

# Supporting Data Workflows at STFC

Brian Matthews  
Scientific Computing Department



Science & Technology  
Facilities Council

- What we do now : Raw Data Management
- What we want to do : Supporting user workflows
- What we want to do : sharing and publishing data
- Coming back to Metadata



- What we do now : Raw Data Management



# Supporting Facilities Data Management

- STFC Scientific Computing Department
  - Support three STFC Funded facilities on the RAL campus
  - Provide data archiving and management tools
- ISIS Neutron and Muon Source
  - Provide tools to support ISIS's data workflows
  - Support through the science lifecycle
  - Rich metadata
  - Provide Data archiving
- DLS Synchrotron Light Source
  - Data Archiving
  - Limited metadata
  - Managing the scale of the archive
- Central Laser Facility
  - Real-time data management and feedback to users
  - Rich metadata on laser configuration
  - Access to data



# DLS Archive Architecture

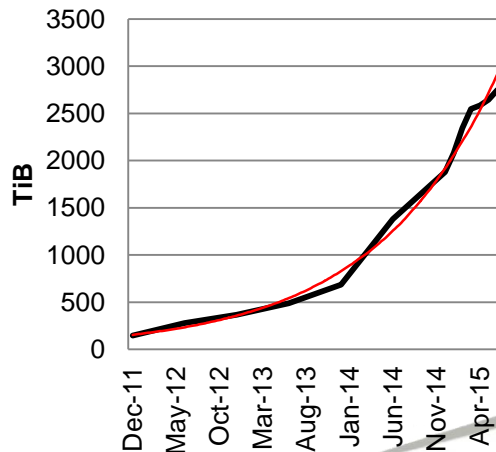
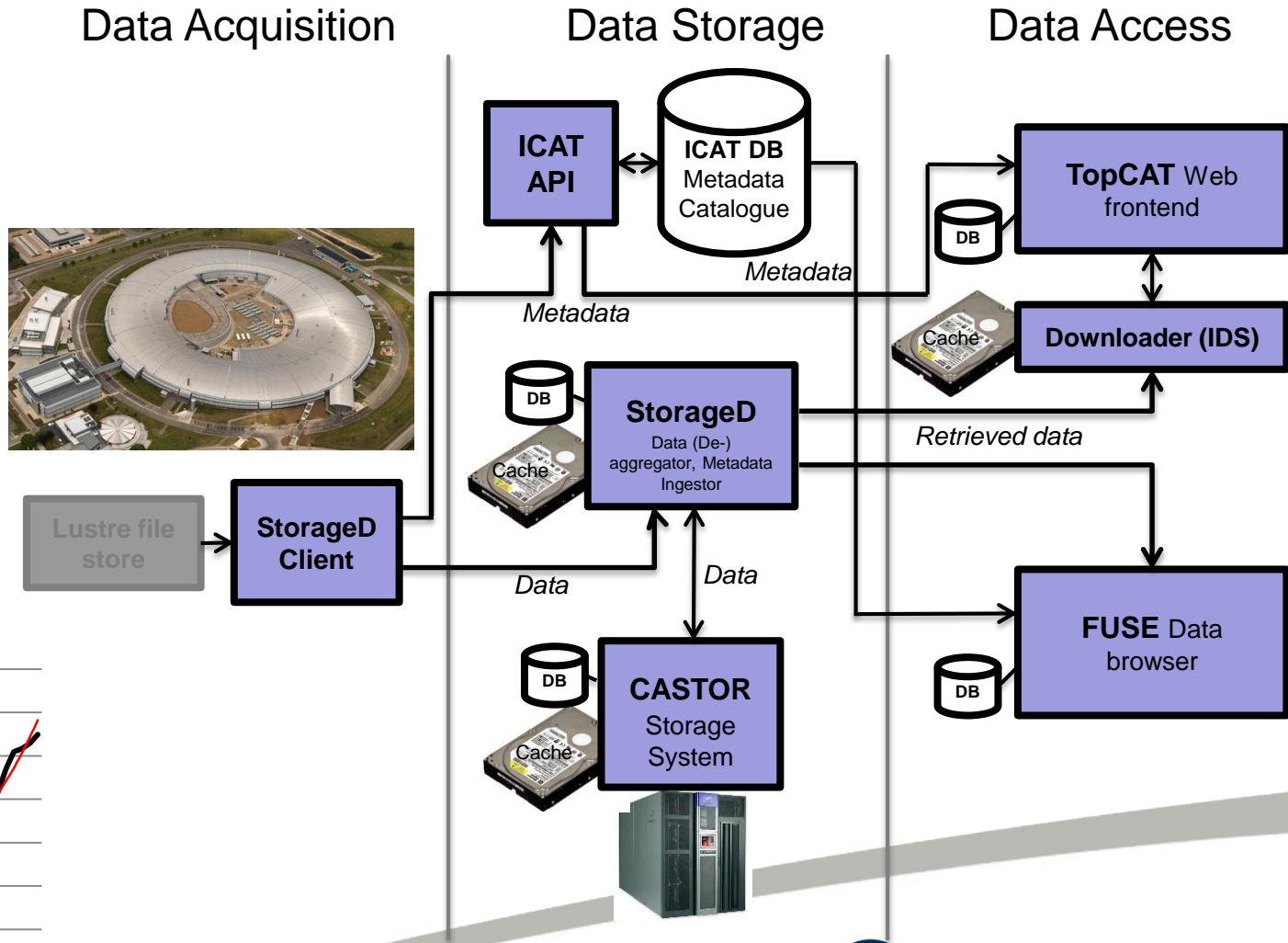


Archive of

- 3.3PB
- 846 million files in total in archive (July 2015)

(cf 2.2PB,  
620m Jan 2015)

- Cataloguing 12000 Files per minute



Science & Technology  
Facilities Council

# Supporting Data Management for STFC Facilities

- Integrated data management pipelines for data handling
  - From data acquisition to storage
- A Catalogue of Experimental Data
  - ICAT Tool Suite: *Metadata as Middleware*
  - Automated metadata capture
  - Integrated with the User Office and data acquisition system
- Providing access to the user
  - TopCat web front end
  - Integrated into Analysis frameworks
    - Mantid for Neutrons, DAWN for X-Rays

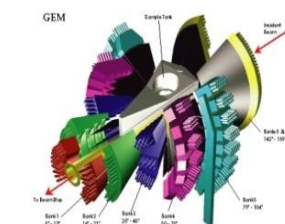


Science & Technology  
Facilities Council

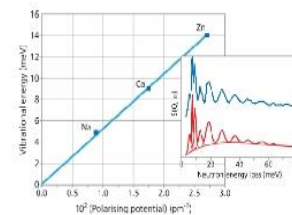
- Secure access to user's data
- Flexible data searching
- Scalable and extensible architecture
- Integration with analysis tools
- Access to high-performance resources
- Linking to other scientific outputs
- Data policy aware



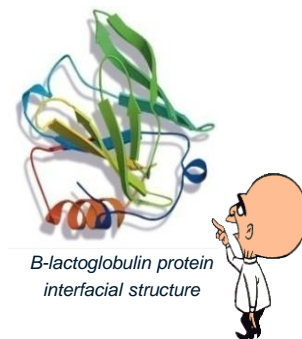
Example ISIS Proposal



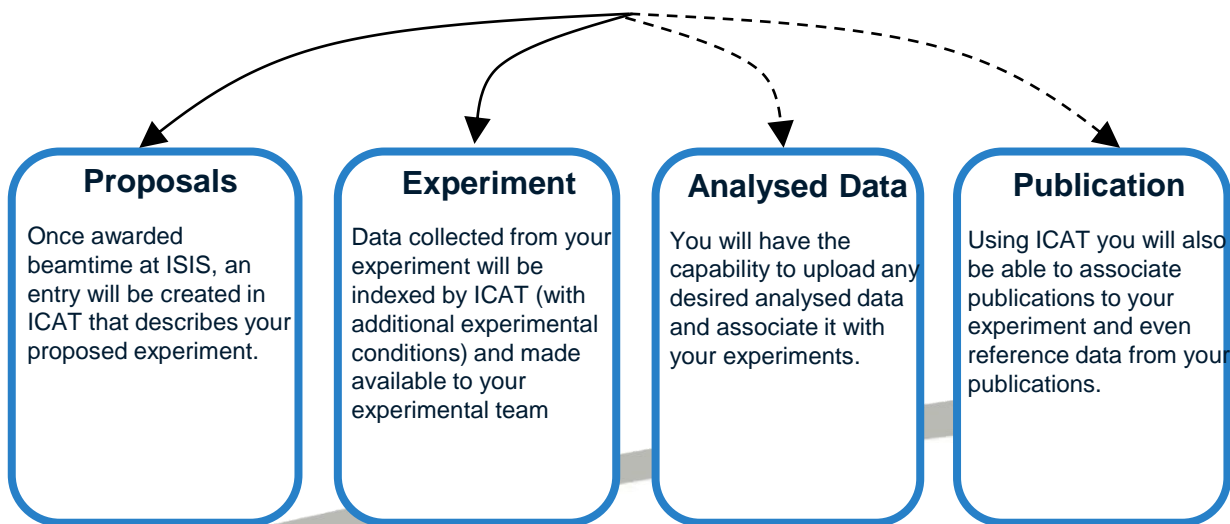
GEM – High intensity, high resolution neutron diffractometer



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



B-lactoglobulin protein interfacial structure

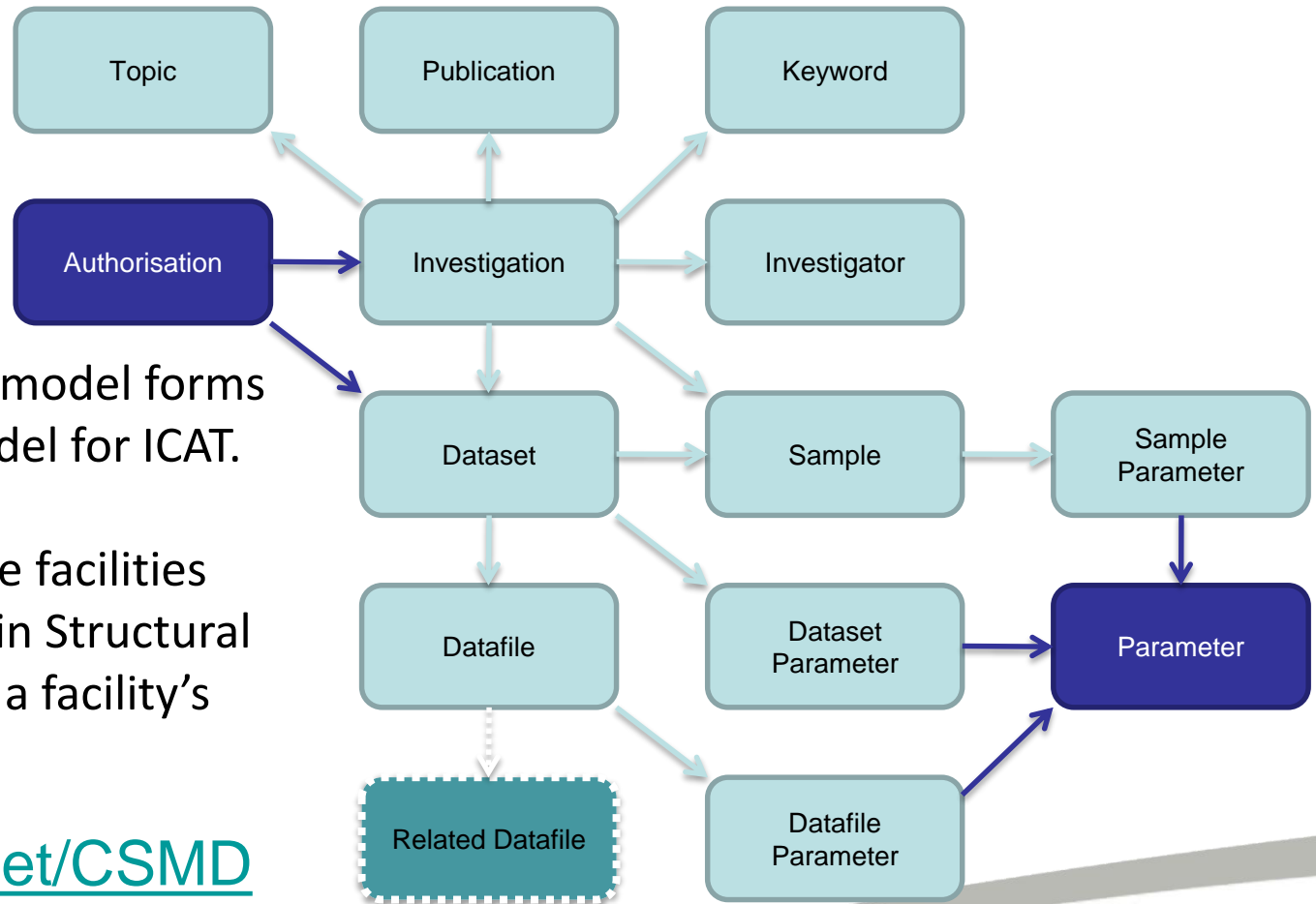


An international collaboration  
<http://icatproject.org>



Science & Technology  
Facilities Council

# Core Scientific Metadata Model (CSMD)



The Core Metadata model forms the information model for ICAT.

Designed to describe facilities based experiments in Structural Science throughout a facility's scientific workflow.

<http://purl.org/net/CSMD>  
<http://icatproject.org/CSMD/>

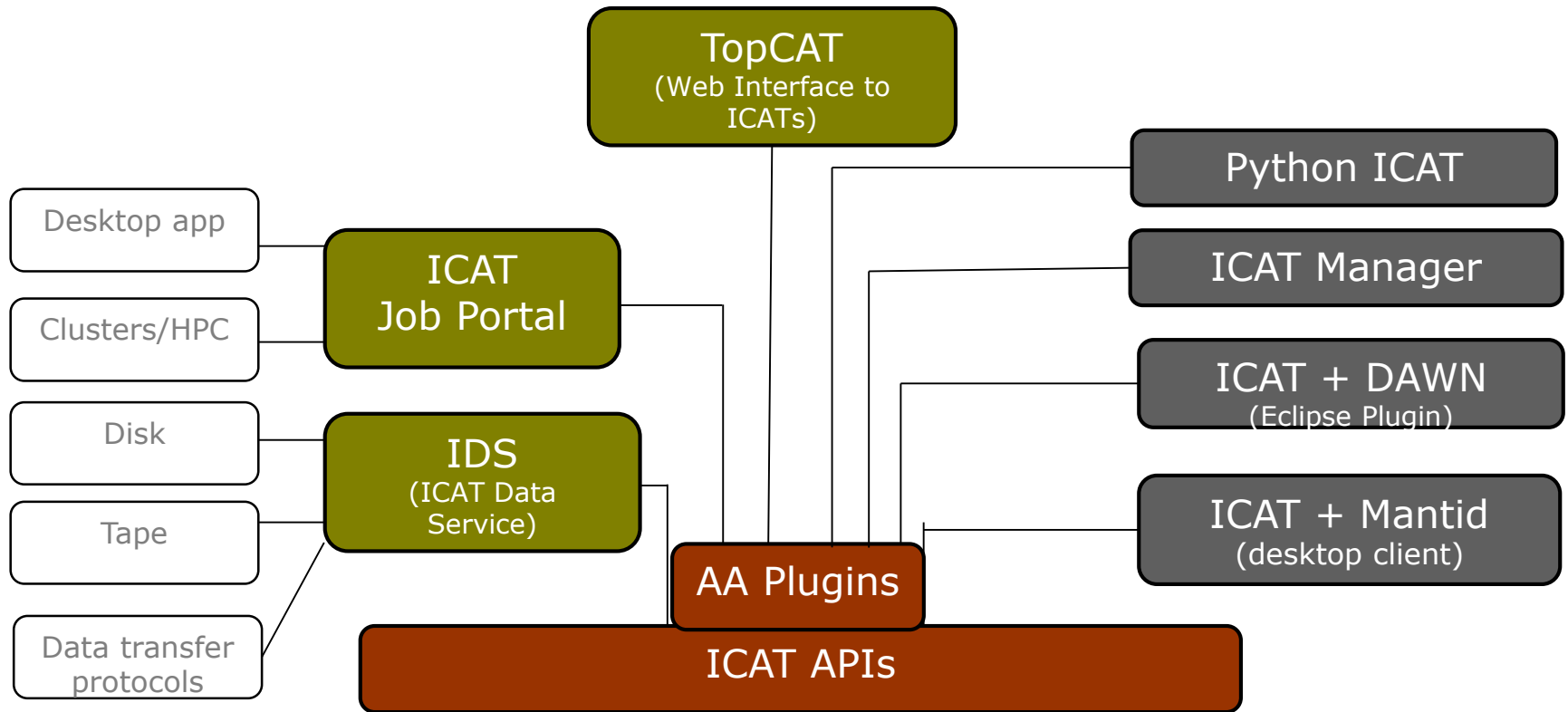


Science & Technology  
Facilities Council



# ICAT Tool Suite and Clients

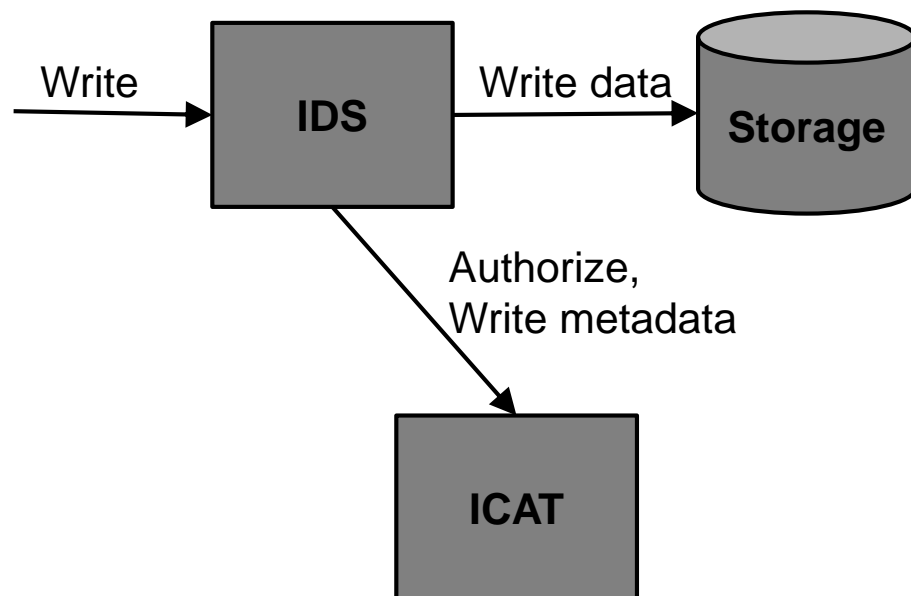
## *Metadata as Middleware*



ICAT

# The ICAT Data Server (IDS)

- ICAT metadata catalogue
  - a SOAP web service interface to metadata
- IDS provides a “RESTful” interface to the data files cataloged by ICAT
  - AA handled via the ICAT
  - Can plugin to different storage infrastructure
  - Can use different data transfer protocols (http, gftp, GlobusOnline ...)
- Separation of concerns: metadata management vs data ingest/access
- Manage data scaling issues



# ICAT: An international collaboration

- In daily production use on the RAL campus:
  - CLF, ISIS, DLS
- Also internationally:
  - In production: ESRF, ILL, SNS,
  - Pre-production: HZB, ALBA
  - Development: PSI, ELLETRA (FERMI)
  - PaNData Consortium
- Actively contributing to tool development
  - E.g. python library
- ICAT steering committee has been established.
  - Andy Götz (ESRF) the chairman

<http://icatproject.org>

<http://code.google.com/p/icatproject/>

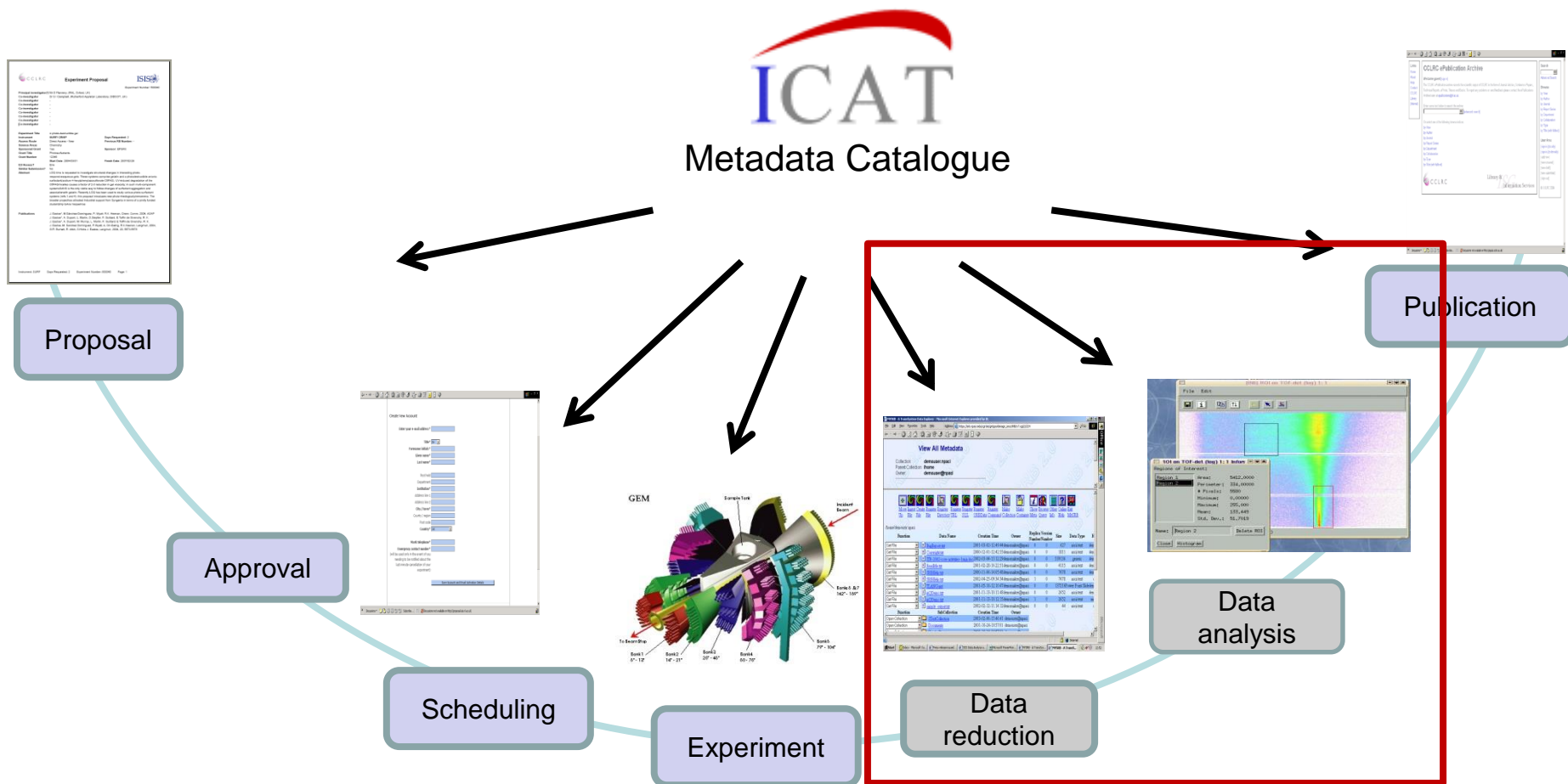


Science & Technology  
Facilities Council

- What we want to do : Supporting user workflows



# Facility Data Lifecycle



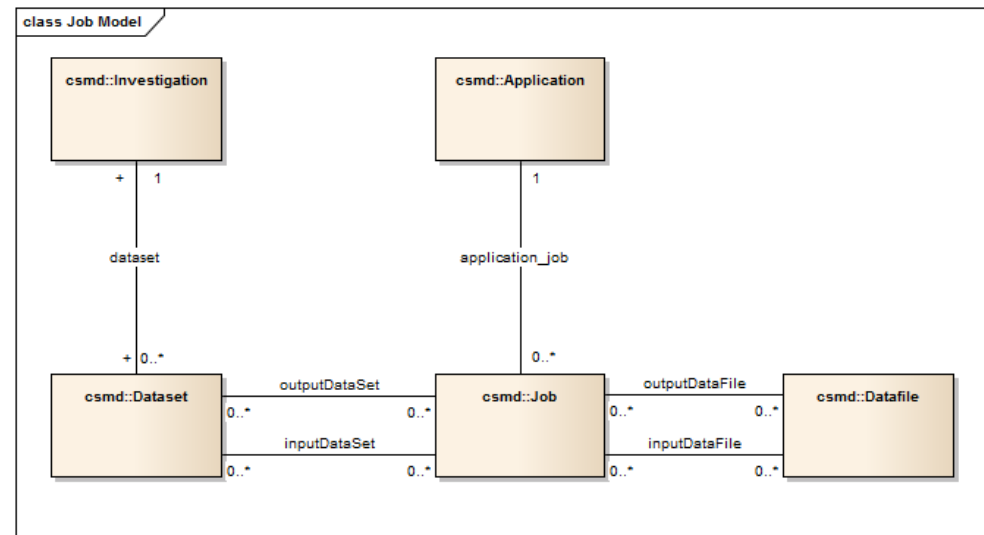
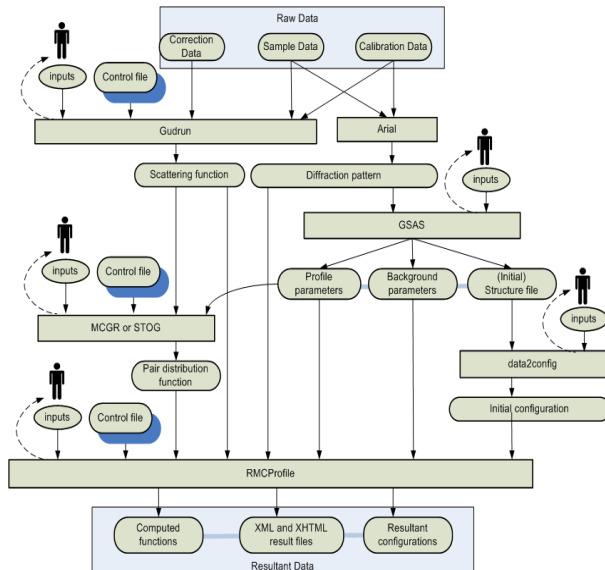
Traditionally, these steps are decoupled from facilities. However, they are key to derive useful insights.

# Data Analysis Challenges

- Diverse science
- Varying levels of expertise
  - Help users through the analysis
- Data getting bigger – too big too move
- High CPU / memory requirements
- Complex software environments
- Open data / reuse – provenance
- Automation

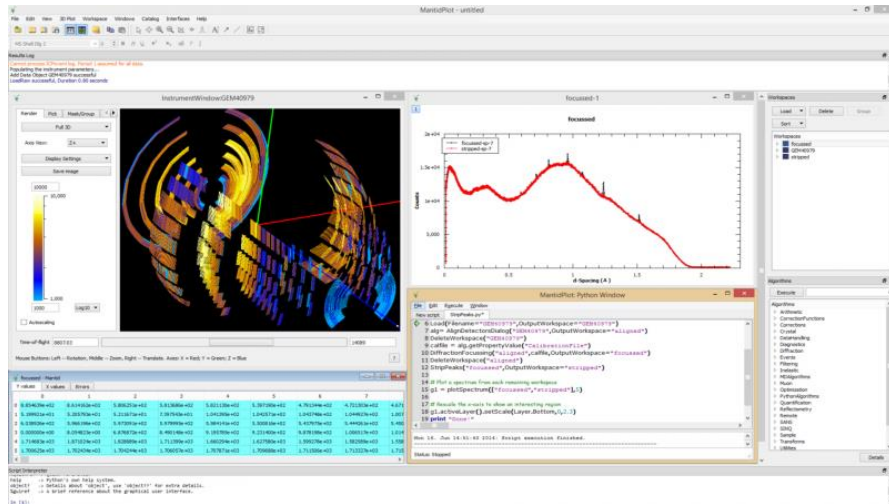
# Supporting Data Analysis

- Managing analysis codes for external users
- Accessing HPC
- Tracking provenance
- Modified ICAT to support:
  - Derived data
  - Software, jobs
  - Linking between these
- Modification to the metadata model



# Tools to support analysis processes

- MANTID
- ICAT Job Portal
- ISIS Auto-Reduction
- All use ICAT to :
  - Access data
  - Record Provenance Steps



ICAT Job Portal

7 datasets found

Name	Description	Users
20120624_0002_0001_632c1ef9-9f52-4a39-a649-85ed5692c27	OctopusSM2	
20120625_0004_0001_088336de-8d79-4c13-84ca-72af8a86ed34	coinc 3 Affibodyx T47D	
20120624_0002_0001_e4210ec3-d7eb-4e33-baea-b969ed31988	T47D 3 Affibodyx 639 nm laser	
20120625_0004_0001_6e2b3e05-9a99-4d5a-9307-d1a9f2a35912	coinc 3 Affibodyx T47D	
20120624_0002_0001_11b3b055-07e5-4aaf-ba3f-a926391812e	T47D 3 Affibodyx 639 nm laser	
20120624_0002_0001_c5a9e970-8451-4d6f-9e06-b23878096d1d	T47D 3 Affibodyx 639 nm laser	
20120624_0002_0001_aee07cbe-dc7d-4b6c-a599-6e2e64829e	T47D 3 Affibodyx 639 nm laser	

endDate	2012-11-27T14:18:17Z
experiment_type	Undefined
id	7201
instrument	OctopusSM3
location	Dummy investigation 1/20120624_0002_0001_aee07cbe-dc7d-4b6c-a599-6e2e64829e
name	20120624_0002_0001_aee07cbe-dc7d-4b6c-a599-6e2e64829e
nchannels	1
nframes	571
sampledescription	T47D 3 Affibodyx 639 nm laser
startDate	2012-11-27T14:16:21Z

Science & Technology  
Fusion Council

Welcome, Marcus [All Jobs](#) [Job Queue](#) [Admin](#) [Logout](#)

## ISIS Auto-reduction

### MARI

Status: All reduction jobs complete

View upcoming saved variable changes

By Experiment Reference Number  By Run Number Range

Run Number Start:  Finished (Optional):

sample\_mn:

energy\_bins:

incident\_energy:

sample\_mass:

sum\_runs: ☐

monovan\_run:

wb\_run:

sample\_run:

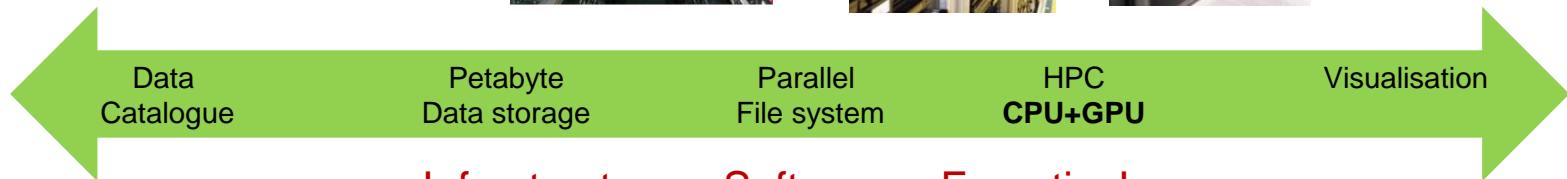
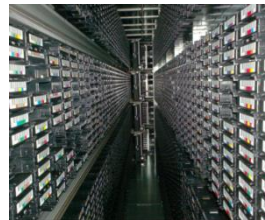
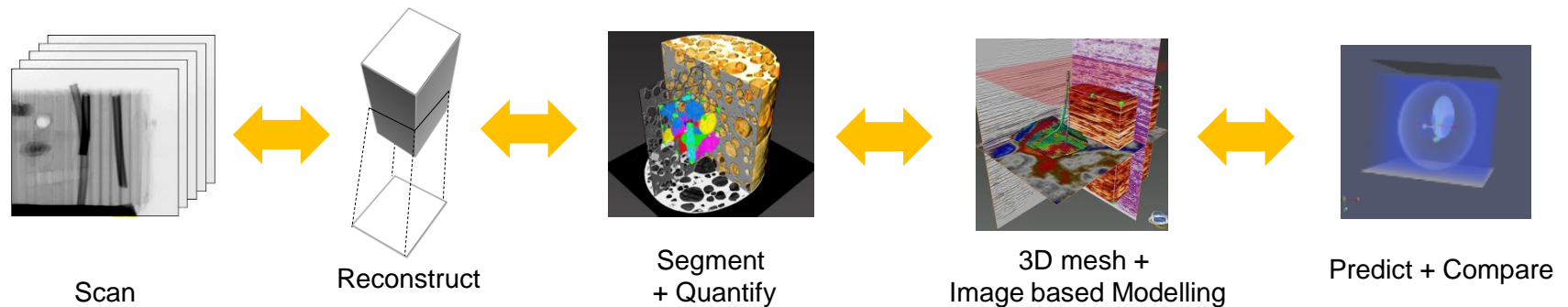
Advanced Variables

Cancel Submit Changes

If you require any help with this website please visit the [Help Pages](#) or [Contact Us](#)

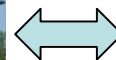
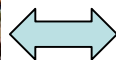


# In- and Post-experimental support



**Infrastructure + Software + Expertise!**

**ISIS:IMAT**



**DLS:I12/I13**

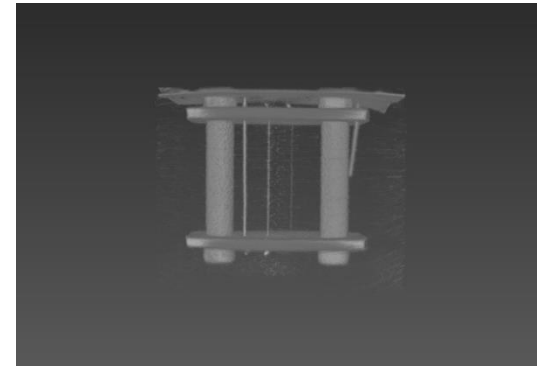


- **Tomography**: Dealing with high data volumes – 200Gb/scan, ~5 TB/day (one experiment at DLS)
- **MX**: high data volumes, smaller files, but a lot more experiments
- Hard to move the data – needs to be handled at the facility?

Erica Yang, Sri Nagella

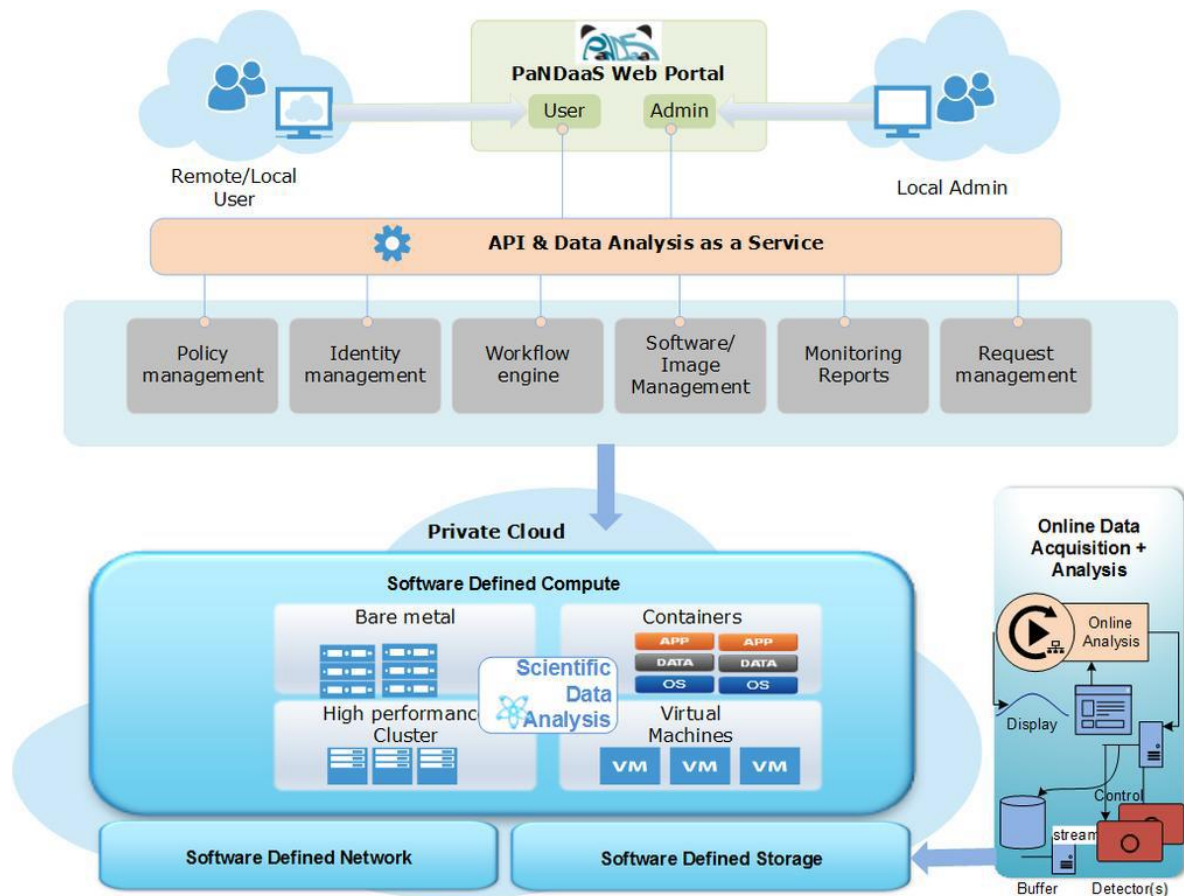
# Tomography Reconstruction for IMAT

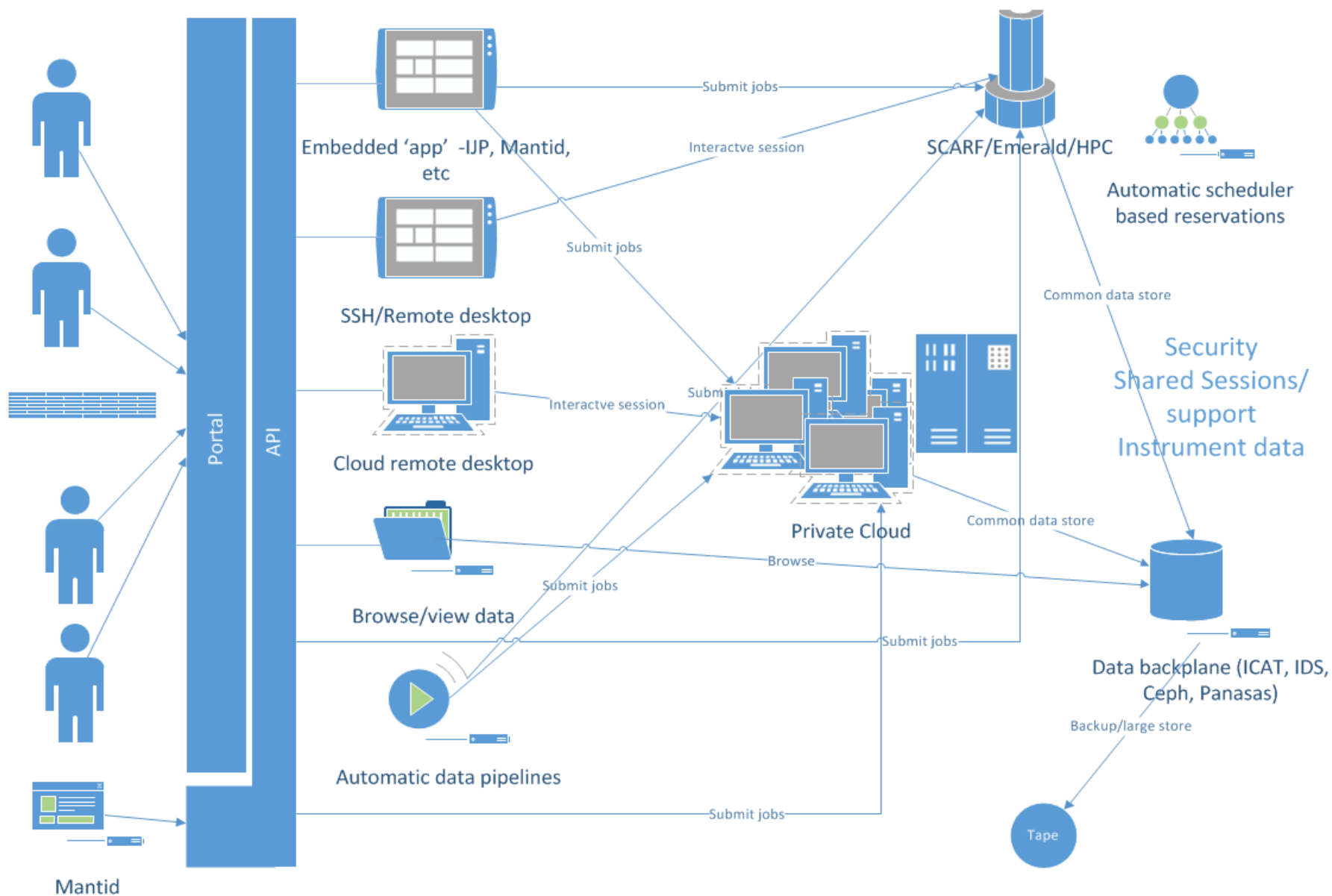
- In- (ISIS) and post-experiment (ISIS and DLS) data processing.
  - IMAT is a new neutron imaging instrument on ISIS
- HPC integration with experiments;
  - Using SCARF CPU and GPU clusters
- A tomographic image reconstruction toolbox
  - With supported algorithms;
- High throughput image reconstruction framework;
  - With fast 3D visualisation;
- An integral component of IMAT's in-experiment data analysis capability through Mantid (ISIS) and DAWN (DLS),
- Maximise the science resulting from Data collected on facility instruments.
- Towards a service in 2015/2016



# PanDaas

- Data Analysis as a Service
  - Led by ESRF
  - 18 institutes worldwide
- Data reduction and analysis platform  
Photon and Neutron analytical facilities
- Not funded, but a continuing need
  - Looking to continue.





- Sharing and Publishing Data





# Data Publication

PHYSICAL REVIEW B 84, 075219 (2011)

## Thickness-dependent magnetic properties of oxygen-deficient EuO

M. Barbagallo,<sup>1,\*</sup> T. Stollenwerk,<sup>2</sup> J. Kroha,<sup>2</sup> N.-J. Steinke,<sup>1</sup> N. D. M. Hine,<sup>1,3</sup> J. F. K. Cooper,<sup>1</sup> C. H. W. Barnes,<sup>1,3</sup> A. Ionescu,<sup>1</sup> P. M. D. S. Monteiro,<sup>1</sup> J.-Y. Kim,<sup>1</sup> K. R. A. Ziebeck,<sup>1</sup> C. J. Kinane,<sup>4</sup> R. M. Dalgliesh,<sup>4</sup> T. R. Charlton,<sup>4</sup> and S. Langridge<sup>4</sup>

<sup>1</sup>Cambridge Laboratory, Physics Department, University of Cambridge, Cambridge CB3 0HE, United Kingdom

<sup>2</sup>Physikalisches Institut und Bethe Center for Theoretical Physics, Universität Bonn, D-53115 Bonn, Germany

<sup>3</sup>Thomas Young Centre, Department of Materials and Department of Physics, Imperial College London, Exhibition Road SW7 2AZ, United Kingdom

<sup>4</sup>ISIS, Harwell Science and Innovation, Chilton, Didcot, Oxfordshire OX11 0QX, United Kingdom  
(Received 24 March 2011)

We have studied how the magnetic properties of thickness. The magnetic moment, measured found to decrease with reducing thickness. On the reduced number of nearest neighbors, band

DOI: 10.1103/PhysRevB.84.075219

## I. INTRODUCTION

Electron-doped EuO is a semiconductor which undergoes simultaneous ferromagnetic and insulating-conducting transition, across which the resistivity drops by 8 orders of magnitude<sup>1,2</sup> and the conduction electrons are nearly 100% spin polarized,<sup>3,4</sup> making EuO a candidate for efficient spin filtering.<sup>5,6</sup> Electron doping increases Curie temperature of EuO thin films to above 200 K/70 K for undoped EuO, and also increases the magnetic moment up to 7.13  $\mu_B$  from the intrinsic value of 7. This is due to the enhanced, conduction-electron-mediated Ruderman-Kittel-Kasuya-Yosida (RKKY) coupling between the Eu 4f spins.<sup>8,9</sup> In thin films and interfaces, these additional factors, such as surface-induced modification of crystalline environment and of the band structure,<sup>10</sup> as well as magnetic proximity effects,<sup>11-13</sup> These interface effects have been studied experimentally mainly in 3d systems, but itinerant ferromagnets<sup>10</sup> or transition metal oxides,<sup>14,15</sup> interfaces of the 4f compound EuO have only been studied theoretically.<sup>16,17</sup>

We have studied systematically the Curie temperature and magnetic moment per Eu atom,  $m(d)$ , in dependence of thickness  $d$  of layers of oxygen-deficient EuO<sub>0.96</sub>, interleaved with Pt capping layers. In the thickness range from 2 to 12 nm we find a systematic reduction of both  $T_C(d)$  and  $m(d)$  decreasing  $d$ , while our previous investigation in the range from 7 to 12 nm for various oxygen-vacancy concentrations did not show a thickness-dependent variation of these magnetic properties. We find that band bending, the reduced number of nearest neighbors at the interface, and a spatially nonuniform spin-exchange coupling are the primary causes of the thickness dependence of  $T_C(d)$  and  $m(d)$ , due to the increased importance of the interface. We are then able to estimate the extension of 9 nm for the effective spin coupling in EuO.

This paper is organized as follows. In Sec. II, we discuss the growth process and the experimental details of the measurement techniques. Section III discusses the experimental results, in particular the thickness-dependent measurement of the magnetic moment and  $T_C(d)$  of EuO<sub>1-x</sub>.

1098-0121/2011/84(7)075219(5)

## Metadata Search beta

DataCite  
Filter  
allocator  
datacentre  
prefix  
resourceType  
contributor  
creator  
publicationYear  
publisher  
language  
refQuality  
has\_metadata

No active filters  
Documents found in 31ms  
Page 1 of 1

RB820322: Magnetic moment of EuO in spin

doi:10.5286/ISIS.E.24066298

Easton, S. • Barnes, C. H. W. • Ionescu, A.

BL STFC - Science and Technology Facilities Council

RB820486: Electric field effect on the

doi:10.5286/ISIS.E.24066298

Steinke, N. J.

BL STFC - Science and Technology Facilities Council

RB1010380: Interaction of the conduction

doi:10.5286/ISIS.E.24079772

Meenman, F. P. S.

BL STFC - Science and Technology Facilities Council

GBS 20.7GHz slant path radio propagation

doi:10.5286/ISIS.E.24079772

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Science and Technology Facilities Council

Firefox

STFC Data home | Science & Technol...

https://data.isis.stfc.ac.uk/doi/INVESTIGATION/2406

Science & Technology Facilities Council

Funding Research Innovation Skills Public Engagement

ISIS Data

Investigation title: Magnetic moment of EuO in spin

Creator: Easton, S.  
Creator: Griffin, T.  
Creator: Barnes, C. H. W.  
Creator: Ionescu, A.

DOI: 10.5286/ISIS.E.24066298

Date of Experiment: Thu Feb 19 13:34:31 GMT 2009

Publisher: STFC ISIS Facility

Data format: RAW/Nexus

Select the data format above to find out more about it.

Data Citation

The recommended format for citing this dataset in a [author], [date], [title], [publisher], [doi]

For Example:  
Easton, S., et al; (2009); 820232, STFC ISIS Facility, doi:10.5286/ISIS.E.24066298

Abstract

EuO is the ferromagnetic oxide semiconductor with the makes it at present one of the most promising materials for study the tunnelling of single electrons in quantum devices. Our group on ferromagnetic thin film structures with the Laboratory in Cambridge. In this light we strongly believe electrodes such as NiFe, Co and Y, and with substrate of EuO is influenced by and influences the adjacent layers.

Contact us Cookies/Privacy Terms & conditions Cymru

Science & Technology Facilities Council

Browse All Data

Download

ISIS

- ALF
- ARGUS
- CRISP
- EMU
- ENGINX
- EVS
- GEM
  - cycle\_11\_4
  - cycle\_11\_3
    - BaruO3 8mm pos 8 (id: CAL\_GEM\_2011-10-31T09:01:37)
      - GEM56174.raw
      - GEM56174.log
      - GEM56174\_ICPreout.bit
      - GEM56174\_ICPrevent.bit
      - GEM56174\_ICStatus.bit
      - GEM56174\_Status.bit
    - GEM56175.raw
    - GEM56176.raw
    - GEM56174.mvs
    - GEM56175.mvs
    - GEM56176.mvs
  - Empty 6mm can 620 (id: CAL\_GEM\_2011-10-10T17:23:28)
  - Empty GEM (position 4) (id: CAL\_GEM\_2011-10-10T13:53:49)
  - Empty SampleChanger pos 1 (id: CAL\_GEM\_2011-10-30T14:51:29)
  - NIIST Silicon SRM 640c 8mm can (id: CAL\_GEM\_2011-10-11T08:58:43)
  - PtTiO3 pos 4 (id: CAL\_GEM\_2011-10-30T19:03:23)

Username

Password

Login Cancel

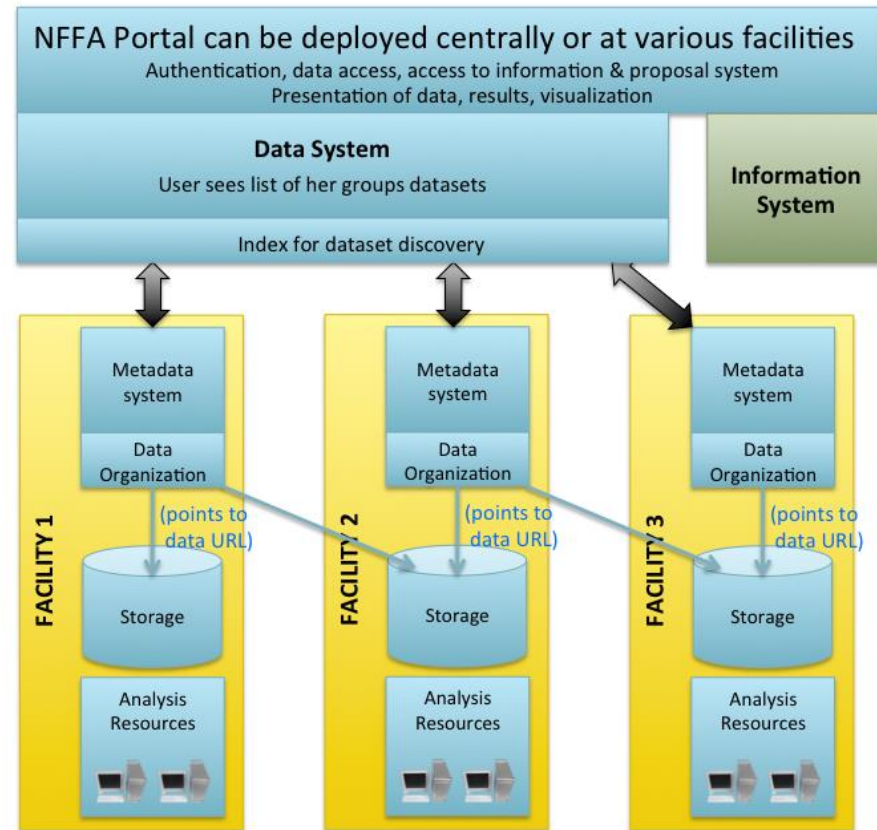
# Publishing and Sharing Metadata

- Publish metadata to general purpose harvesters, and search engines which provide search tools across disciplines
  - Being developed by other e-Infrastructure projects
- Worked with the EUDat project
  - B2Find Data Discovery Service
  - [www.eudat.eu](http://www.eudat.eu)
- Made core metadata available to B2Find
  - OAI-PMH interface
  - Published Data (e.g. with DOIs).
- Mapping of CSMD metadata to Dublin Core and EUDat metadata requirements

	EUDAT Field	ICAT Field(s)
dc:identifier	-	Investigation->doi
dc:title	title	Investigation->title
dc:description	notes	Investigation->summary
dc:relation	tags	Instrument->fullName Investigation->name InvestigationParameter->name (multiple)
dcterms:references	URL	"dx.doi.org/" + Investigation->doi
dc:creator	author	User->fullName
-	spatial	-
dc:contributor	maintainer	Science and Technology Facility Council, ISIS
dc:subject	discipline	"Clean energy and the environment, pharmaceuticals and health care, nanotechnology and materials engineering, catalysis and polymers, fundamental studies of materials"
-	PublicationYear	-
dcterms:issued	PublicationTimestamp	Investigation->releaseDate
dcterms:temporal	TemporalCoverage:EndDate	Investigation->startDate Investigation->endDate

# NFFA-EUROPE

- Nanoscience Foundries and Fine Analysis
  - Research and Innovation actions
- Integrated, distributed research infrastructure
  - for multidisciplinary research at the nanoscale
  - from synthesis and nano-lithography
  - Nano-characterization, theoretical modelling and numerical simulation,
  - coordinated open-access to complementary facilities
- Information and Data management Repository Platform (IDRP)
  - CNR-IOM, ESRF, STFC, KIT
- RDA standardisation





- Back to metadata



# 3 Levels of Metadata

- Discovery
  - General low-detail metadata
  - search engines and aggregators
  - Dublin Core, CKAN, EUDat, DataCite
  - Dryad, Figshare, Zenodo
  - PIDs and DOIs
  - Domain specific terms
- Access
  - How data is organised
  - Who it belongs to and how to access
  - What was done to it – provenance
  - Can be used in data management processes.
  - CSMD, DCAT, CERIF, PROV-O
- Usage
  - Sample, instrument, technique details
  - Controlled vocabularies
  - ESRF approach
  - CIF, NeXus



# NFFA-Europe: Metadata Management

*To develop metadata standards for the cataloguing, access and exchange of data and associated information describing nano-science experiments*

- In support of Information and Data management Repository Platform Underpins the data discovery and sharing services
- Work within the Research Data Alliance [www.rd-alliance.org](http://www.rd-alliance.org)
  - Organisation for sharing and developing best practise in research data management
  - Working with the existing Materials IG and Photon and Neutron Science IG, Metadata WG - may work through these groups
- Starting points :
  - EUDat, CSMD, CIF, Nexus
  - COData Framework for Nanostructures



# Plug: CoData Data Journal

- Recently Relaunched
  - dedicated to the advancement of data science and its application in policies, practices and management as Open
  - descriptions of data systems, their implementations and their publication, applications, infrastructures, software, legal, reproducibility and transparency issues, the availability and usability of complex datasets,
  - principles, policies and practices for data.
- Section Editor for large scale data facilities, data intensive research and data management



# Conclusion

- Management of large amounts of raw data complex
  - Good systematic metadata collection
  - Automation
  - Track what happens to data too
- Need to extend support across the lifecycle
  - Data analysis and publication
  - Support the whole research object
- Metadata at different levels,
  - Discovery, Access, Use
- MetaData as an active part of the computing infrastructure

