation Center (UK)
2. Microsoft Research
3. Interuniversity Consortium for Political and Social Research (ICPS
1. Korea Institute of Science and Technology Information (KISTI)
5. Bejiing Genomic Institute (BGI)
6. IEEE
7. Harvard University Library
8. World Data System (WDS)
9. GWDG

DataCite structure



DataCite

DataCite in 2014

Over 3,200,000 DOI names registered so far.

290 data centers.

8,000,000 resolutions in 2013.

DataCite Metadata schema published (in cooperation with all members) http://schema.datacite.org

DataCite MetadataStore

http://search.datacite.org

OAI and Statistics

OAI Harvester

- http://oai.datacite.org

DataCite statistics (resolution and registration)

- http://stats.datacite.org

ODIN project

ORCID and DataCite interoperability network. Funded under FP7

http://www.odin-project.eu

Claim your DataCite DOI with your ORCID profile:

http://datacite.labs.orcid-eu.org/

2012: STM, CrossRef and DataCite Joint Statement

1. To improve the availability and findability of research data, the signers encourage authors of research papers to **deposit researcher validated data in trustworthy and reliable Data Archives**.

2. The Signers encourage Data Archives to **enable bi-directional linking between datasets and publications** by using established and community endorsed unique persistent identifiers such as database accession codes and DOI's.

3. The Signers encourage publishers and data archives to make visible or increase **visibility of these links** from publications to datasets and vice versa

DataCite

Example

The dataset:

Storz, D et al. (2009):

Planktic foraminiferal flux and faunal composition of sediment trap L1_K276 in the northeastern Atlantic.

http://dx.doi.org/10.1594/PANGAEA.724325

Is supplement to the article:

Storz, David; Schulz, Hartmut; Waniek, Joanna J; Schulz-Bull, Detlef; Kucera, Michal (2009): *Seasonal and interannual variability of the planktic foraminiferal flux in the vicinity of the Azores Current.*

Deep-Sea Research Part I-Oceanographic Research Papers, **56(1),** 107-124,

http://dx.doi.org/10.1016/j.dsr.2008.08.009

Cooperation

MoU with ORCID

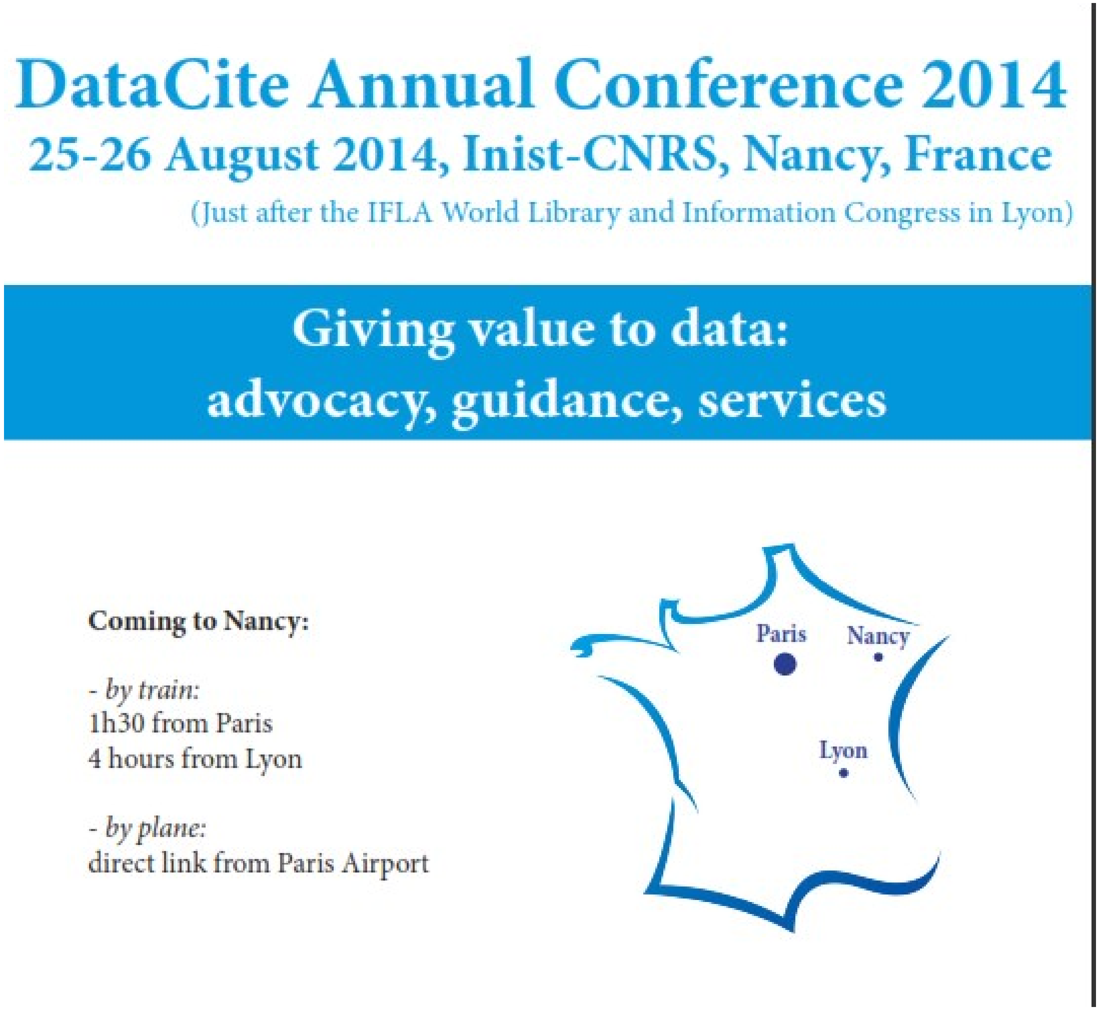
Agreement with Re3Data and DataBib to include their service in 2016

MoU with RDA to become organisational affiliate

Joint Declaration of Data Citation Principles

https://www.force11.org/datacitation

2014 Annual conference



DataCite Annual Conference 2014

25-26 August 2014, Inist-CNRS, Nancy, France

(Just after the IFLA World Library and Information Congress in Lyon)

**Giving value to data:
advocacy, guidance, services**

Coming to Nancy:

- *by train:*
1h30 from Paris
4 hours from Lyon

- *by plane:*
direct link from Paris Airport

Paris
Nancy
Lyon

Thank you!

ZB MED Knowledge for Life

gesis Leibniz Institute for the Social Sciences

ZBW Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

TIB Technische Informationsbibliothek

สำนักงาน
คณะกรรมการวิจัยแห่งชาติ
วช. NRCT
National Research Council of Thailand

CERN

NRC·CNRC CISTI ICIST

inist

TU Delft

VE RI TAS

BRITISH LIBRARY

SAEON South African Environmental Observation Network

IEEE

MAGYAR TUDOMÁNYOS AKADÉMIA 1825

TARTU ÜLIKOOL UNIVERSITAS TARTUENSIS 1632

DTU

GWDG Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

KiSTi www.kisti.re.kr

SND Swedish National Data Service

华大基因 BGI

ETH Eidgenössische Technische Hochschule Zürich Swiss Federal Institute of Technology Zurich

University of California

Japan Link Center

Welcome to Japan Link Center, JaLC!

ICSU WORLD DATA SYSTEM

osti.gov

PURDUE UNIVERSITY LIBRARIES

DCC

ands AUSTRALIAN NATIONAL DATA SERVICE

ICPSR A PARTNER IN SOCIAL SCIENCE RESEARCH INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

Microsoft Research

CRUI Conferenza dei Rettori delle Università Italiane

Data analysis issues and frameworks

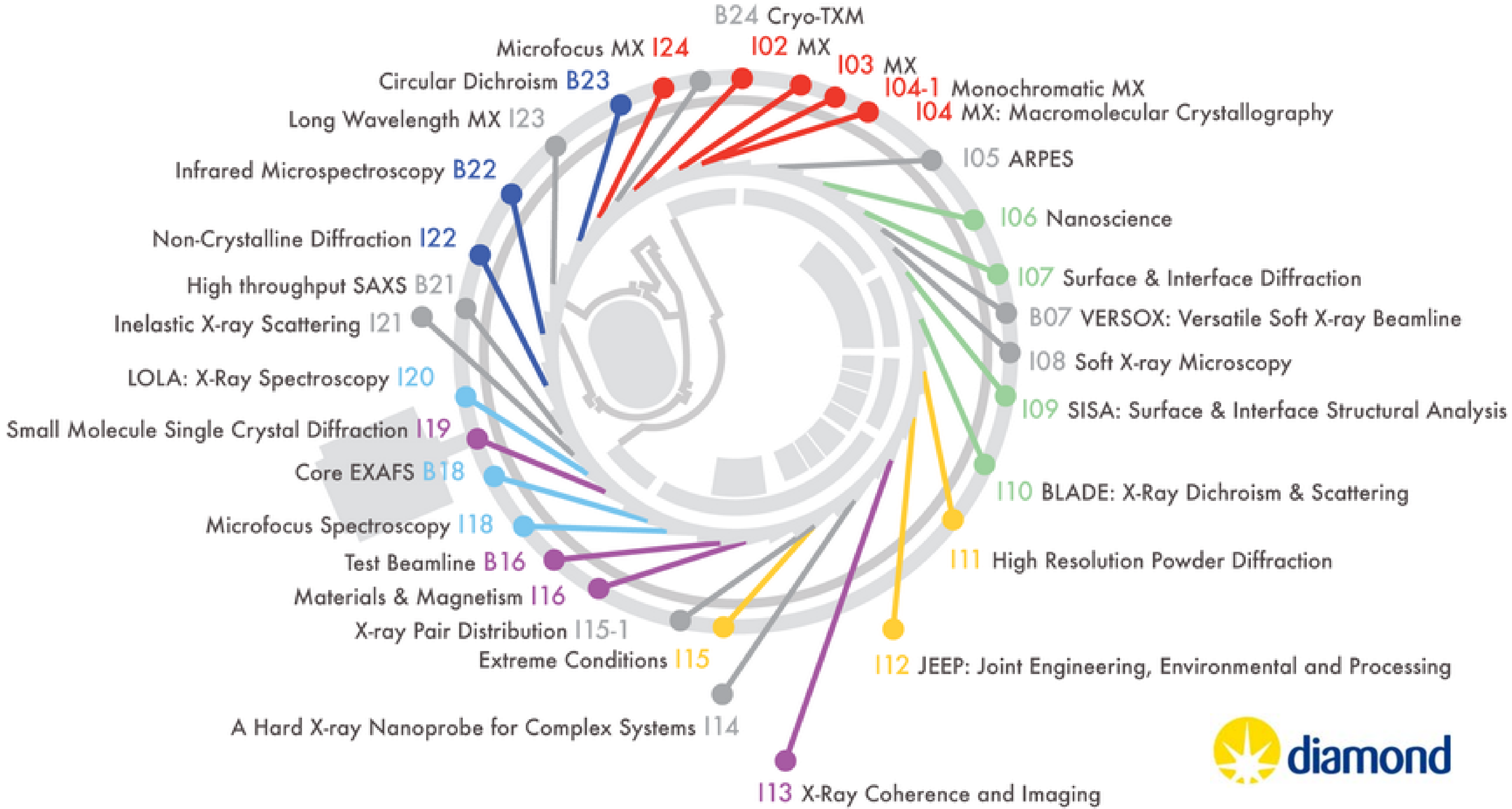
Alun Ashton

Group Leader, Data Analysis Software

Diamond Light Source

- EM Facility
- SFX XFEL UK Hub

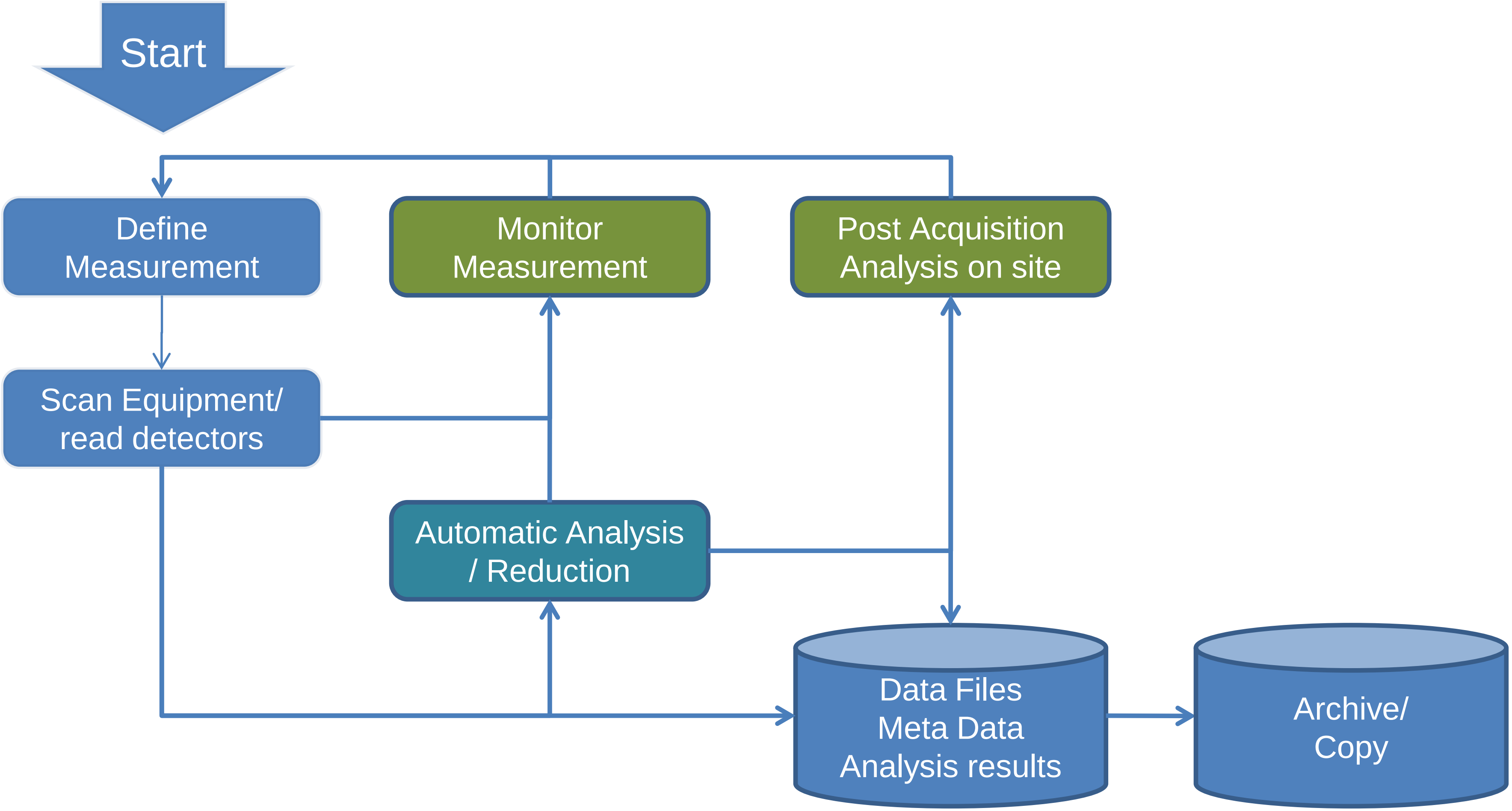
Data Analysis – Experimental Challenges Diversity: 33 beamlines > 120 experiment techniques


B24 Cryo-TXM
Microfocus MX I24
I02 MX
I03 MX
I04-1 Monochromatic MX
I04 MX: Macromolecular Crystallography
Circular Dichroism B23
Long Wavelength MX I23
I05 ARPES
Infrared Microspectroscopy B22
I06 Nanoscience
Non-Crystalline Diffraction I22
I07 Surface & Interface Diffraction
High throughput SAXS B21
B07 VERSOX: Versatile Soft X-ray Beamline
Inelastic X-ray Scattering I21
I08 Soft X-ray Microscopy
LOLA: X-Ray Spectroscopy I20
I09 SISA: Surface & Interface Structural Analysis
Small Molecule Single Crystal Diffraction I19
Core EXAFS B18
I10 BLADE: X-Ray Dichroism & Scattering
Microfocus Spectroscopy I18
I11 High Resolution Powder Diffraction
Test Beamline B16
Materials & Magnetism I16
X-ray Pair Distribution I15-1
Extreme Conditions I15
I12 JEEP: Joint Engineering, Environmental and Processing
A Hard X-ray Nanoprobe for Complex Systems I14
I13 X-Ray Coherence and Imaging
diamond

Data Analysis – Experimental Challenges

BEFORE	**IMMEDIATE**	**SHORT TERM**	**LONG TERM**
(FROM DLS/USER'S INSTITUTION)	(DURING EXPERIMENTS)	(BEFORE THE USER GOES HOME)	(FROM DLS/USER'S INSTITUTION)
Simulations Processing of older datasets	"Real time" data processing, analysis and visualisation – to make experimental decisions	Data reduction and processing – Users go home with clean data free of instrument artefacts. Preliminary data analysis – helpful, but may require significant processing power and know-how	Detailed analysis – from data to information. Incorporating results from other techniques. Experiments: ➢Provide parameters for a model. ➢Test/verify a model or theory. ➢Show where a new theory or model is required.

Data Analysis – Experimental Challenges


Start
Define Measurement
Monitor Measurement
Post Acquisition Analysis on site
Scan Equipment/ read detectors
Automatic Analysis / Reduction
Data Files Meta Data Analysis results
Archive/ Copy

The ‘Customer’

Data Analysis – Experimental Challenges

How facility users want to analyse their data (sample):

- Command line, interactive analysis
- ‘scripts’
- Black box
- Wizards
- GUI
- Automatically
- Someone else do it
- Don't you look at it

- Quickly
- Estimates
- Properly
- Publication quality
- The way I used to do it
- Use the new stuff
- On all the data
- On some of the data

Data Analysis – Experimental Challenges

How facility users want to analyse their data (sample):

- Only when I am at the facility
- When I am at home
- On a web page
- On my laptop
- On your computers

- Yesterday
- Next year
- By the student
- By the expert
- Securely
- Shared data
- ASCII not binary...

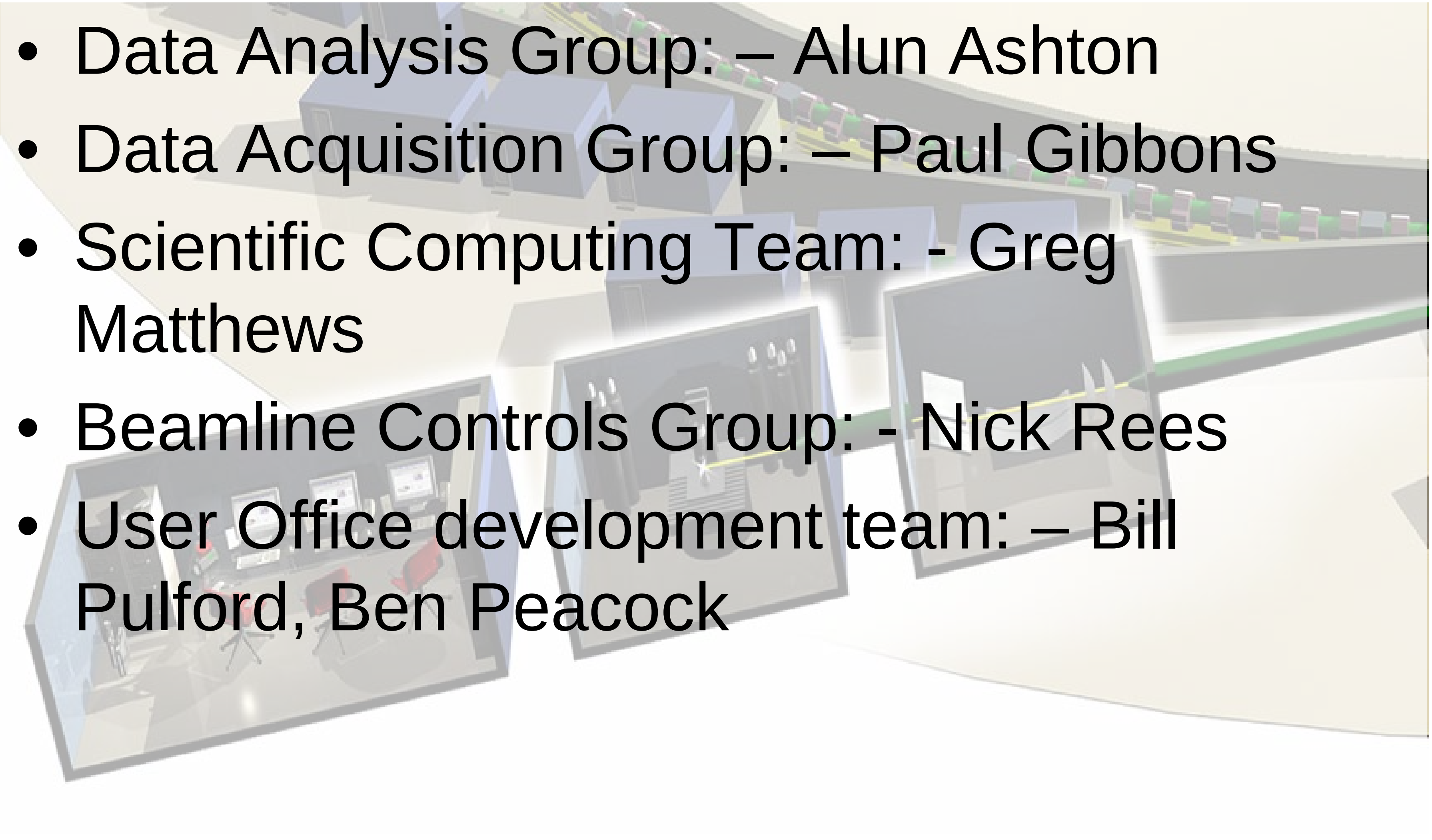
Other Factors

- Existing codes and expertise
- Plethora of data structures/file formats/data rates
- Available computing resources
  - Enough computing
  - Quick access to Data
- Information management
- Collaborations

How we are dealing with the problem in Diamond

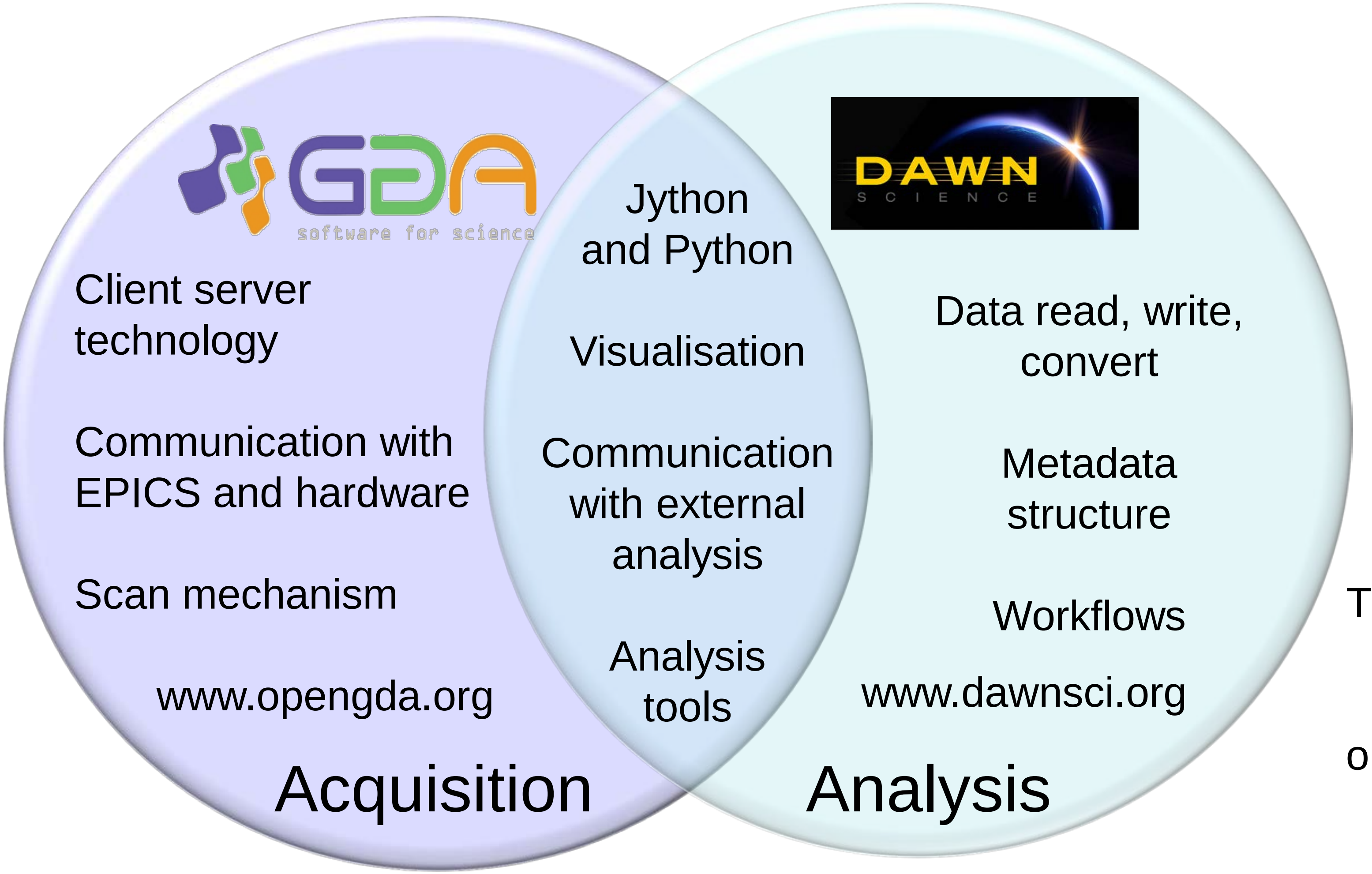
examples

Beamline Computing/Software Support



- Data Analysis Group: – Alun Ashton
- Data Acquisition Group: – Paul Gibbons
- Scientific Computing Team: - Greg Matthews
- Beamline Controls Group: - Nick Rees
- User Office development team: – Bill Pulford, Ben Peacock

Acquisition and Analysis in the Eclipse Framework
(www.eclipse.org)



Client server technology

Communication with EPICS and hardware

Scan mechanism

www.opengda.org

Acquisition

Jython and Python

Visualisation

Communication with external analysis

Analysis tools

Data read, write, convert

Metadata structure

Workflows

www.dawnsci.org

Analysis

DAWN is a collection of generic and bespoke ‘views’ collated into ‘perspectives’.

The perspectives and views can be used in part or whole in either the GDA or DAWN.

All core technologies open source

diamond

An open source collaboration with a core of generic tools with local extensions and implementations.

- Visualisation of scientific data
- Workflow tool for the treatment/manipulation of scientific data
- Integrated python environment

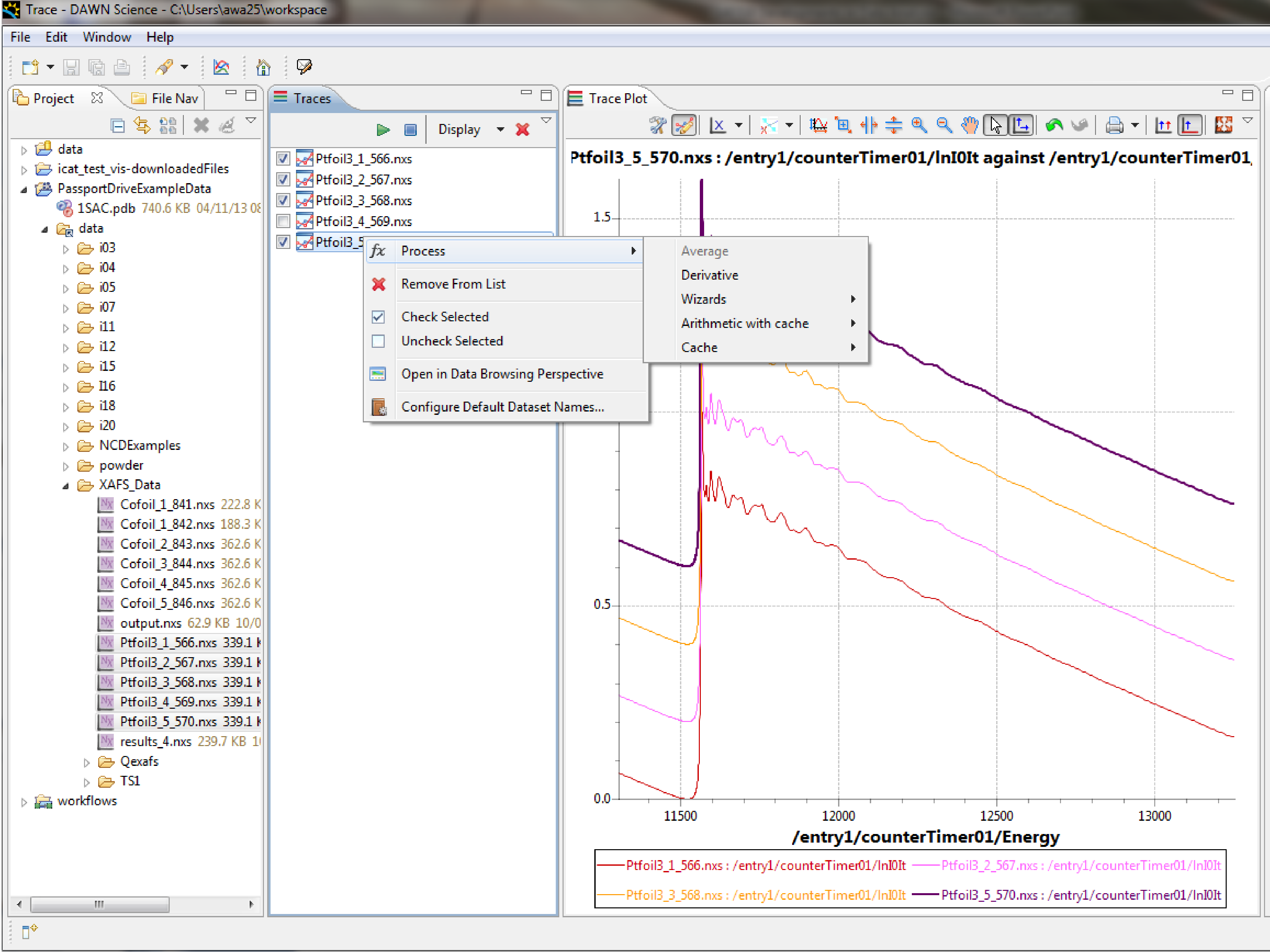
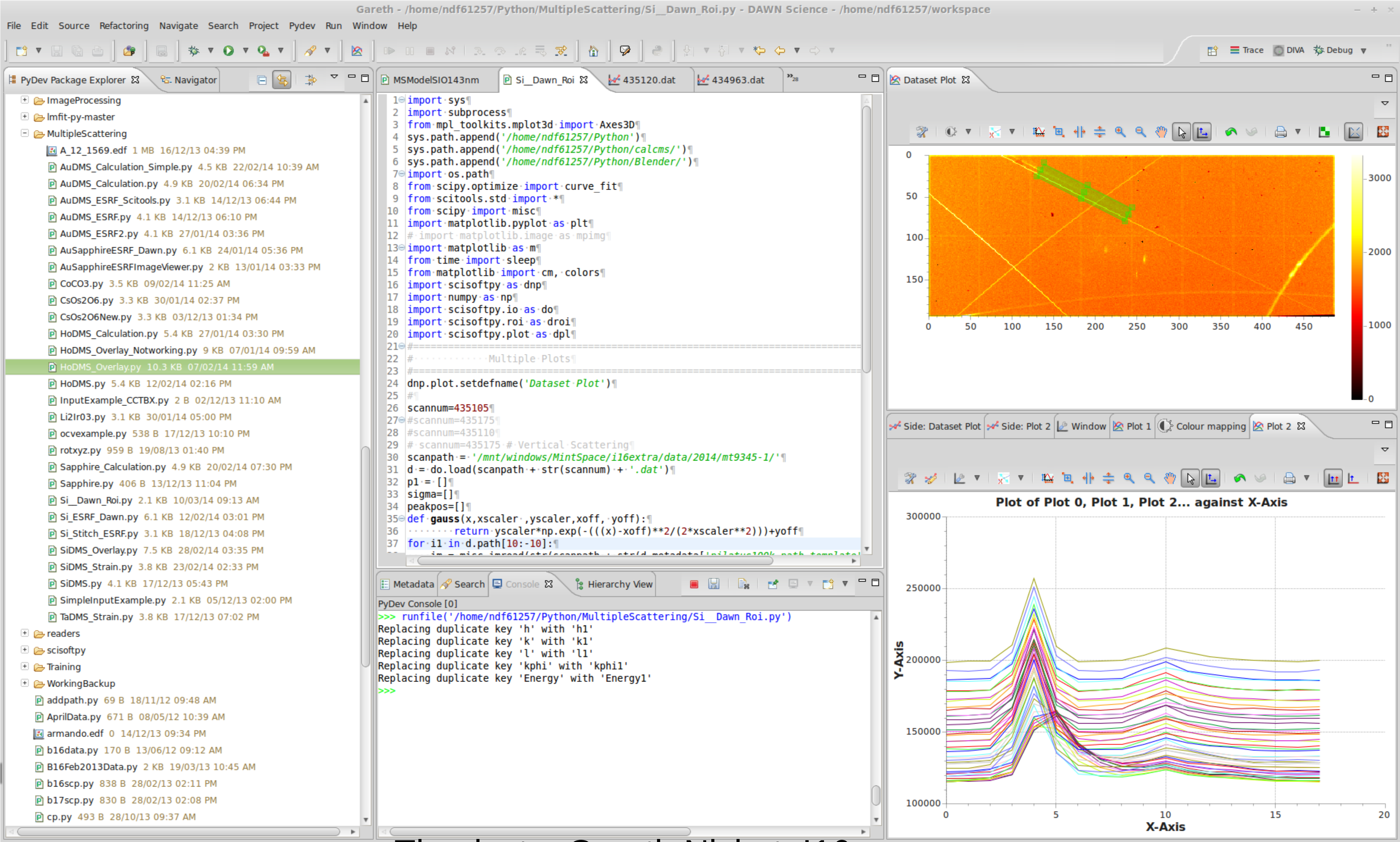

Trace - DAWN Science - C:\Users\awa25\workspace
File
Edit
Window
Help
Project
File Nav
data
icat_test_vis-downloadedFiles
PassportDriveExampleData
1SAC.pdb 740.6 KB 04/11/13 08
data
i03
i04
i05
i07
i11
i12
i15
I16
i18
i20
NCDExamples
powder
XAFS_Data
Cofoil_1_841.nxs 222.8 K
Cofoil_1_842.nxs 188.3 K
Cofoil_2_843.nxs 362.6 K
Cofoil_3_844.nxs 362.6 K
Cofoil_4_845.nxs 362.6 K
Cofoil_5_846.nxs 362.6 K
output.nxs 62.9 KB 10/0
Ptfoil3_1_566.nxs 339.1 K
Ptfoil3_2_567.nxs 339.1 K
Ptfoil3_3_568.nxs 339.1 K
Ptfoil3_4_569.nxs 339.1 K
Ptfoil3_5_570.nxs 339.1 K
results_4.nxs 239.7 KB 1(
Qexafs
TS1
workflows
Traces
Display
☑ Ptfoil3_1_566.nxs
☑ Ptfoil3_2_567.nxs
☑ Ptfoil3_3_568.nxs
□ Ptfoil3_4_569.nxs
☑ Ptfoil3_5
Process
Remove From List
Check Selected
Uncheck Selected
Open in Data Browsing Perspective
Configure Default Dataset Names...
Average
Derivative
Wizards
Arithmetic with cache
Cache
Trace Plot
Ptfoil3_5_570.nxs : /entry1/counterTimer01/lnI0It against /entry1/counterTimer01,
1.5
0.5
0.0
11500
12000
12500
13000
/entry1/counterTimer01/Energy
Ptfoil3_1_566.nxs : /entry1/counterTimer01/lnI0It
Ptfoil3_2_567.nxs : /entry1/counterTimer01/lnI0It
Ptfoil3_3_568.nxs : /entry1/counterTimer01/lnI0It
Ptfoil3_5_570.nxs : /entry1/counterTimer01/lnI0It

Image stack analysis



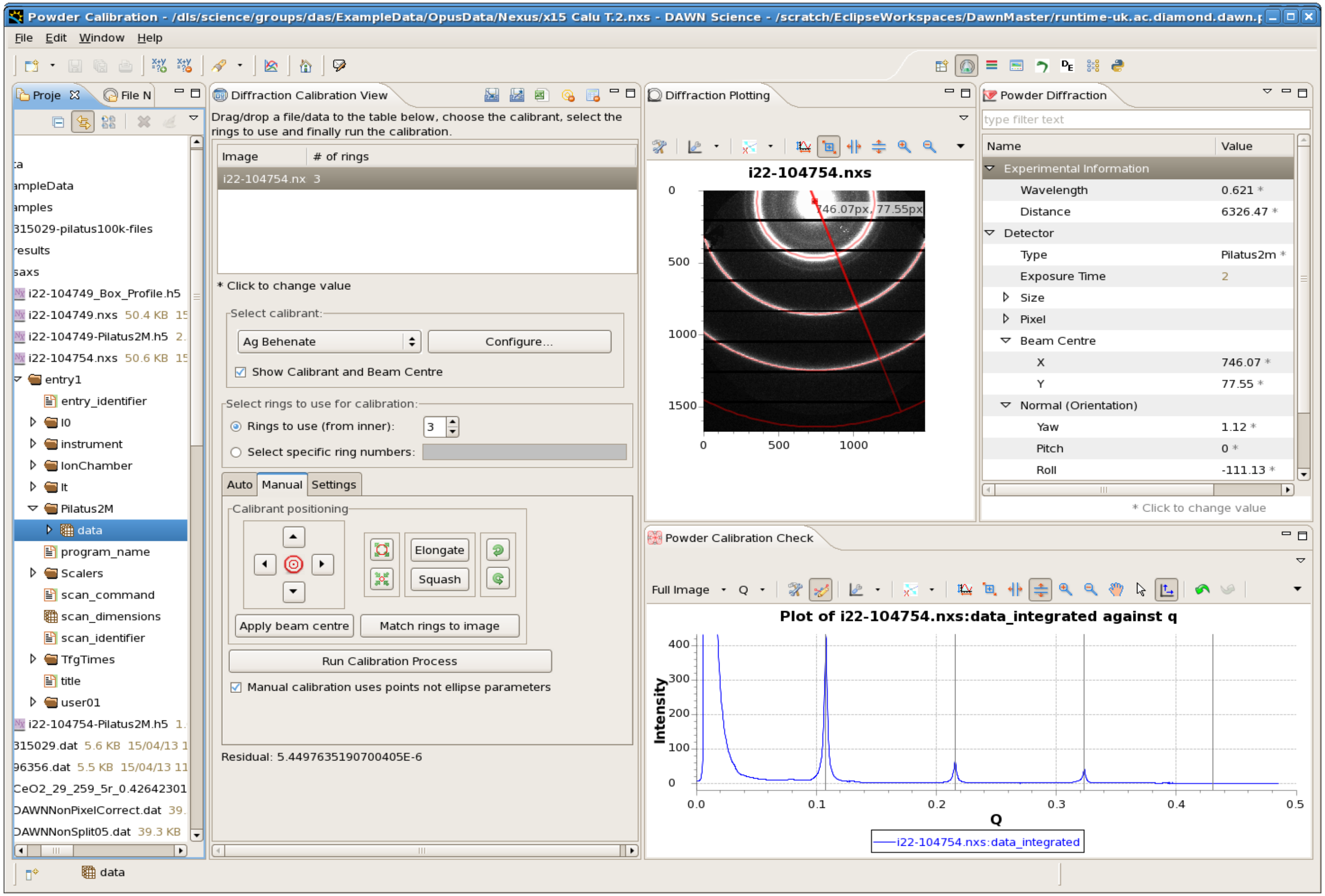
ImageProcessing		
lmfit-py-master		
MultipleScattering		
A_12_1569.edf	1 MB	16/12/13 04:39 PM
AuDMS_Calculation_Simple.py	4.5 KB	22/02/14 10:39 AM
AuDMS_Calculation.py	4.9 KB	20/02/14 06:34 PM
AuDMS_ESRF_Scitools.py	3.1 KB	14/12/13 06:44 PM
AuDMS_ESRF.py	4.1 KB	14/12/13 06:10 PM
AuDMS_ESRF2.py	4.1 KB	27/01/14 03:36 PM
AuSapphireESRF_Dawn.py	6.1 KB	24/01/14 05:36 PM
AuSapphireESRFImageViewer.py	2 KB	13/01/14 03:33 PM
CoCO3.py	3.5 KB	09/02/14 11:25 AM
CsOs2O6.py	3.3 KB	30/01/14 02:37 PM
CsOs2O6New.py	3.3 KB	03/12/13 01:34 PM
HoDMS_Calculation.py	5.4 KB	27/01/14 03:30 PM
HoDMS_Overlay_Notworking.py	9 KB	07/01/14 09:59 AM
HoDMS_Overlay.py	10.3 KB	07/02/14 11:59 AM
HoDMS.py	5.4 KB	12/02/14 02:16 PM
InputExample_CCTBX.py	2 B	02/12/13 11:10 AM
Li2Ir03.py	3.1 KB	30/01/14 05:00 PM
ocvexample.py	538 B	17/12/13 10:10 PM
rotxyz.py	959 B	19/08/13 01:40 PM
Sapphire_Calculation.py	4.9 KB	20/02/14 07:30 PM
Sapphire.py	406 B	13/12/13 11:04 PM
Si_Dawn_Roi.py	2.1 KB	10/03/14 09:13 AM
Si_ESRF_Dawn.py	6.1 KB	12/02/14 03:01 PM
Si_Stitch_ESRF.py	3.1 KB	18/12/13 04:08 PM
SiDMS_Overlay.py	7.5 KB	28/02/14 03:35 PM
SiDMS_Strain.py	3.8 KB	23/02/14 02:33 PM
SiDMS.py	4.1 KB	17/12/13 05:43 PM
SimpleInputExample.py	2.1 KB	05/12/13 02:00 PM
TaDMS_Strain.py	3.8 KB	17/12/13 07:02 PM
readers		
scisoftpy		
Training		
WorkingBackup		
addpath.py	69 B	18/11/12 09:48 AM
AprilData.py	671 B	08/05/12 10:39 AM
armando.edf	0	14/12/13 09:34 PM
b16data.py	170 B	13/06/12 09:12 AM
B16Feb2013Data.py	2 KB	19/03/13 10:45 AM
b16scp.py	838 B	28/02/13 02:11 PM
b17scp.py	830 B	28/02/13 02:08 PM
cp.py	493 B	28/10/13 09:37 AM

```python
import sys
import subprocess
from mpl_toolkits.mplot3d import Axes3D
sys.path.append('/home/ndf61257/Python')
sys.path.append('/home/ndf61257/Python/calcms/')
sys.path.append('/home/ndf61257/Python/Blender/')
import os.path
from scipy.optimize import curve_fit
from scitools.std import *
from scipy import misc
import matplotlib.pyplot as plt
# import matplotlib.image as mpimg
import matplotlib as m
from time import sleep
from matplotlib import cm, colors
import scisoftpy as dnp
import numpy as np
import scisoftpy.io as do
import scisoftpy.roi as droi
import scisoftpy.plot as dpl
#===========================================================================
#            Multiple Plots
#===========================================================================
dnp.plot.setdefname('Dataset Plot')
#
scannum=435105
#scannum=435175
#scannum=435110
# scannum=435175 # Vertical Scattering
scanpath = '/mnt/windows/MintSpace/i16extra/data/2014/mt9345-1/'
d = do.load(scanpath + str(scannum) + '.dat')
p1 = []
sigma=[]
peakpos=[]
def gauss(x,xscaler ,yscaler,xoff, yoff):
        return yscaler*np.exp(-(((x)-xoff)**2/(2*xscaler**2)))+yoff
for i1 in d.path[10:-10]:
```

```
>>> runfile('/home/ndf61257/Python/MultipleScattering/Si__Dawn_Roi.py')
Replacing duplicate key 'h' with 'h1'
Replacing duplicate key 'k' with 'k1'
Replacing duplicate key 'l' with 'l1'
Replacing duplicate key 'kphi' with 'kphi1'
Replacing duplicate key 'Energy' with 'Energy1'
>>>
```

Thanks to: Gareth Nisbet, I16

2D Diffraction Processing



Powder Calibration - /dls/science/groups/das/ExampleData/OpusData/Nexus/x15 Calu T.2.nxs - DAWN Science - /scratch/EclipseWorkspaces/DawnMaster/runtime-uk.ac.diamond.dawn.p

Drag/drop a file/data to the table below, choose the calibrant, select the rings to use and finally run the calibration.

Image	# of rings
i22-104754.nx	3

* Click to change value

Select calibrant: Ag Behenate

☑ Show Calibrant and Beam Centre

Select rings to use for calibration:

● Rings to use (from inner): 3

○ Select specific ring numbers:

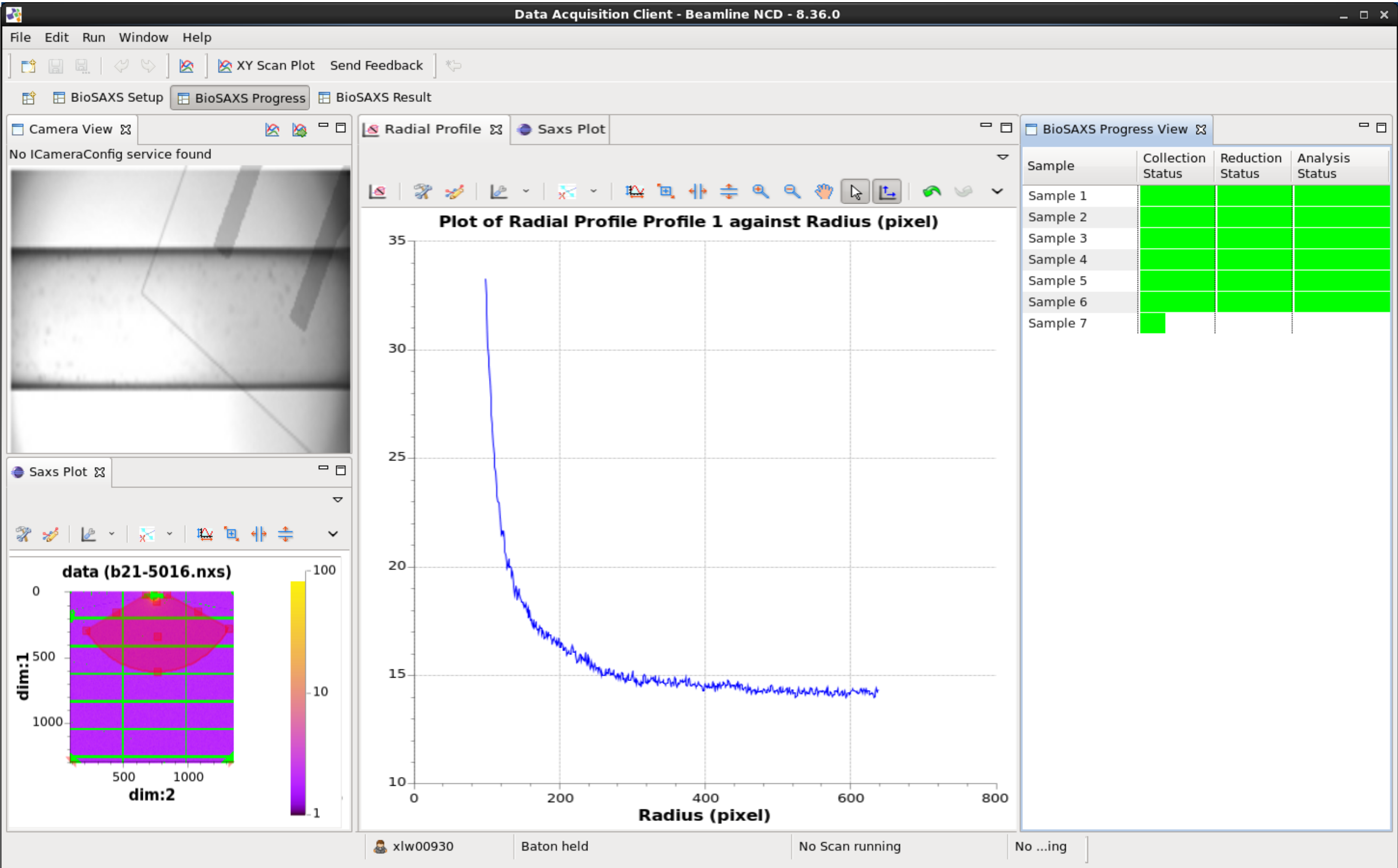
☑ Manual calibration uses points not ellipse parameters

Residual: 5.4497635190700405E-6

Name	Value
Experimental Information	
Wavelength	0.621 *
Distance	6326.47 *
Detector	
Type	Pilatus2m *
Exposure Time	2
Size	
Pixel	
Beam Centre	
X	746.07 *
Y	77.55 *
Normal (Orientation)	
Yaw	1.12 *
Pitch	0 *
Roll	-111.13 *

* Click to change value

BioSAXS experiment



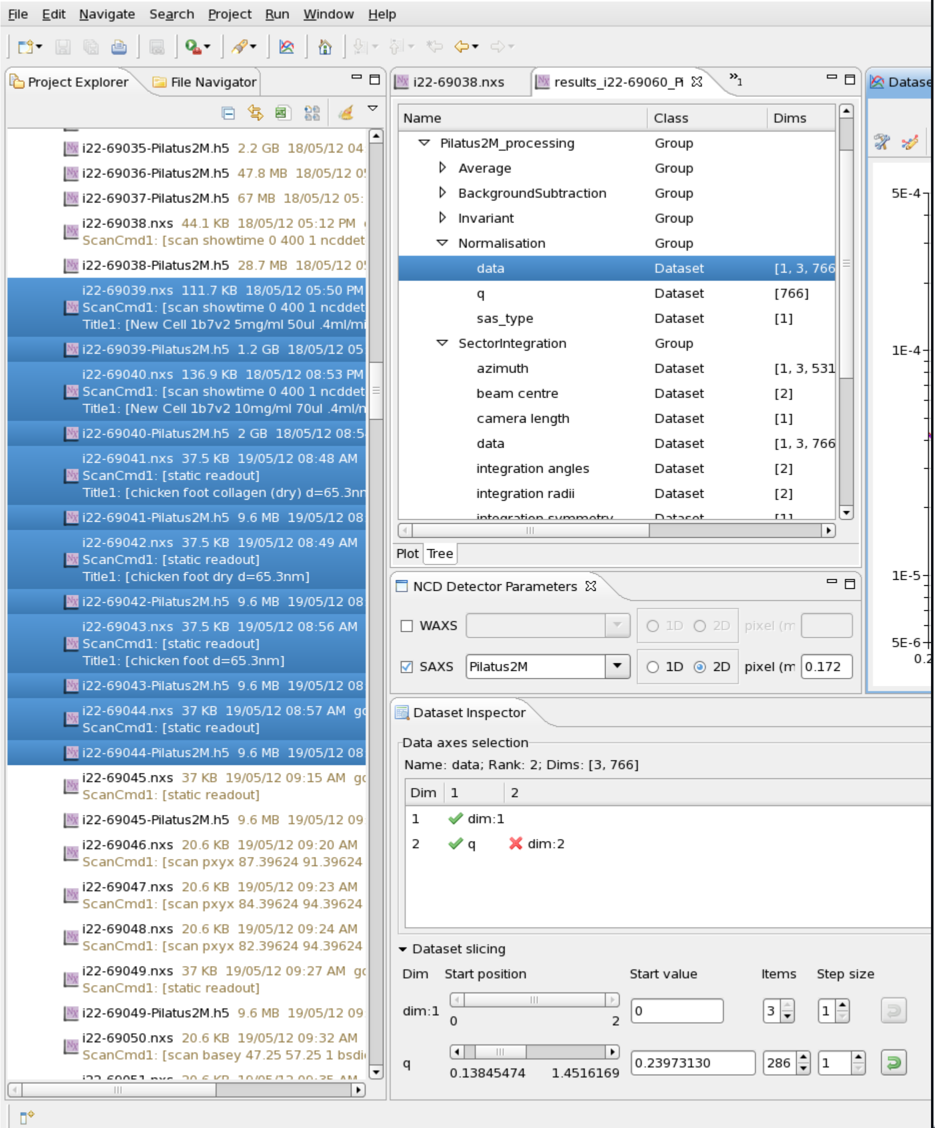
Data Acquisition Client - Beamline NCD - 8.36.0

No ICameraConfig service found

Sample	Collection Status	Reduction Status	Analysis Status
Sample 1	🟩	🟩	🟩
Sample 2	🟩	🟩	🟩
Sample 3	🟩	🟩	🟩
Sample 4	🟩	🟩	🟩
Sample 5	🟩	🟩	🟩
Sample 6	🟩	🟩	🟩
Sample 7	🟩		

xlw00930 | Baton held | No Scan running | No ...ing

SAXS Data Reduction



i22-69035-Pilatus2M.h5 2.2 GB 18/05/12 04
i22-69036-Pilatus2M.h5 47.8 MB 18/05/12 0
i22-69037-Pilatus2M.h5 67 MB 18/05/12 05:
i22-69038.nxs 44.1 KB 18/05/12 05:12 PM
ScanCmd1: [scan showtime 0 400 1 ncddet
i22-69038-Pilatus2M.h5 28.7 MB 18/05/12 0
i22-69039.nxs 111.7 KB 18/05/12 05:50 PM
ScanCmd1: [scan showtime 0 400 1 ncddet
Title1: [New Cell 1b7v2 5mg/ml 50ul .4ml/m
i22-69039-Pilatus2M.h5 1.2 GB 18/05/12 05
i22-69040.nxs 136.9 KB 18/05/12 08:53 PM
ScanCmd1: [scan showtime 0 400 1 ncddet
Title1: [New Cell 1b7v2 10mg/ml 70ul .4ml/n
i22-69040-Pilatus2M.h5 2 GB 18/05/12 08:5
i22-69041.nxs 37.5 KB 19/05/12 08:48 AM
ScanCmd1: [static readout]
Title1: [chicken foot collagen (dry) d=65.3nn
i22-69041-Pilatus2M.h5 9.6 MB 19/05/12 08
i22-69042.nxs 37.5 KB 19/05/12 08:49 AM
ScanCmd1: [static readout]
Title1: [chicken foot dry d=65.3nm]
i22-69042-Pilatus2M.h5 9.6 MB 19/05/12 08
i22-69043.nxs 37.5 KB 19/05/12 08:56 AM
ScanCmd1: [static readout]
Title1: [chicken foot d=65.3nm]
i22-69043-Pilatus2M.h5 9.6 MB 19/05/12 08
i22-69044.nxs 37 KB 19/05/12 08:57 AM g
ScanCmd1: [static readout]
i22-69044-Pilatus2M.h5 9.6 MB 19/05/12 08
i22-69045.nxs 37 KB 19/05/12 09:15 AM g
ScanCmd1: [static readout]
i22-69045-Pilatus2M.h5 9.6 MB 19/05/12 09
i22-69046.nxs 20.6 KB 19/05/12 09:20 AM
ScanCmd1: [scan pxyx 87.39624 91.39624
i22-69047.nxs 20.6 KB 19/05/12 09:23 AM
ScanCmd1: [scan pxyx 84.39624 94.39624
i22-69048.nxs 20.6 KB 19/05/12 09:24 AM
ScanCmd1: [scan pxyx 82.39624 94.39624
i22-69049.nxs 37 KB 19/05/12 09:27 AM g
ScanCmd1: [static readout]
i22-69049-Pilatus2M.h5 9.6 MB 19/05/12 09
i22-69050.nxs 20.6 KB 19/05/12 09:32 AM
ScanCmd1: [scan basey 47.25 57.25 1 bsdi

Name	Class	Dims
Pilatus2M_processing	Group	
Average	Group	
BackgroundSubtraction	Group	
Invariant	Group	
Normalisation	Group	
data	Dataset	[1, 3, 766
q	Dataset	[766]
sas_type	Dataset	[1]
SectorIntegration	Group	
azimuth	Dataset	[1, 3, 531
beam centre	Dataset	[2]
camera length	Dataset	[1]
data	Dataset	[1, 3, 766
integration angles	Dataset	[2]
integration radii	Dataset	[2]
integration symmetry	Dataset	[1]

NCD Detector Parameters

□ WAXS ○ 1D ○ 2D pixel (m

☑ SAXS Pilatus2M ○ 1D ● 2D pixel (m 0.172

Dataset Inspector

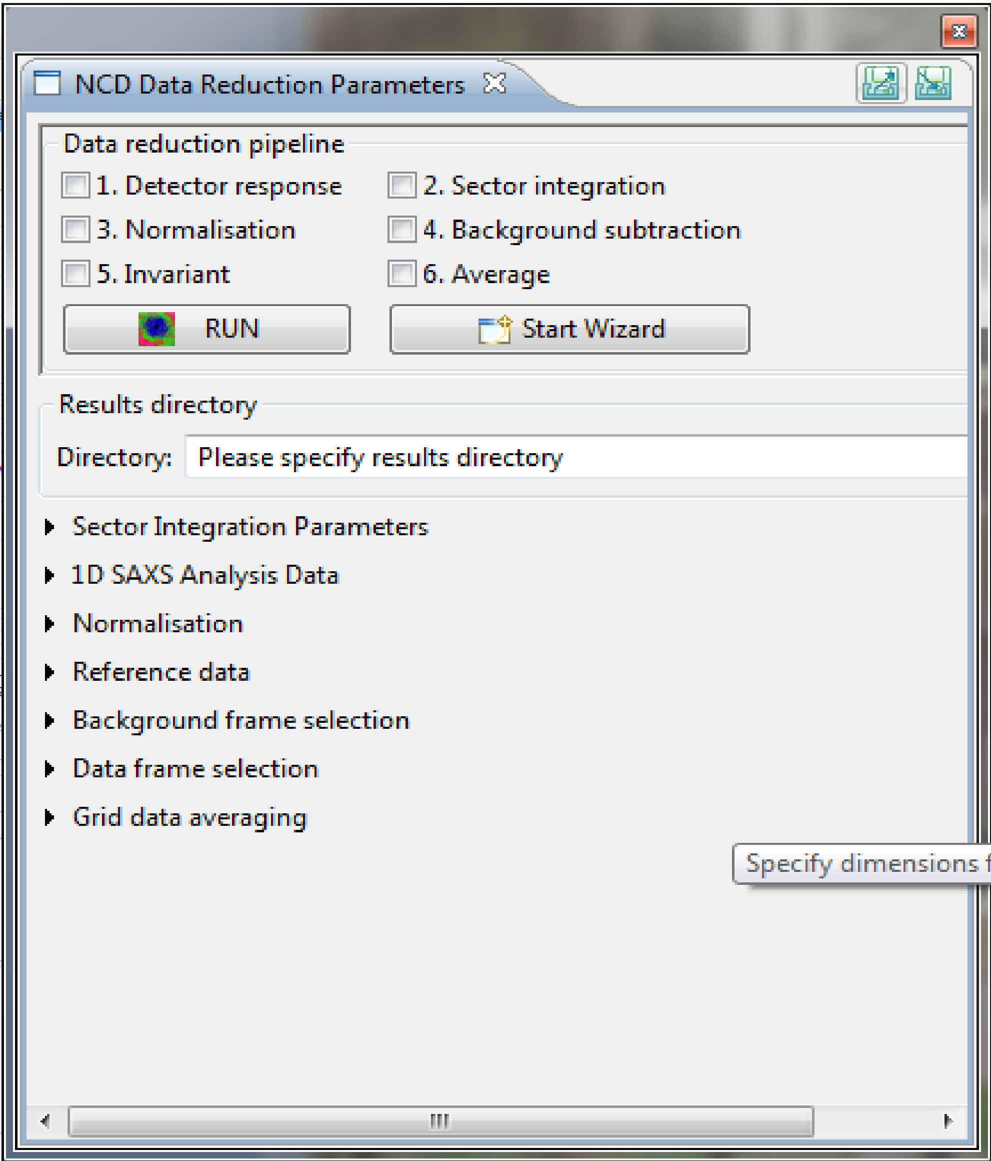
Data axes selection

Name: data; Rank: 2; Dims: [3, 766]

Dim	1	2
1	✓ dim:1	
2	✓ q	✗ dim:2

Dataset slicing

Dim	Start position	Start value	Items	Step size
dim:1	0 … 2	0	3	1
q	0.13845474 … 1.4516169	0.23973130	286	1



NCD Data Reduction Parameters

Data reduction pipeline

□ 1. Detector response □ 2. Sector integration

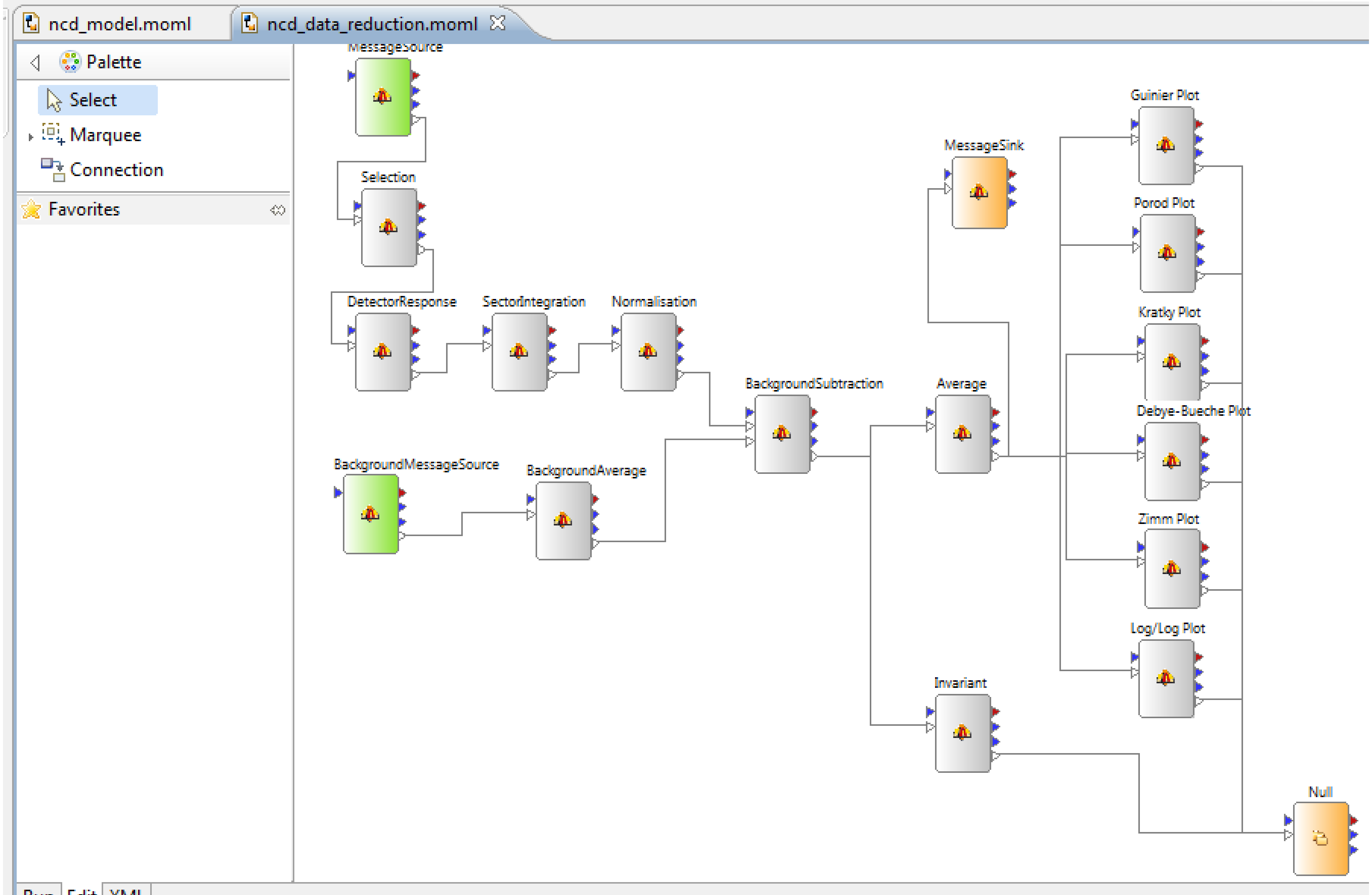
□ 3. Normalisation □ 4. Background subtraction

□ 5. Invariant □ 6. Average

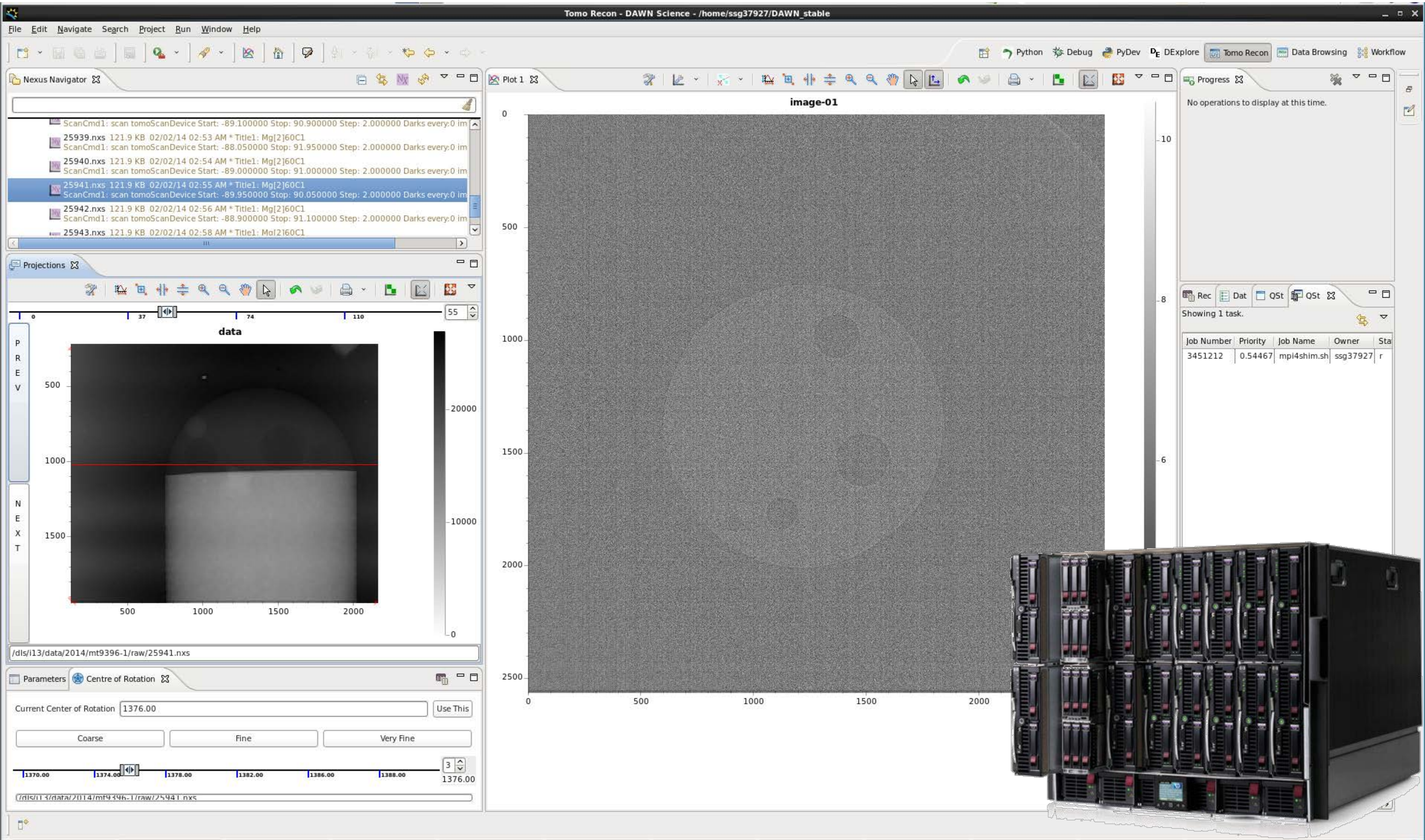
Results directory

Directory: Please specify results directory

Automated Data Reduction



Tomography Reconstruction GUI



Tomo Recon - DAWN Science - /home/ssg37927/DAWN_stable

ScanCmd1: scan tomoScanDevice Start: -89.100000 Stop: 90.900000 Step: 2.000000 Darks every:0 im
25939.nxs 121.9 KB 02/02/14 02:53 AM * Title1: Mg[2]60C1
ScanCmd1: scan tomoScanDevice Start: -88.050000 Stop: 91.950000 Step: 2.000000 Darks every:0 im
25940.nxs 121.9 KB 02/02/14 02:54 AM * Title1: Mg[2]60C1
ScanCmd1: scan tomoScanDevice Start: -89.000000 Stop: 91.000000 Step: 2.000000 Darks every:0 im
25941.nxs 121.9 KB 02/02/14 02:55 AM * Title1: Mg[2]60C1
ScanCmd1: scan tomoScanDevice Start: -89.950000 Stop: 90.050000 Step: 2.000000 Darks every:0 im
25942.nxs 121.9 KB 02/02/14 02:56 AM * Title1: Mg[2]60C1
ScanCmd1: scan tomoScanDevice Start: -88.900000 Stop: 91.100000 Step: 2.000000 Darks every:0 im
25943.nxs 121.9 KB 02/02/14 02:58 AM * Title1: Mg[2]60C1

/dls/i13/data/2014/mt9396-1/raw/25941.nxs

Current Center of Rotation: 1376.00

1376.00

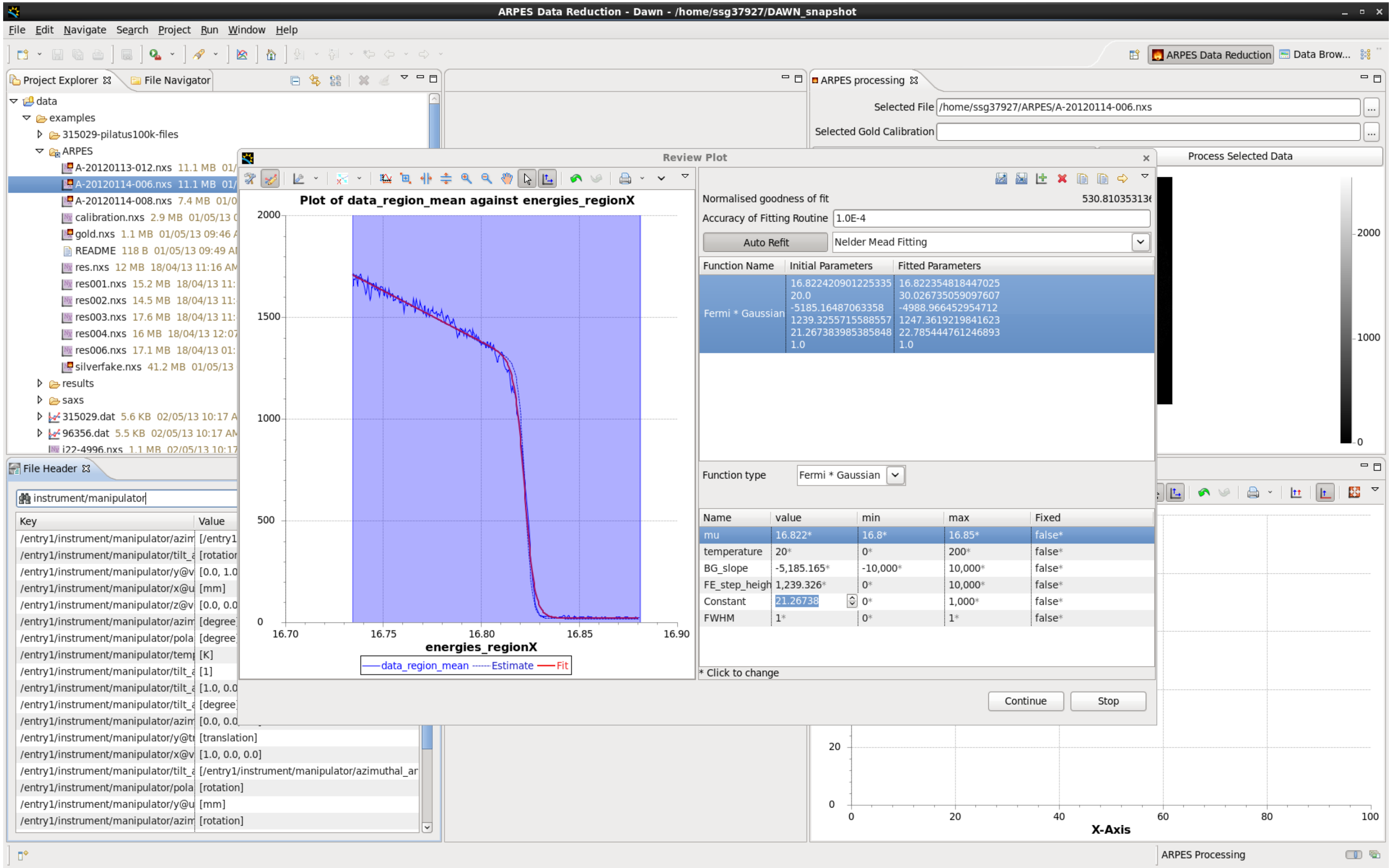
/dls/i13/data/2014/mt9396-1/raw/25941.nxs

No operations to display at this time.

Showing 1 task.

Job Number	Priority	Job Name	Owner	Sta
3451212	0.54467	mpi4shim.sh	ssg37927	r

ARPES Angle-Resolved Photo-Emission Spectroscopy



ARPES Data Reduction - Dawn - /home/ssg37927/DAWN_snapshot

- data
  - examples
    - 315029-pilatus100k-files
    - ARPES
      - A-20120113-012.nxs 11.1 MB 01/
      - A-20120114-006.nxs 11.1 MB 01/
      - A-20120114-008.nxs 7.4 MB 01/0
      - calibration.nxs 2.9 MB 01/05/13 0
      - gold.nxs 1.1 MB 01/05/13 09:46 A
      - README 118 B 01/05/13 09:49 AM
      - res.nxs 12 MB 18/04/13 11:16 AM
      - res001.nxs 15.2 MB 18/04/13 11:
      - res002.nxs 14.5 MB 18/04/13 11:
      - res003.nxs 17.6 MB 18/04/13 11:
      - res004.nxs 16 MB 18/04/13 12:07
      - res006.nxs 17.1 MB 18/04/13 01:
      - silverfake.nxs 41.2 MB 01/05/13
    - results
    - saxs
    - 315029.dat 5.6 KB 02/05/13 10:17 A
    - 96356.dat 5.5 KB 02/05/13 10:17 AM
    - i22-4996.nxs 1.1 MB 02/05/13 10:17

File Header

instrument/manipulator

Key	Value
/entry1/instrument/manipulator/azim	[/entry1
/entry1/instrument/manipulator/tilt_a	[rotatior
/entry1/instrument/manipulator/y@v	[0.0, 1.0
/entry1/instrument/manipulator/x@u	[mm]
/entry1/instrument/manipulator/z@v	[0.0, 0.0
/entry1/instrument/manipulator/azim	[degree
/entry1/instrument/manipulator/pola	[degree
/entry1/instrument/manipulator/temp	[K]
/entry1/instrument/manipulator/tilt_a	[1]
/entry1/instrument/manipulator/tilt_a	[1.0, 0.0
/entry1/instrument/manipulator/tilt_a	[degree
/entry1/instrument/manipulator/azim	[0.0, 0.0
/entry1/instrument/manipulator/y@tr	[translation]
/entry1/instrument/manipulator/x@v	[1.0, 0.0, 0.0]
/entry1/instrument/manipulator/tilt_a	[/entry1/instrument/manipulator/azimuthal_ar
/entry1/instrument/manipulator/pola	[rotation]
/entry1/instrument/manipulator/y@u	[mm]
/entry1/instrument/manipulator/azim	[rotation]

ARPES processing

Selected File /home/ssg37927/ARPES/A-20120114-006.nxs

Selected Gold Calibration

Process Selected Data

Review Plot

Plot of data_region_mean against energies_regionX

Normalised goodness of fit 530.81035313€

Accuracy of Fitting Routine 1.0E-4

Auto Refit | Nelder Mead Fitting

Function Name	Initial Parameters	Fitted Parameters
Fermi * Gaussian	16.822420901225335 20.0 -5185.16487063358 1239.3255715588557 21.267383985385848 1.0	16.822354818447025 30.026735059097607 -4988.966452954712 1247.3619219841623 22.785444761246893 1.0

Function type Fermi * Gaussian

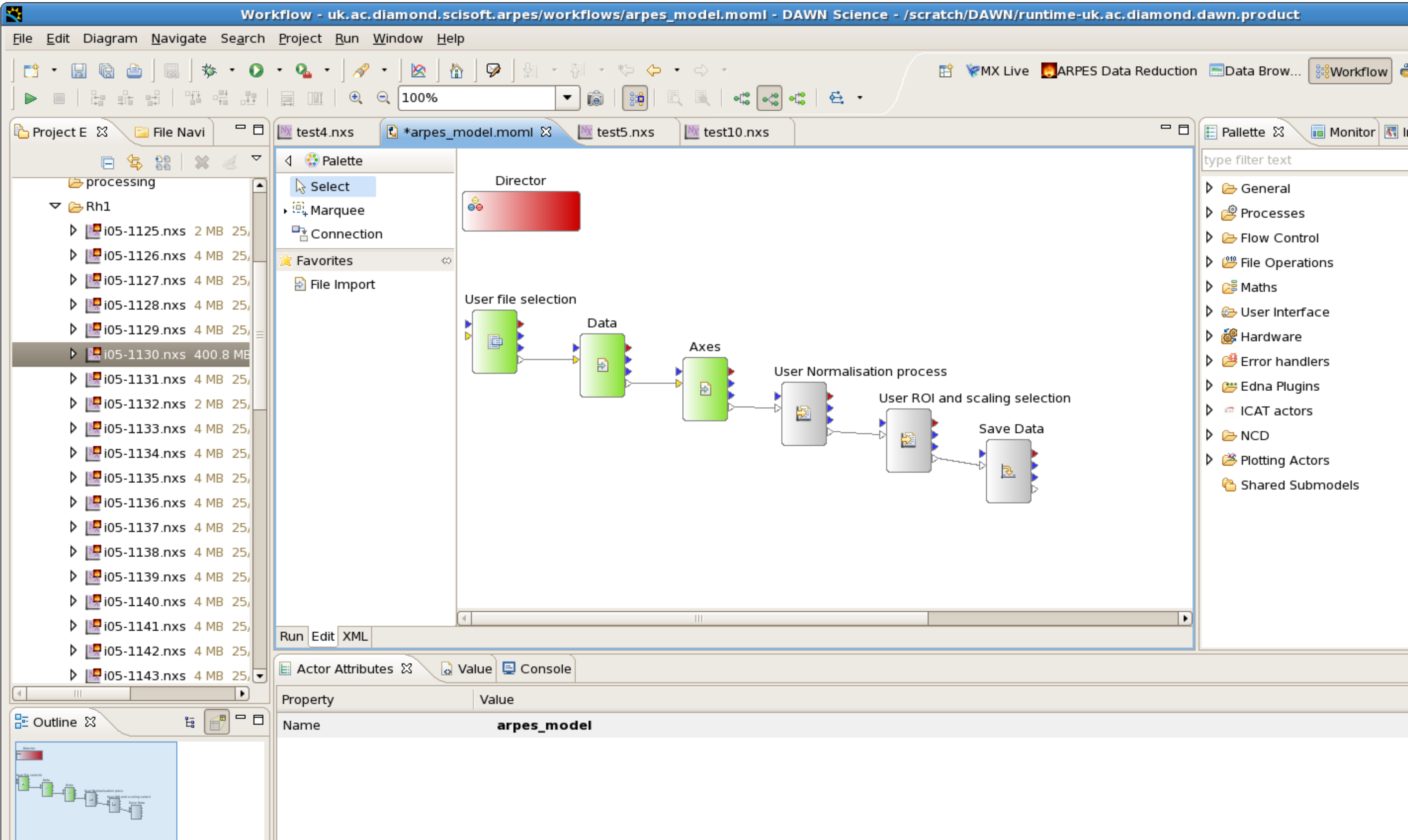
Name	value	min	max	Fixed
mu	16.822*	16.8*	16.85*	false*
temperature	20*	0*	200*	false*
BG_slope	-5,185.165*	-10,000*	10,000*	false*
FE_step_heigh	1,239.326*	0*	10,000*	false*
Constant	21.26738	0*	1,000*	false*
FWHM	1*	0*	1*	false*

* Click to change

Continue | Stop

ARPES Processing

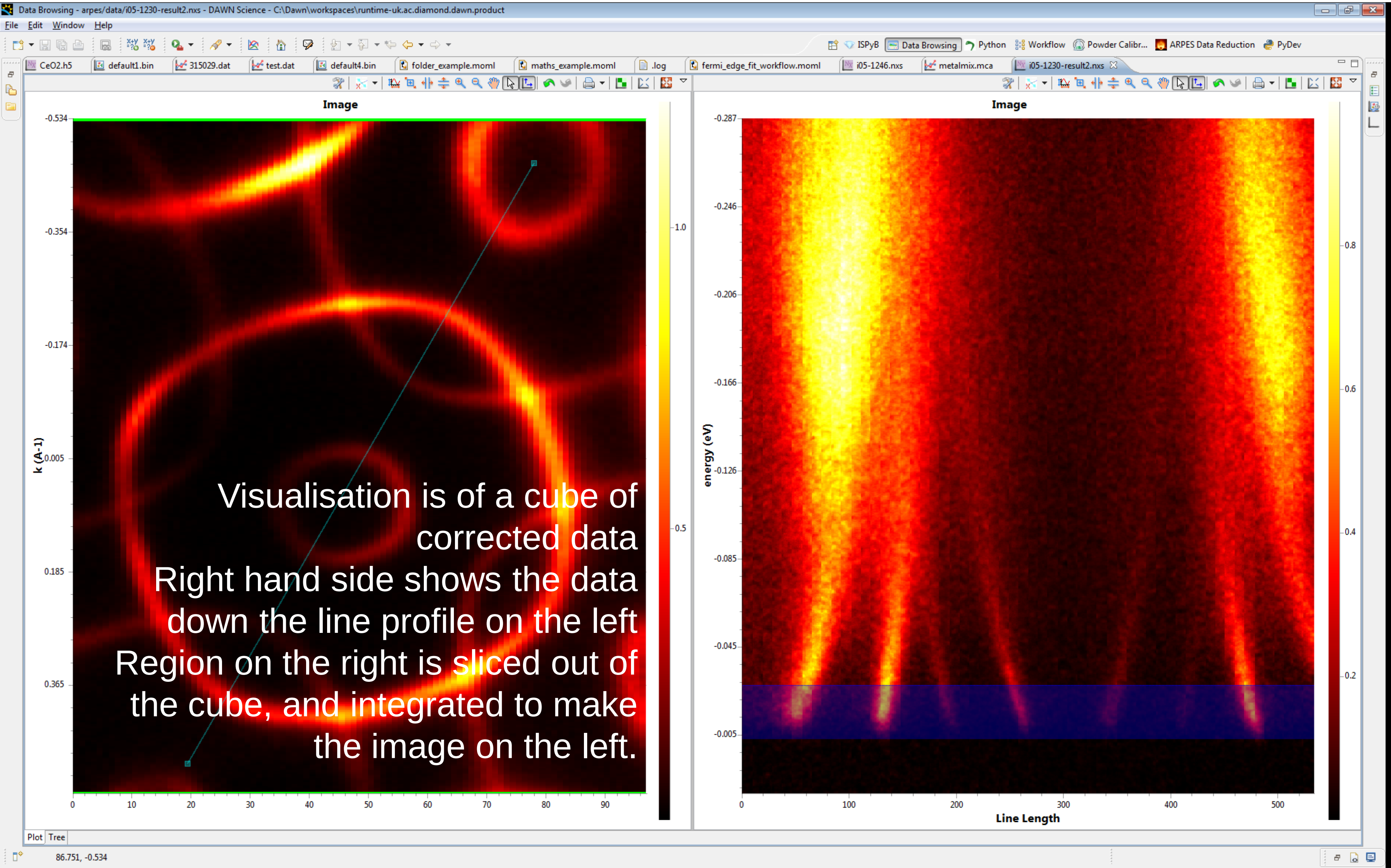
ARPES



Workflow - uk.ac.diamond.scisoft.arpes/workflows/arpes_model.moml - DAWN Science - /scratch/DAWN/runtime-uk.ac.diamond.dawn.product

Property	Value
Name	**arpes_model**

ARPES

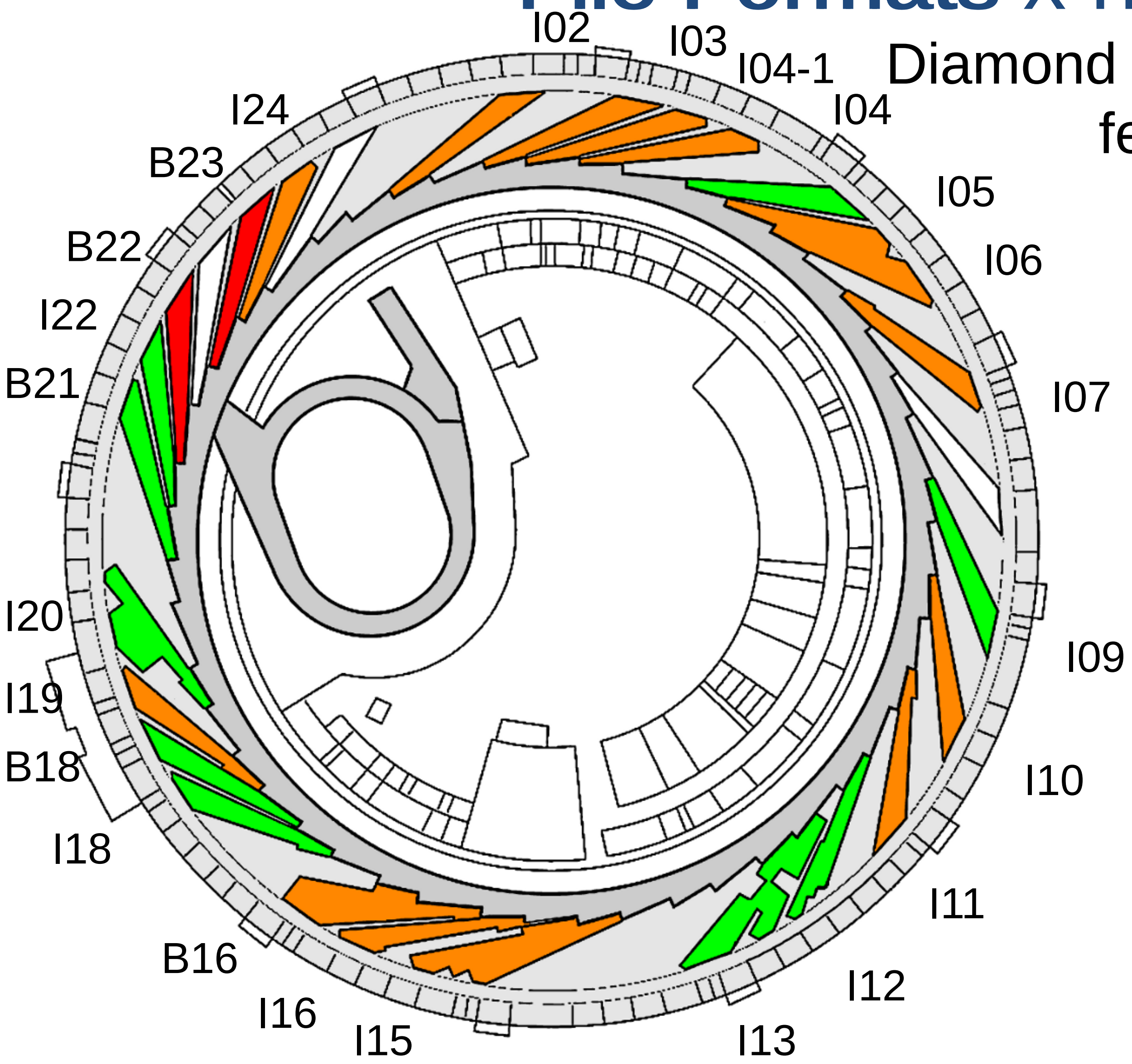


Visualisation is of a cube of corrected data

Right hand side shows the data down the line profile on the left

Region on the right is sliced out of the cube, and integrated to make the image on the left.

File Formats x n



Diamond has a policy of, where feasible, to standardise on file formats, the choice being NeXus/HDF5

Green: predominantly using NeXus.

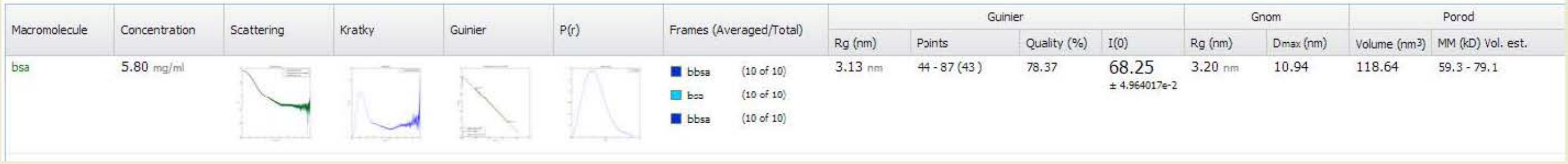
Orange: Mixed NeXus and other formats or considering NeXus in the next 12 months.

Files can be generated by Detector, EPICS or Data Acquisition

ISPyB information Flow: Display of Results

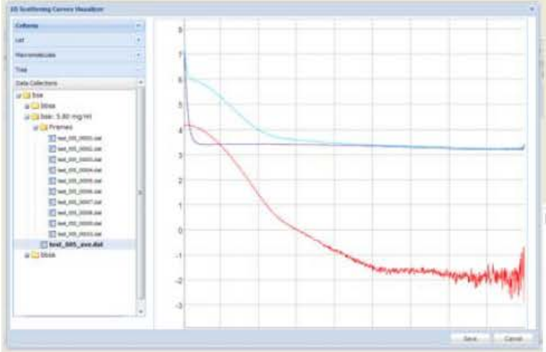
Data tracking and status checking of data acquired:

- Automatic plot generation of Scattering curves, Kratky, Guinier and P(r) plots

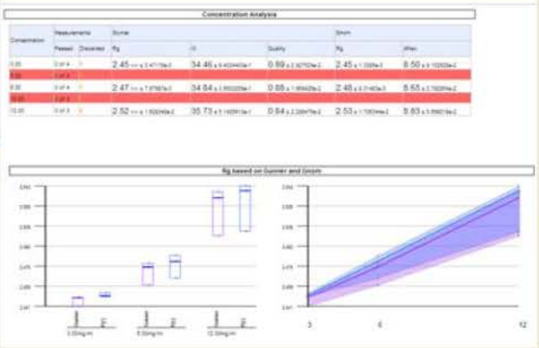


Macromolecule	Concentration	Scattering	Kratky	Guinier	P(r)	Frames (Averaged/Total)		Guinier				Gnom		Porod	
								Rg (nm)	Points	Quality (%)	I(0)	Rg (nm)	Dmax (nm)	Volume (nm3)	MM (kD) Vol. est.
bsa	5.80 mg/ml					■ bbsa ■ bsa ■ bbsa	(10 of 10) (10 of 10) (10 of 10)	3.13 nm	44 - 87 (43)	78.37	68.25 ± 4.964017e-2	3.20 nm	10.94	118.64	59.3 - 79.1

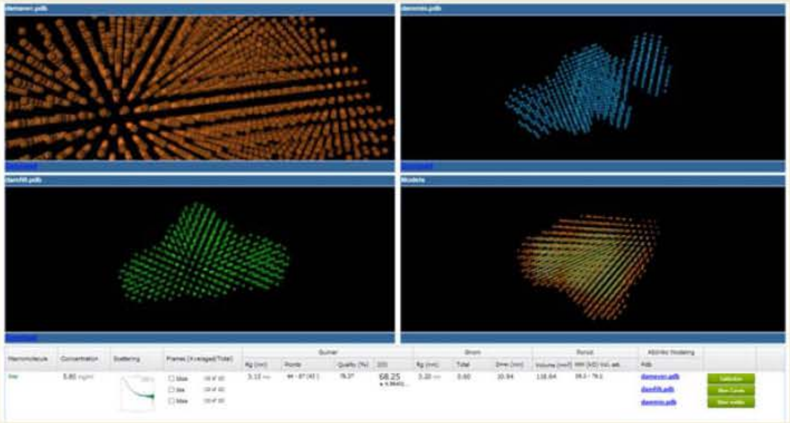
- Display of 1D curves, averaged curves and subtracted curves



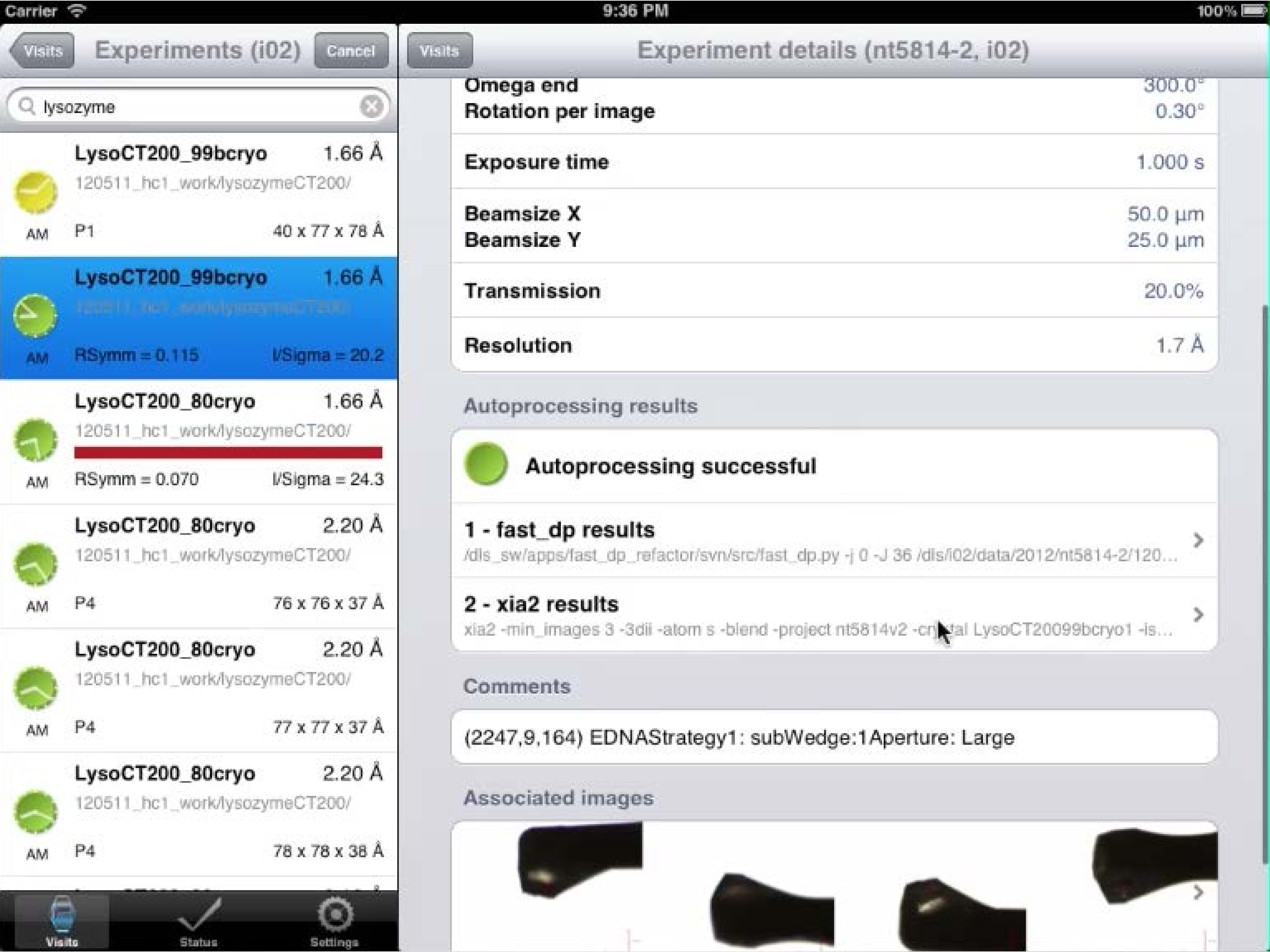
- Concentration Effects



- 3D Models displayed using WebGL



A prototype version of ISPyB for BioSAXS (ISPyBB) database is now available at BM29@ESRF and is being ported to P12@EMBL



Experiments (i02)

lysozyme

AM	LysoCT200_99bcryo 120511_hc1_work/lysozymeCT200/ P1	1.66 Å 40 x 77 x 78 Å
AM	LysoCT200_99bcryo [illegible] RSymm = 0.115	1.66 Å I/Sigma = 20.2
AM	LysoCT200_80cryo 120511_hc1_work/lysozymeCT200/ RSymm = 0.070	1.66 Å I/Sigma = 24.3
AM	LysoCT200_80cryo 120511_hc1_work/lysozymeCT200/ P4	2.20 Å 76 x 76 x 37 Å
AM	LysoCT200_80cryo 120511_hc1_work/lysozymeCT200/ P4	2.20 Å 77 x 77 x 37 Å
AM	LysoCT200_80cryo 120511_hc1_work/lysozymeCT200/ P4	2.20 Å 78 x 78 x 38 Å

Experiment details (nt5814-2, i02)

Omega end	300.0°
Rotation per image	0.30°
Exposure time	1.000 s
Beamsize X	50.0 µm
Beamsize Y	25.0 µm
Transmission	20.0%
Resolution	1.7 Å

Autoprocessing results

● Autoprocessing successful

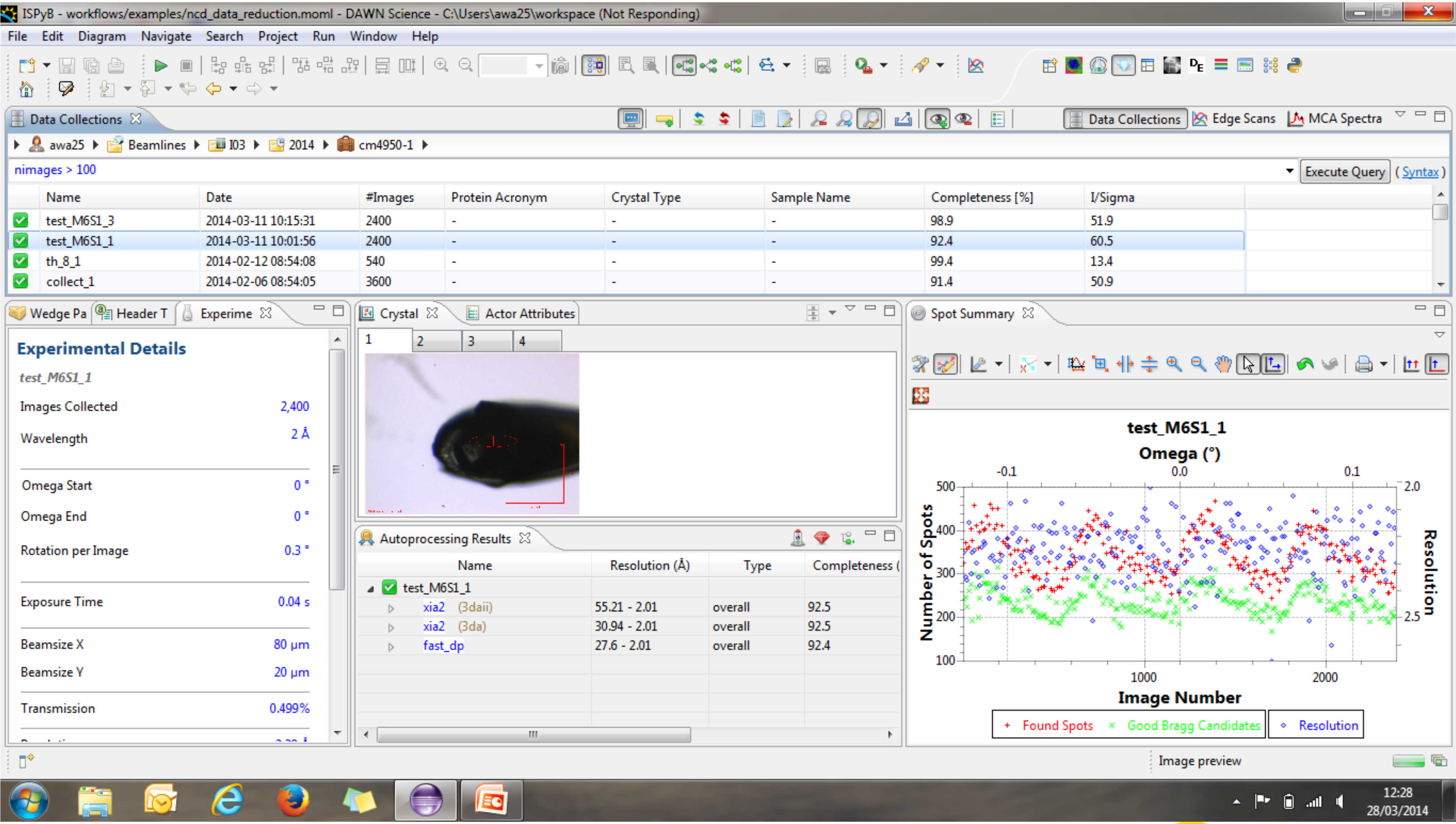
1 - fast_dp results
/dls_sw/apps/fast_dp_refactor/svn/src/fast_dp.py -j 0 -J 36 /dls/i02/data/2012/nt5814-2/120...

2 - xia2 results
xia2 -min_images 3 -3dii -atom s -blend -project nt5814v2 -crystal LysoCT20099bcryo1 -is...

Comments

(2247,9,164) EDNAStrategy1: subWedge:1Aperture: Large

Associated images



ISPyB - workflows/examples/ncd_data_reduction.moml - DAWN Science - C:\Users\awa25\workspace (Not Responding)

Data Collections

awa25 ▸ Beamlines ▸ I03 ▸ 2014 ▸ cm4950-1

nimages > 100

	Name	Date	#Images	Protein Acronym	Crystal Type	Sample Name	Completeness [%]	I/Sigma
☑	test_M6S1_3	2014-03-11 10:15:31	2400	-	-	-	98.9	51.9
☑	test_M6S1_1	2014-03-11 10:01:56	2400	-	-	-	92.4	60.5
☑	th_8_1	2014-02-12 08:54:08	540	-	-	-	99.4	13.4
☑	collect_1	2014-02-06 08:54:05	3600	-	-	-	91.4	50.9

Experimental Details

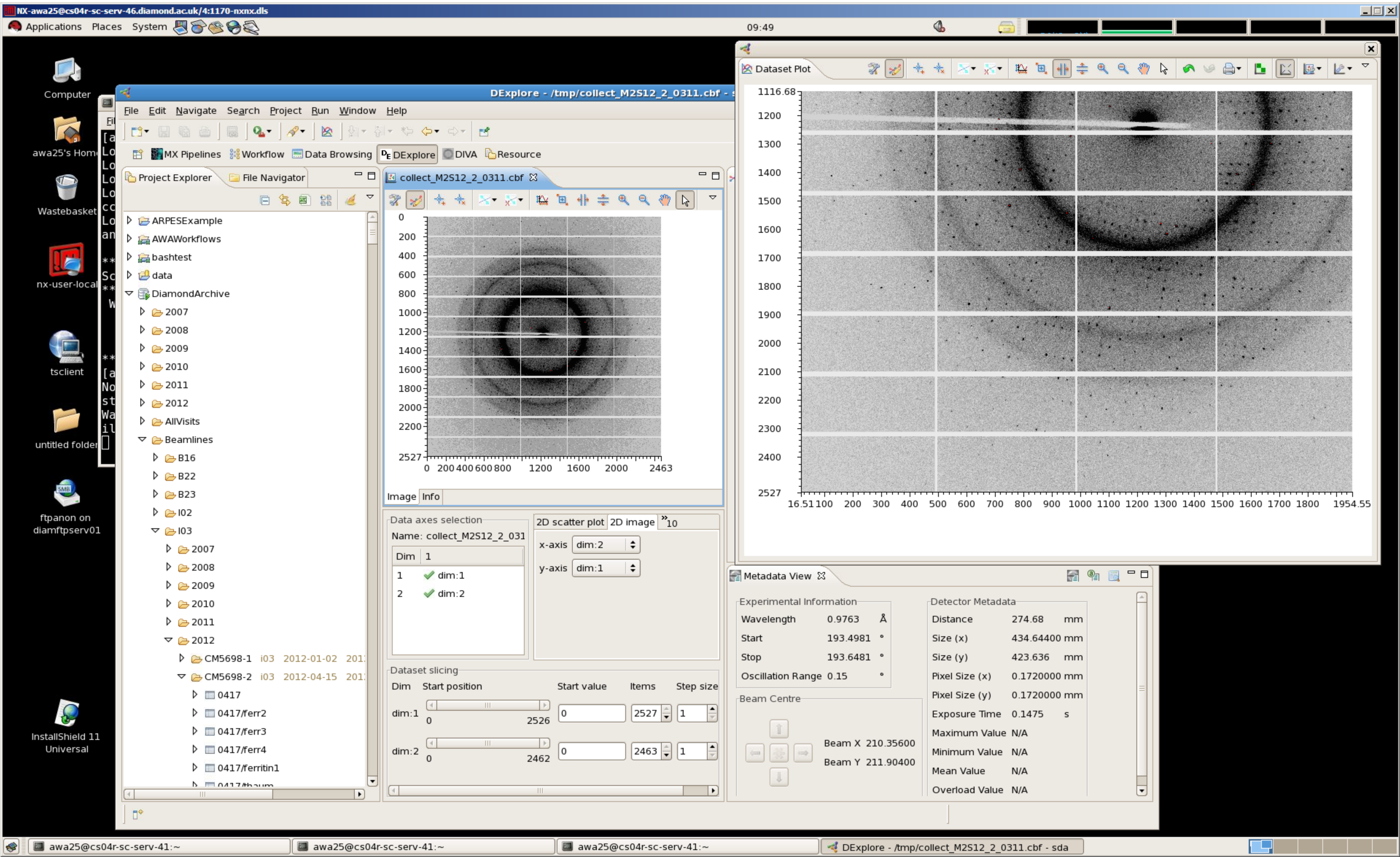
test_M6S1_1

Images Collected	2,400
Wavelength	2 Å
Omega Start	0 °
Omega End	0 °
Rotation per Image	0.3 °
Exposure Time	0.04 s
Beamsize X	80 µm
Beamsize Y	20 µm
Transmission	0.499%

Autoprocessing Results

Name	Resolution (Å)	Type	Completeness (
☑ test_M6S1_1			
xia2 (3daii)	55.21 - 2.01	overall	92.5
xia2 (3da)	30.94 - 2.01	overall	92.5
fast_dp	27.6 - 2.01	overall	92.4

Archive and Reprocessing Service



Data axes selection

Name: collect_M2S12_2_031

Dim	1
1	✓ dim:1
2	✓ dim:2

2D scatter plot | 2D image | »10

x-axis: dim:2

y-axis: dim:1

Dataset slicing

Dim	Start position	Start value	Items	Step size
dim:1	0 2526	0	2527	1
dim:2	0 2462	0	2463	1

Experimental Information

Wavelength	0.9763	Å
Start	193.4981	°
Stop	193.6481	°
Oscillation Range	0.15	°

Beam Centre

Beam X 210.35600

Beam Y 211.90400

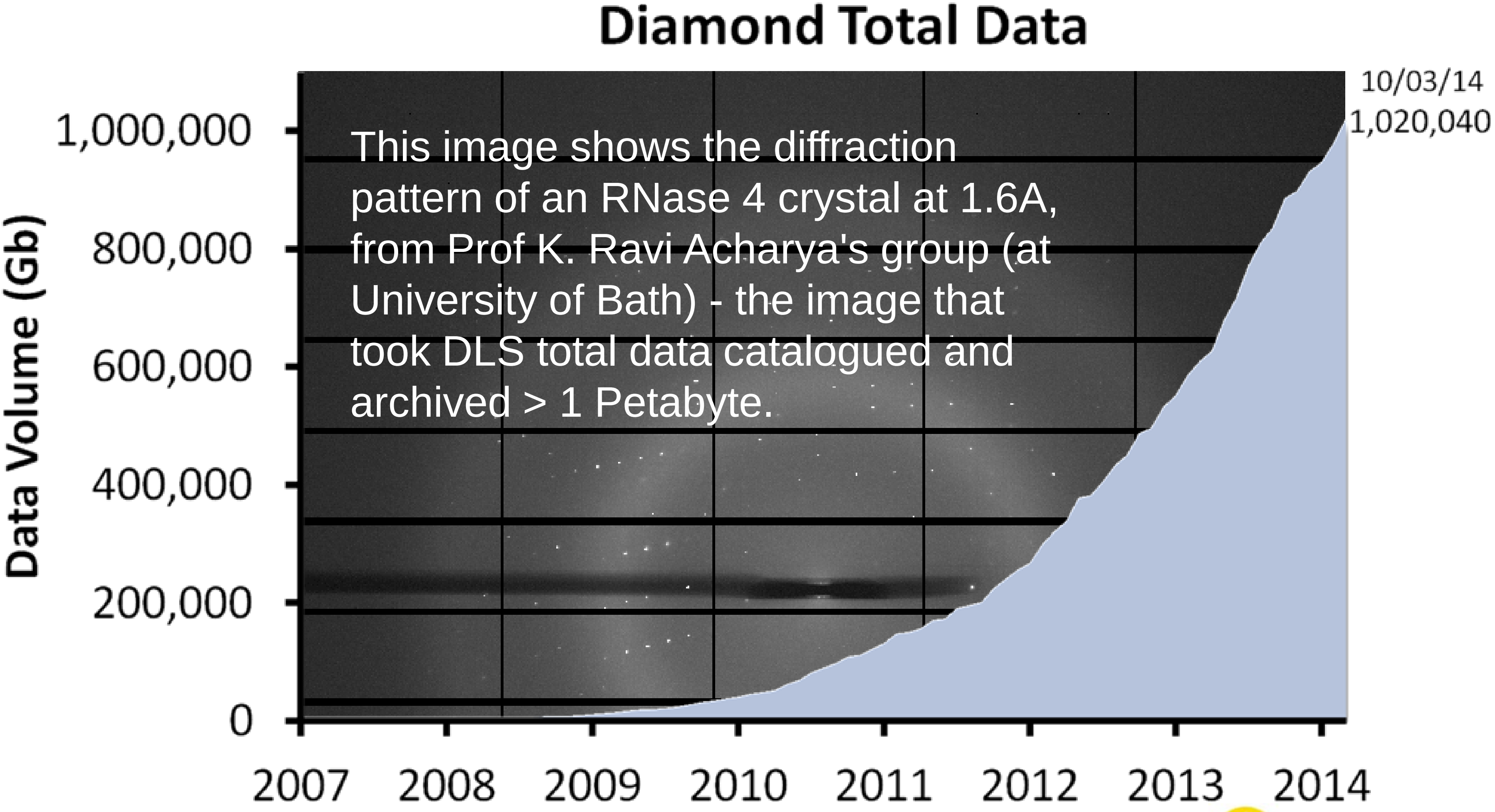
Detector Metadata

Distance	274.68	mm
Size (x)	434.64400	mm
Size (y)	423.636	mm
Pixel Size (x)	0.1720000	mm
Pixel Size (y)	0.1720000	mm
Exposure Time	0.1475	s
Maximum Value	N/A	
Minimum Value	N/A	
Mean Value	N/A	
Overload Value	N/A	

Collaboration experience

- Collaborating Inside the facility is often as challenging as with external groups.
- ‘customer’ involvement essential
- Management enthusiasm and support essential
- The scope of a collaborative project must fit in with the ambitions of your facility.
- Collaborating is expensive and ideally all parties must have adequate and commensurate resources though a centre of mass can produce results quicker, but at the cost of less well resourced partners.
- Although service providers are often happy to collaborate, our customers might not....

Current status/volumes in Diamond ICAT = 285,198,074 files



This image shows the diffraction pattern of an RNase 4 crystal at 1.6A, from Prof K. Ravi Acharya's group (at University of Bath) - the image that took DLS total data catalogued and archived > 1 Petabyte.

Novel developments in HDF5

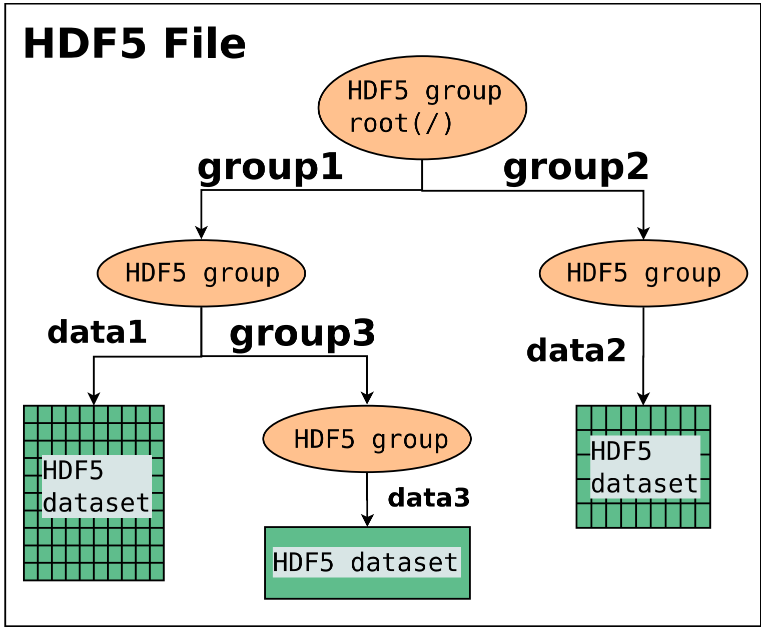
An overview on new developments in the HDF5 library

Eugen Wintersberger
Dublin, 28.03.2014

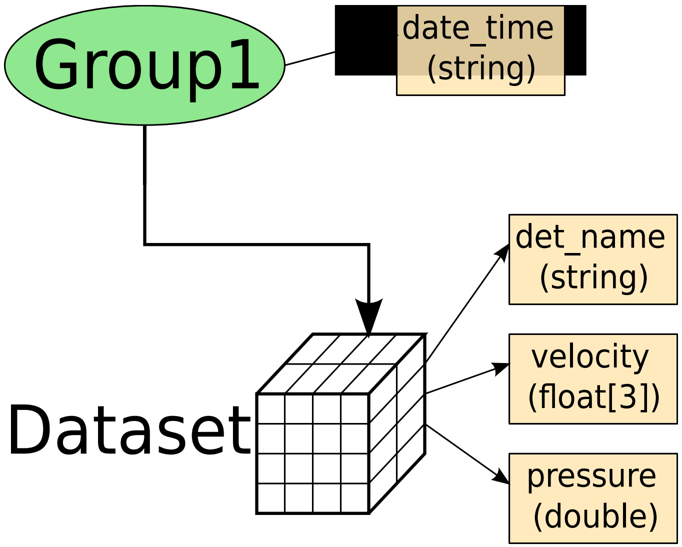
HDF5 at a glance ...

HDF5 is a binary file format for large datasets

Data organized as trees by means of groups and datasets



Groups and datasets can be annotated by attributes



Developed by the HDF Group in Champaign, IL, USA

HDF5 key features

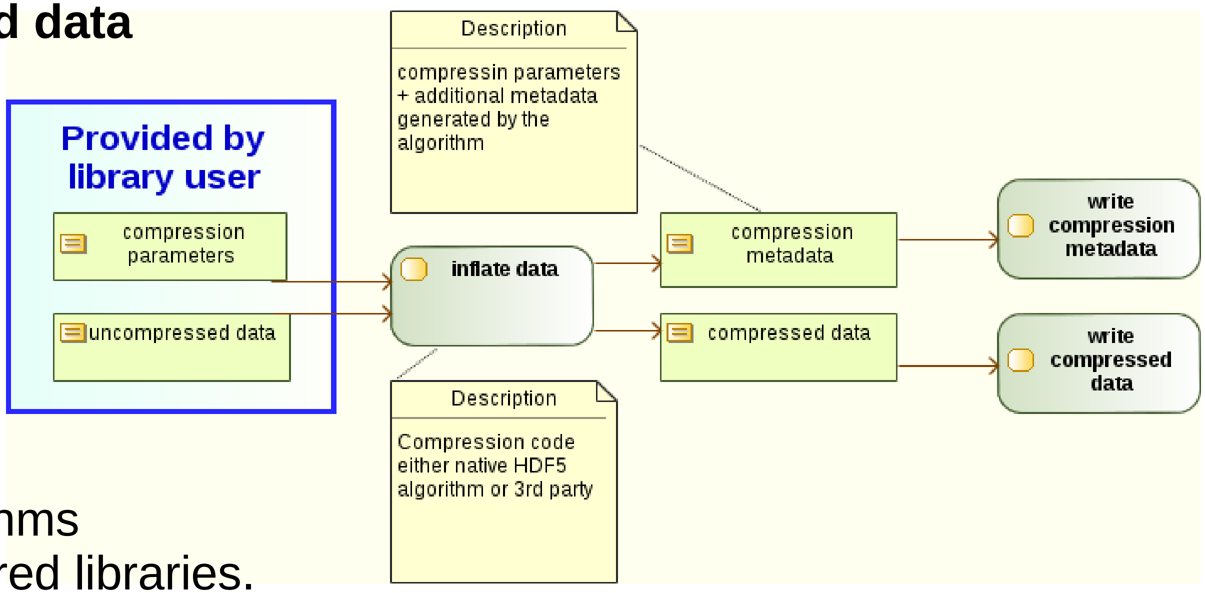
- Transparent compression of individual datasets
- platform independent (Windows, Linux, AIX, Solaris, OSX, OpenVMS)
- library is implemented in C (easy to interface)
- bindings to many languages: C++, Fortran, Python, Java, Perl, R, Go
- Supported by commercial products like: Matlab, Mathematica, IDL

Easy parsing: data access via library → it is not about how data is stored but rather where (path).

Recently developed features already available in 1.8.12

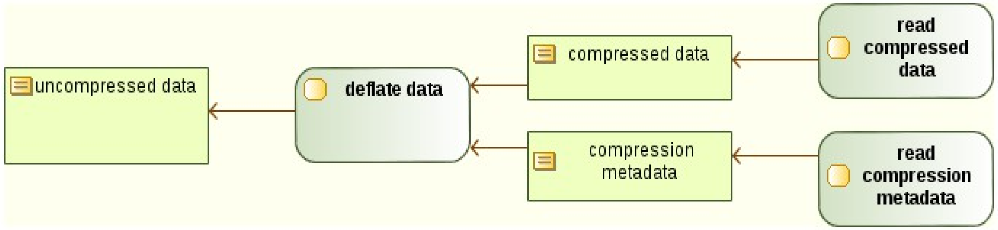
How compression works in HDF5

Writing compressed data


Provided by library user
compression parameters
uncompressed data
Description
compressin parameters + additional metadata generated by the algorithm
inflate data
Description
Compression code either native HDF5 algorithm or 3rd party
compression metadata
compressed data
write compression metadata
write compressed data

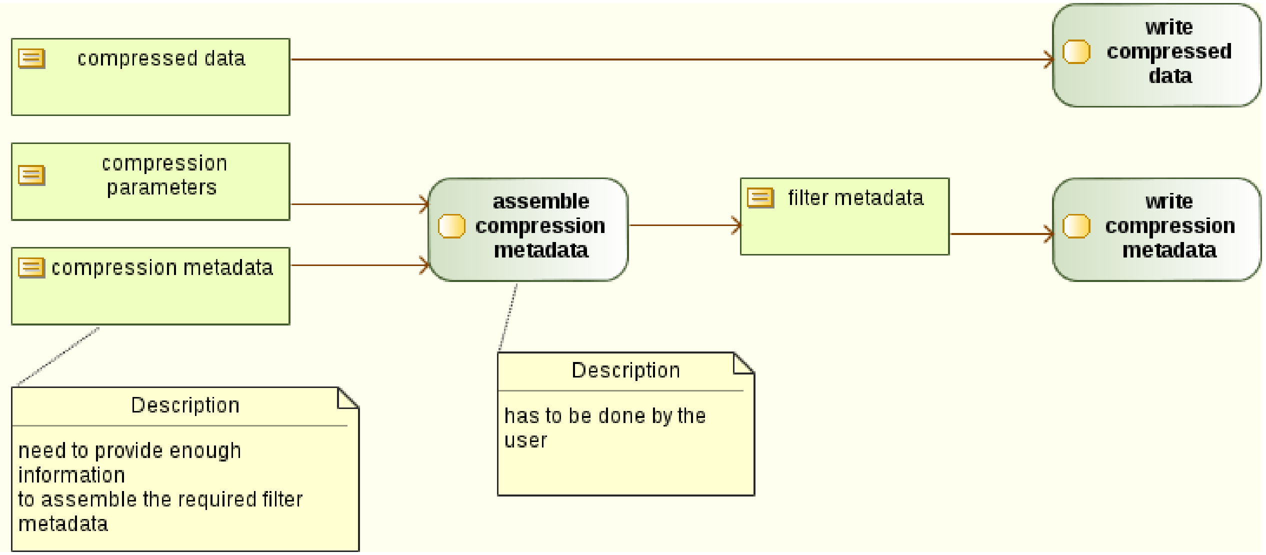
Compression algorithms implemented as shared libraries.

Reading compressed data



Writing precompressed data ...

Sometimes compression can be done better in hardware (FPGAs)
→ reduced bandwidth requirements for network and disk I/O!



Description

need to provide enough information
to assemble the required filter metadata

Description

has to be done by the user

Feature funded by DECTRIS

DECTRIS
Next Generation X-Ray Detectors

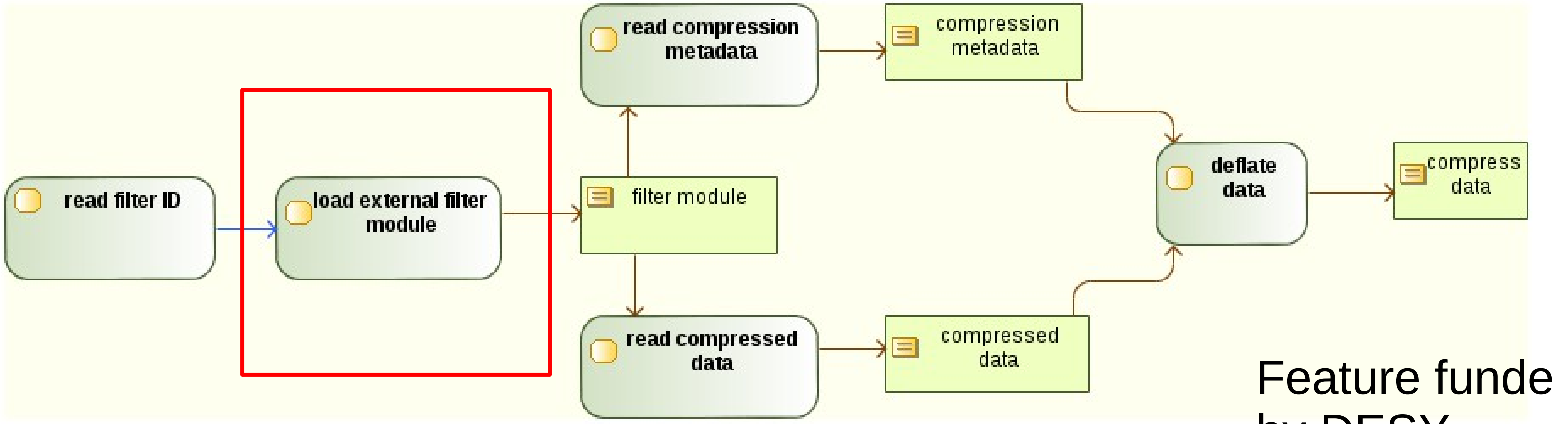
This features allows bypassing the filter chain and write compressed data directly.

Reading the data is the same as for conventionally written data!

External filter API

Detector vendors or facilities may want to use custom compression algorithms.

Possible with all HDF5 1.8.X versions but: reading software needs to be recompiled => commercial applications do not have access to the data.



Feature funded by DESY

External filters can be loaded at runtime on demand.

Requirement:
Detector vendors and facilities provide libraries for all target platforms.

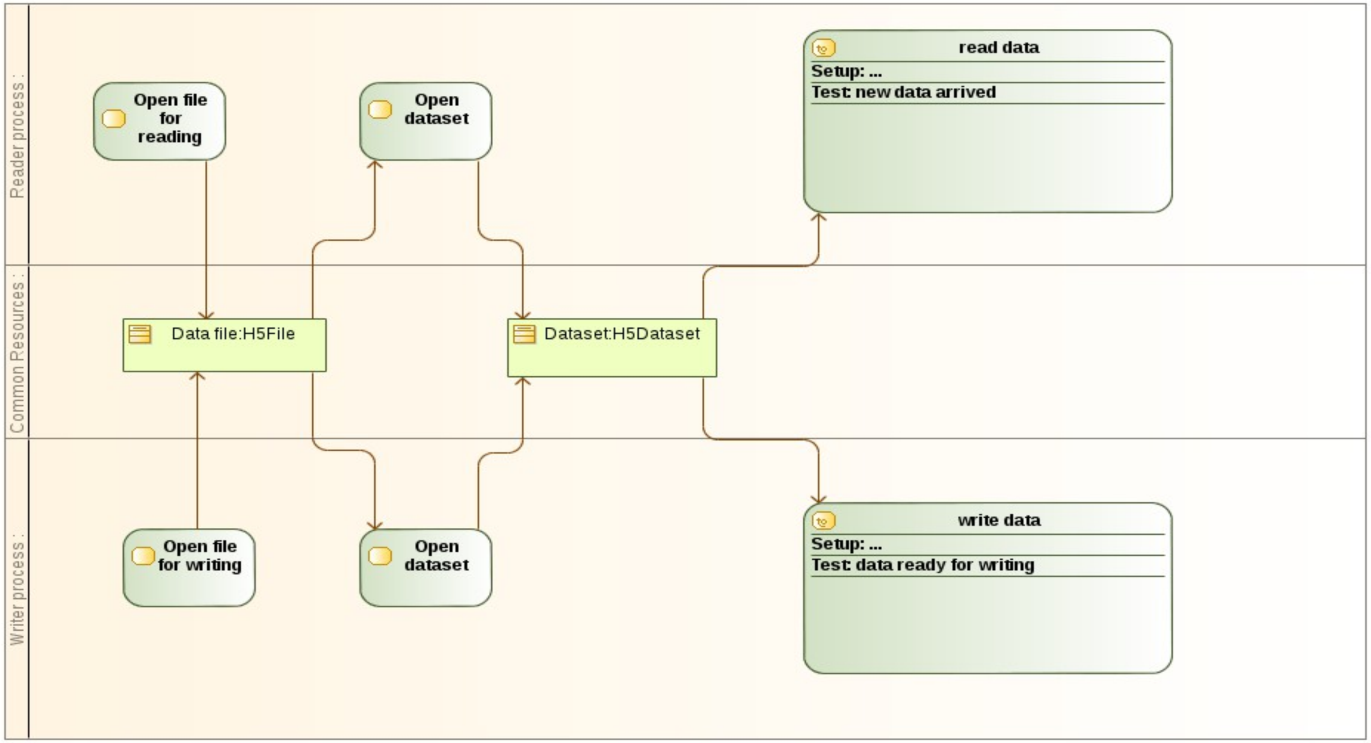
New features addressing concurrent data access

Reading while Writing (SWMR) ...

Single Writer – Multiple Reader = SWMR

Current situation: cannot read data from a file while it is written by two different processes!

SWMR would make this possible!



Status: A prototype is available – missing funding for library integration.

Multithreading support

Option 1: Make HDF5 fully threadsafe!

- would give best performance
- major rewrite of 300K lines of C-Code
- costs of 4-6 FTEs
- significant future maintenance efforts

Very unlikely to happen without extensive funding!

Option 2: Use threads within the library!

- use multiple threads for compression
- asynchronous (non-blocking) I/O
- costs ~1.5 FTEs
- lower maintenance effort
- could improve continuously

Remote access to data stored in an HDF5 file

OPeNDAP

Open Source Project for a Network Data Access Protocol

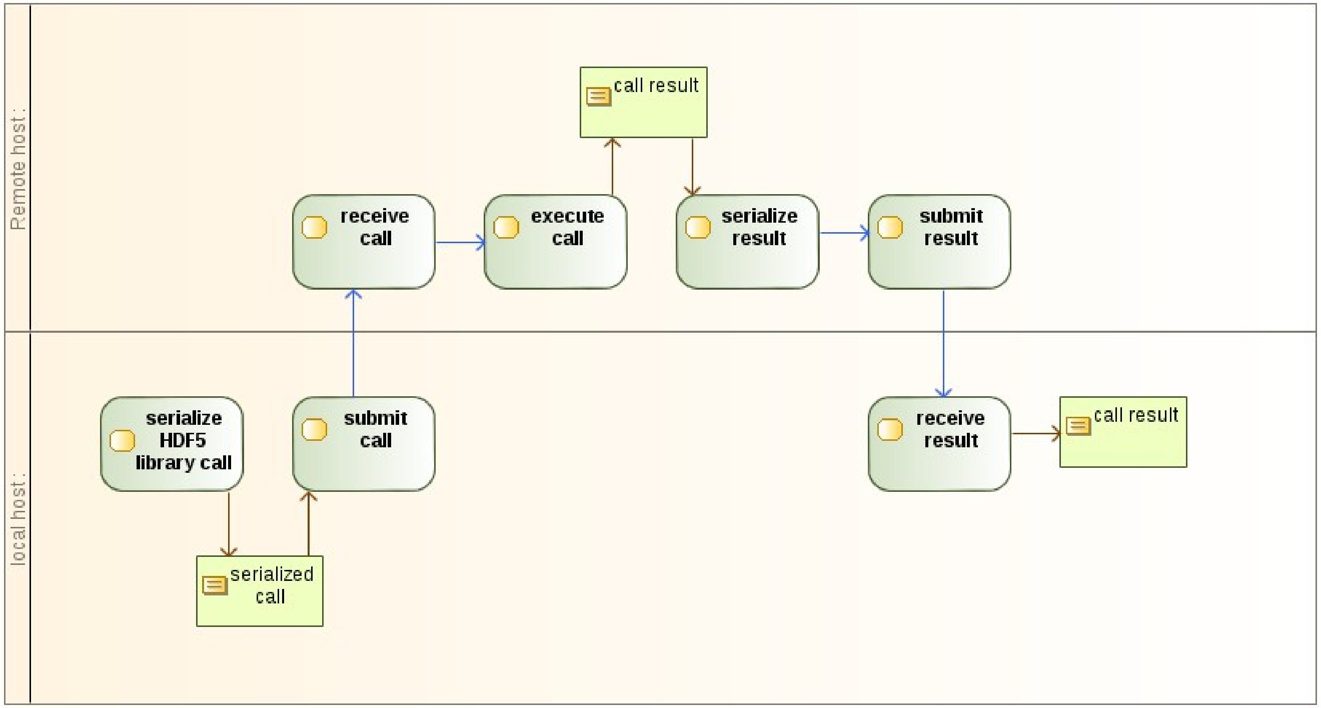
Advantages

- Already existing and actively developed protocol
- Data access via HTTP requests where URLs encode the information of what shall be retrieved from a data source
- Independent of the dataformat (other formats can easily be supported)
- Data can be accessed with any software that can dereference URLS (webbrowser, Excel, Libreoffice,...)
- Matlab 2012a provides native support for OPeNDAP
- C++, Java, and Python libraries available

www.opendap.org

HDF5 data server

Serialization of HDF5 library calls rather than using HTTP and URLs



Status: only as a design draft – no implementation will be done without funding.

Licenses and funding issues

Currently the HDF5 group is entirely project funded!

… they are looking for new funding (licensing) models

➔ the core library will always remain open source

➔ dual licensing would be an alternative

➔ some features may become commercial products in future

As a user community we should take funding issues of the HDF5 group serious and think about how we can improve their financial backup!

Conclusions: what the RDA could do

- Establish an HDF5 special interest group as a hub for HDF5 related work
- collect feature requests from various user communities
- organize funding activities if a new feature should be implemented
- manage filters (compression algorithms) which should be available for HDF5 and provide hosting resources
- Provided solutions based on HDF5 for certain selected use-cases
- Organize workshops around HDF5 (maybe as satellite events around meetings)

pandata

Photon and Neutron
Data Infrastructure

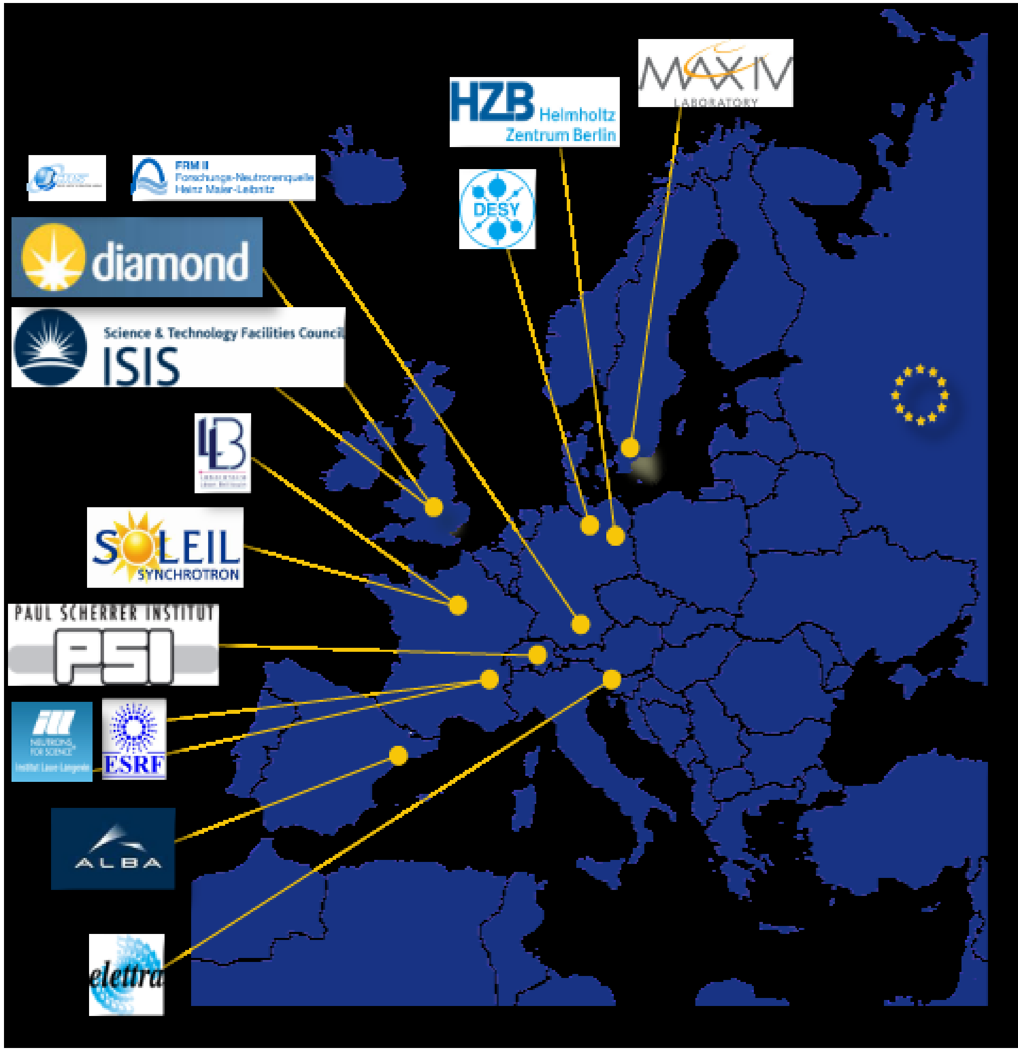
Brian Matthews
STFC

PaNdata

- Photon and Neutron Data Infrastructure
- Established in 2007 with 4 facilities
  - now standing at 13
  - With “friends” around the world
- Combined Number of Unique Users more than 35000 in 2011
- Combines Scientific and IT staff from the collaborating facilities
- European Framework 7 Projects
  - PaNdata-Europe: SA, 2009-11
  - PaNdata-Open Data Infrastructure, IP, 2011-14

–Guestimates

- Investment > €4.000.000.000*
- Running costs > €500.000.000/yr*
- Publications > 10.000/yr*
- RCosts/Publication ~ €50.000*%
- Data volume >> 10PB/yr*


MAX IV LABORATORY
HZB Helmholtz Zentrum Berlin
FRM II Forschungs-Neutronenquelle Heinz Maier-Leibnitz
DESY
diamond
Science & Technology Facilities Council ISIS
SOLEIL SYNCHROTRON
PAUL SCHERRER INSTITUT PSI
NEUTRONS FOR SCIENCE Institut Laue-Langevin
ESRF
ALBA
elettra

Counting Users

Number of Users shared between facilities																
	ALBA	BER II	DESY	DLS	ELETTRA	ESRF	FRM-II	ILL	ISIS	LLB	SINQ	SLS	SOLEIL	neutron	photon	all
ALBA	773	7	61	58	51	281	2	51	13	5	10	77	105	69	400	773
BER II	7	1563	115	46	27	179	157	383	198	98	191	62	36	580	329	1563
DESY	61	115	4197	137	222	851	116	255	113	62	95	315	188	469	1294	4197
DLS	58	46	137	4407	102	810	30	267	399	33	52	229	192	546	1130	4407
ELETTRA	51	27	222	102	3167	433	11	77	35	20	18	179	367	141	900	3167
ESRF	281	179	851	810	433	10287	139	900	369	190	174	963	1286	1313	3586	10287
FRM-II	2	157	116	30	11	139	1095	347	137	89	161	33	29	509	259	1095
ILL	51	383	255	267	77	900	347	4649	731	301	395	156	222	1518	1347	4649
ISIS	13	198	113	399	35	369	137	731	2880	89	233	94	56	936	745	2880
LLB	5	98	62	33	20	190	89	301	89	1235	74	39	151	391	323	1235
SINQ	10	191	95	52	18	174	161	395	233	74	1219	224	31	590	415	1219
SLS	77	62	315	229	179	963	33	156	94	39	224	3827	399	371	1470	3827
SOLEIL	105	36	188	192	367	1286	29	222	56	151	31	399	4568	394	1817	4568
neutron	69	1563	469	546	141	1313	1095	4649	2880	1235	1219	371	394	10023	2334	10023
photon	773	329	4197	4407	3167	10287	259	1347	745	323	415	3827	4568	2334	25336	25336
all	773	1563	4197	4407	3167	10287	1095	4649	2880	1235	1219	3827	4568	10023	25336	33025

http://pan-data.eu/Users2012-Results

PaN-Data Integration

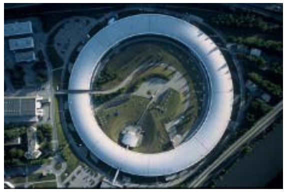
Common data environment, common user experience

Shared Data Policy Framework

Federated User Authentication umbrella

Federated Data Catalogue ICAT

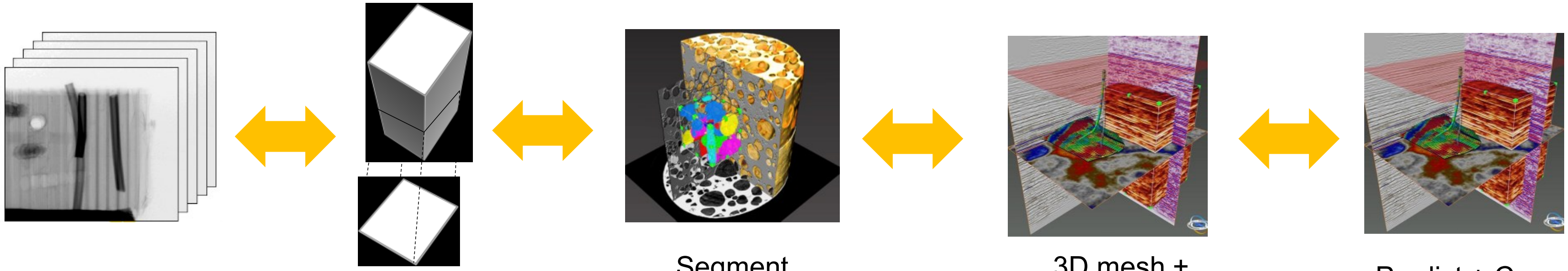
Common Data Format ***NeXus***



Towards the Future

- Provenance
  - Integrating context, analysis and publication into the record
- Preservatiom
  - Long-term need for archiving and curating data
  - Persistence Identifiers, itegtrity, context,
  - Costs and Benefits of data preservation
- Scalability
  - Managing high data rates and volumes
  - Parallel file stores

Infrastructure for managing data flows



Scan

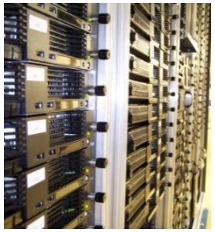
Reconstruct

Segment
+ Quantify

3D mesh +
Image based Modelling

Predict + Compare

ICAT



Data
Catalogue

Petabyte
Data storage

Parallel
File system

HPC
CPU+GPU

Visualisation

Infrastructure + Software + Expertise!

- **Tomography**: Dealing with high data volumes – 200Gb/scan, ~5 TB/day (one experiment)
- **MX**: high data volumes, smaller files, but a lot more experiments
- Hard to move the data – needs to be handled at the facility?

Some mage credit: Avizo, Visualization Sciences Group (VSG)