



Managing Research Data for Diverse Scientific Experiments

Erica Yang

erica.yang@stfc.ac.uk

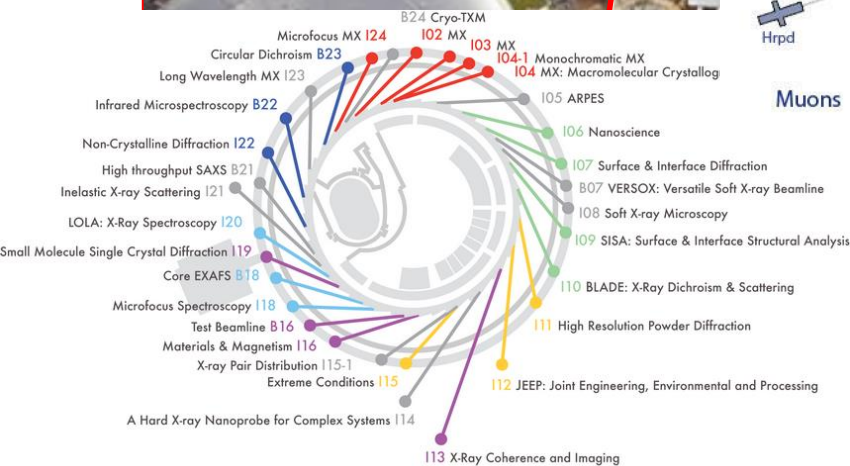
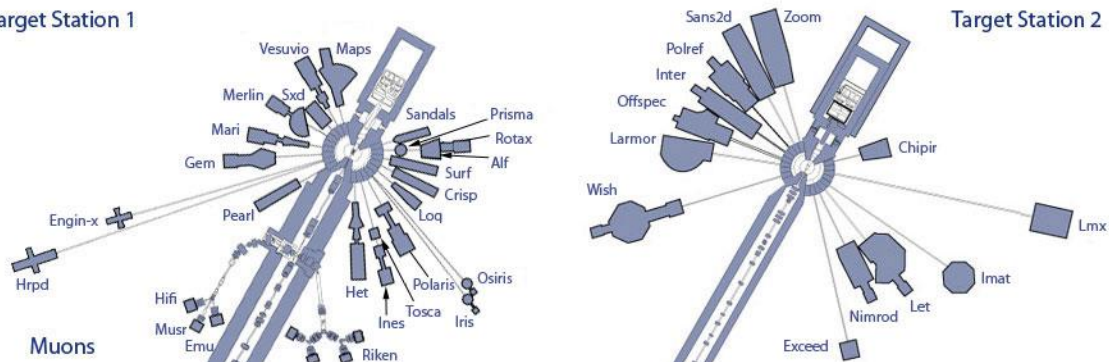
Scientific Computing Department
STFC Rutherford Appleton Laboratory

Crystallographic Information and Data Management Symposium
the 28th European Crystallographic Meeting
25 August 2013, Warwick University, U.K.

STFC Rutherford Appleton Laboratory



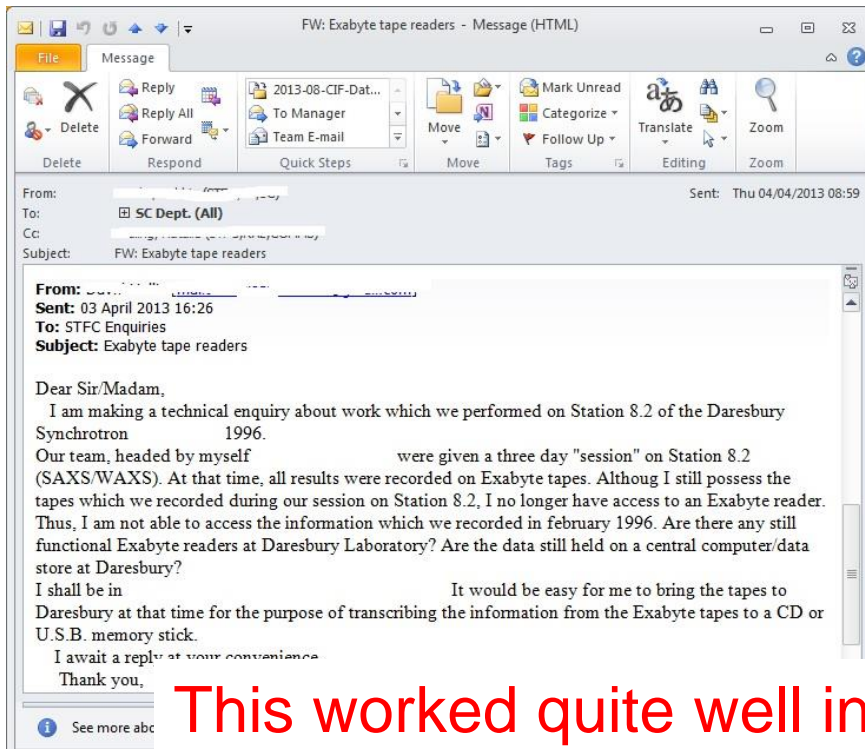
Target Station 1



Once upon a time ...



- Emails, portable disks, a simple web page were all you need.



This worked quite well in the first 20 or so years of ISIS.

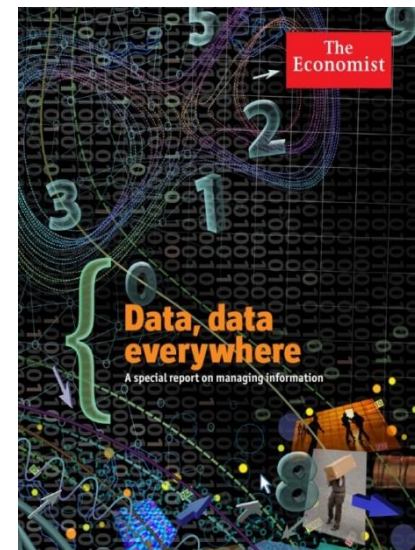
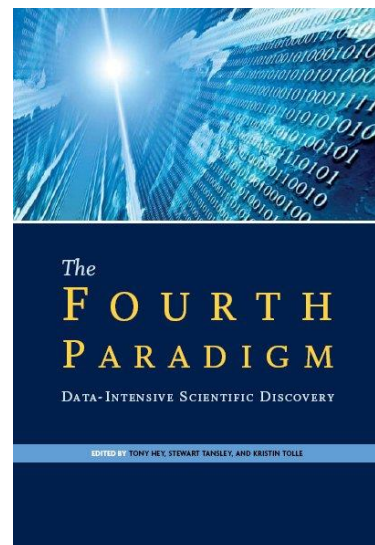
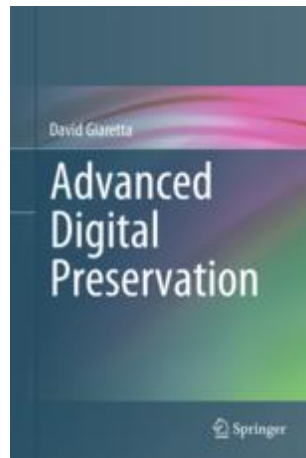
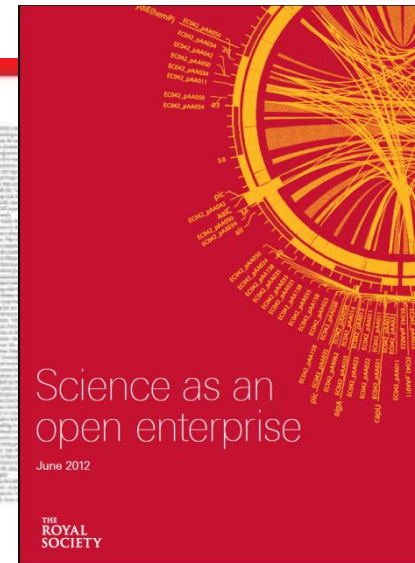


Data Infrastructure

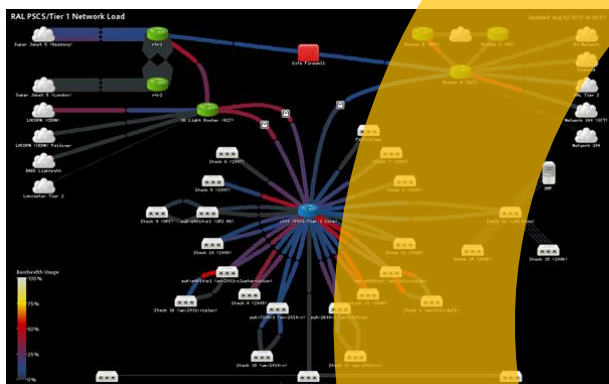
The age of managed data at RAL

The paradigm, societal, and technological changes over time
have made a major impact

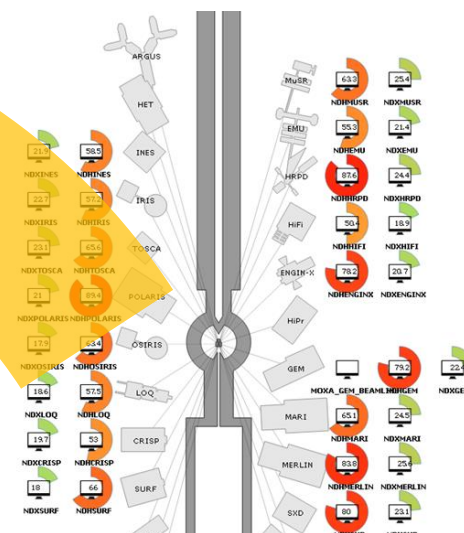
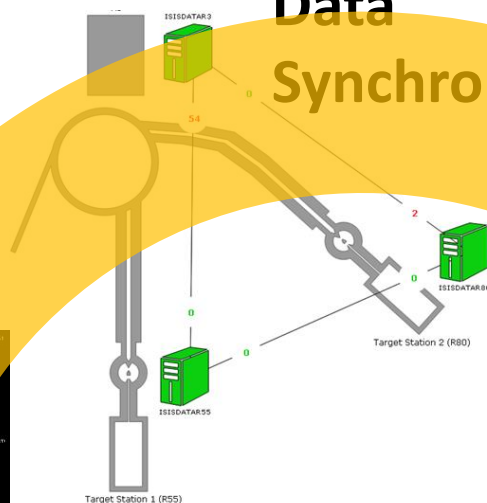
- UK eScience programme
- The 4th paradigm: Data-Intensive Scientific Discovery
- Data, data everywhere
- Digital Preservation
- Royal Society Open Data Report
- Continued developments at the facilities



**Network
monitoring**



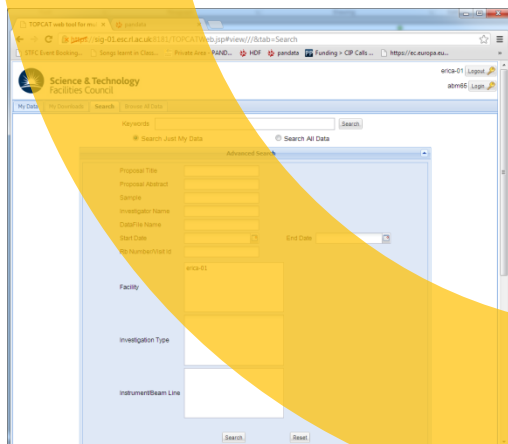
**Data
Synchronisation**



Data monitoring

Now ...

**Data
Cataloguing**

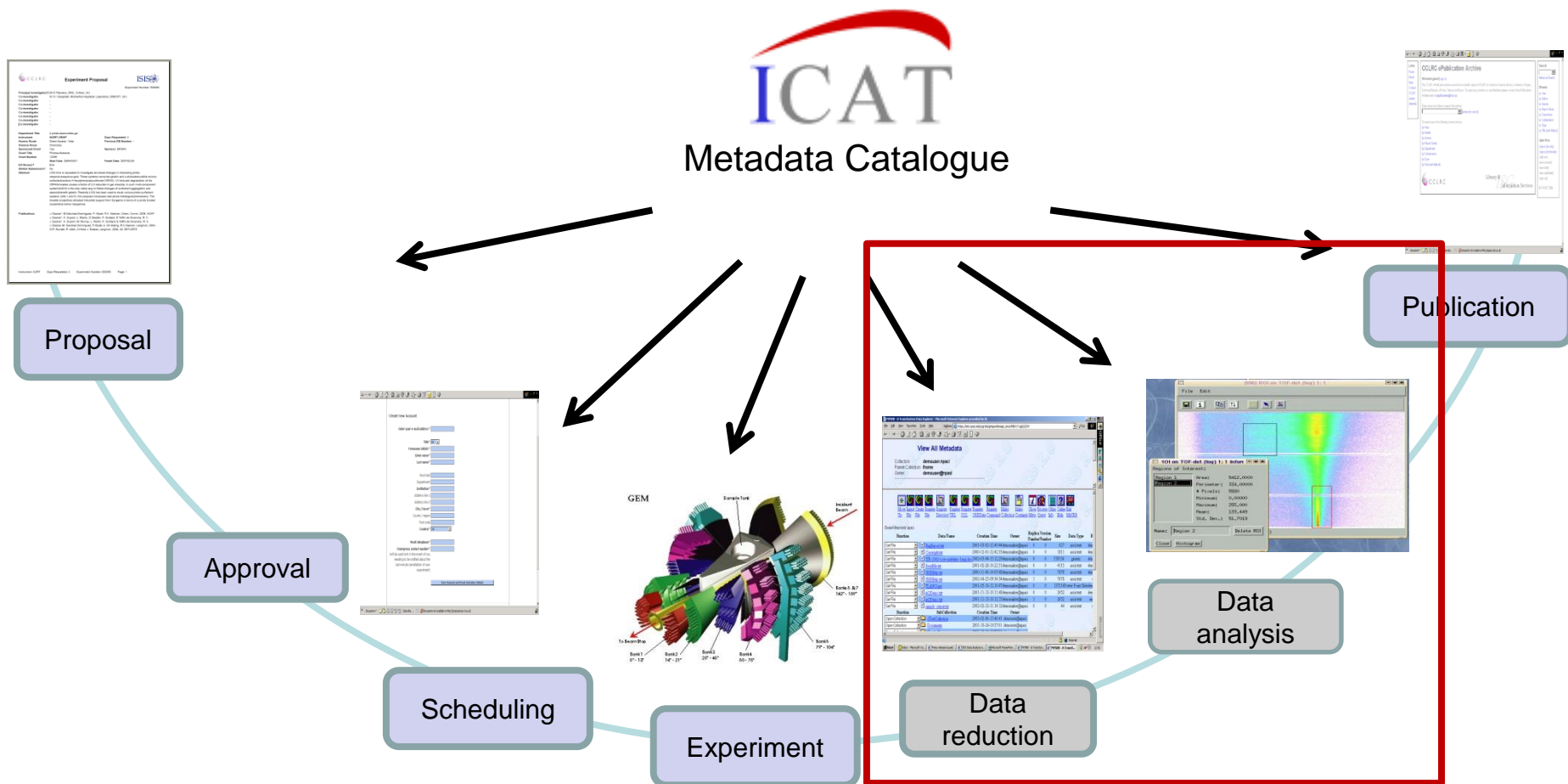


Data archive



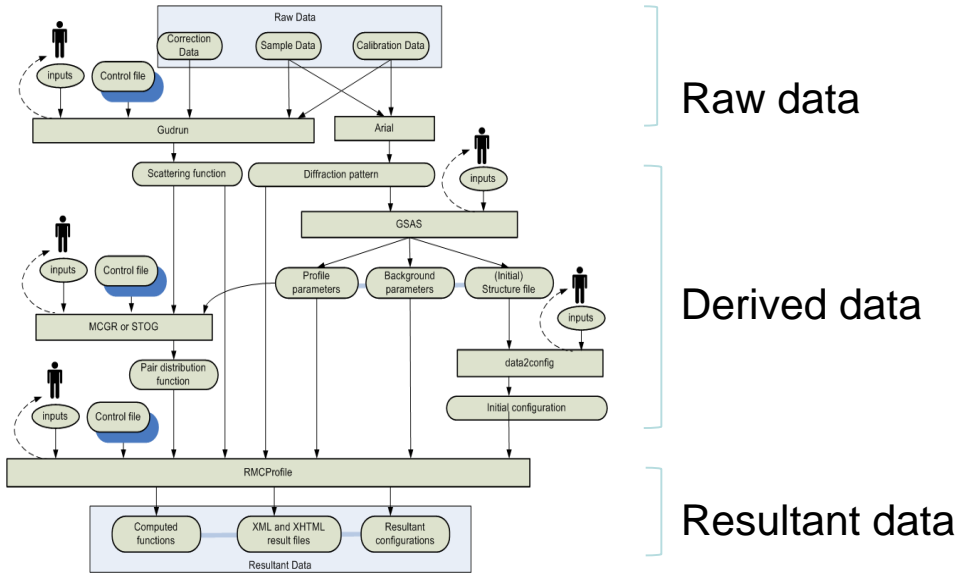
Data Management and Tools

Facility Data Lifecycle



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

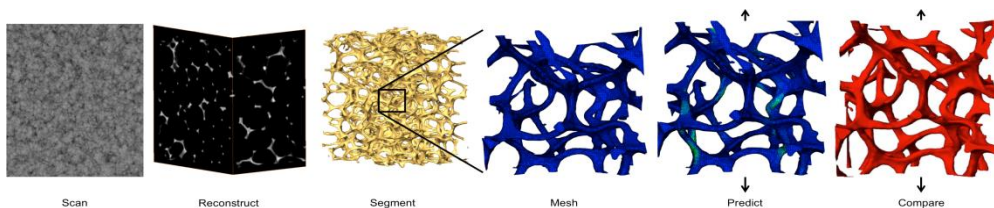
Managing Data Processing Pipelines



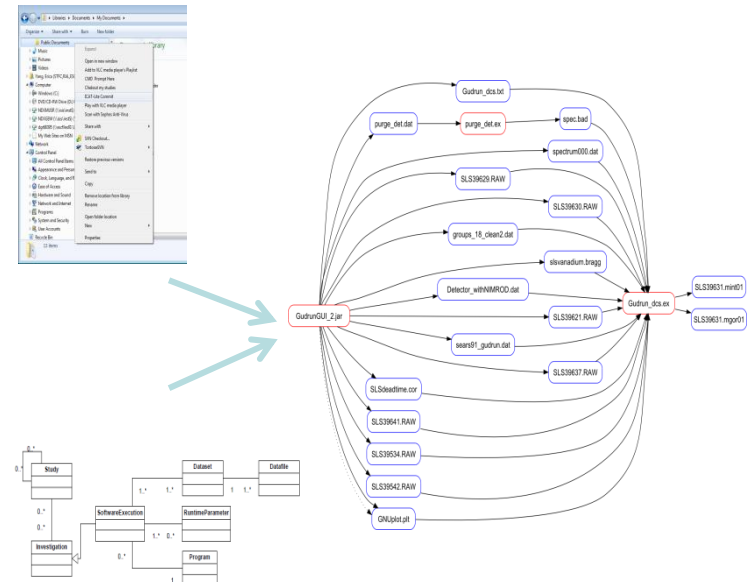
Credits: Martin Dove, Erica Yang (Nov. 2009)

Issues:

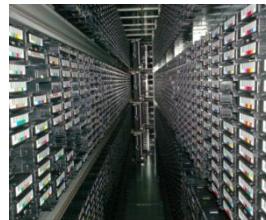
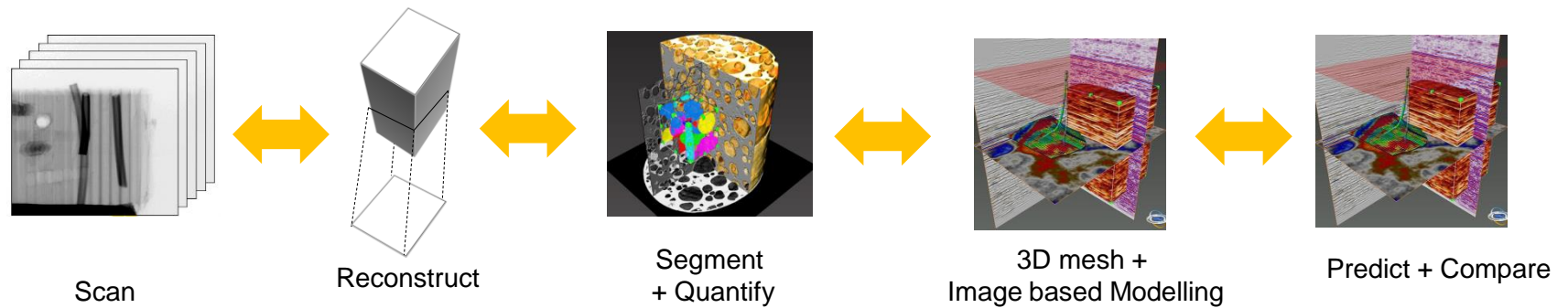
1. Valuable data amongst noise
2. Software version
3. Data provenance
4. Distributed analysis
5. Complex and dynamic workflows
6. Usability of tools



Credit: Phil Withers, Andy Alderson, Sam McDonald



Infrastructure for managing data flows



Data
Catalogue

Petabyte
Data storage

Parallel
File system

HPC
CPU+GPU

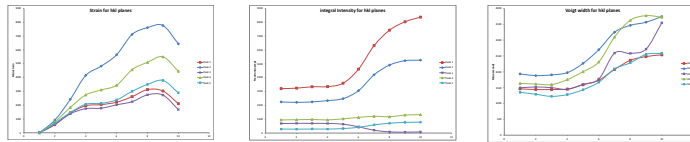
Visualisation

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?

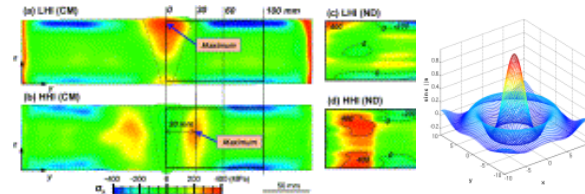
Managing Processed Data



Excel

Standalone
web client

Hosted
web client



Stressing_Pawley_with_data.xlsx - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J
	Run no.	Stress	Macro strain	Lattice a	Lattice a error	Lattice a strain	Lattice b	Lattice b error	Lattice b strain	Lattice c
1	183626	5.012	0	3.59679	8.6188	0	3.59679	8.6188	0	3.5967
2	183627	150.011	0	3.5993	8.89683	697.8444669	3.5993	8.89683	697.8444669	3.599
3	183628	349.997	0	3.60294	11.3991	1709.857957	3.60294	11.3991	1709.857957	3.6029
4	183629	500.01	0	3.60613	19.1838	2596.759889	3.60613	19.1838	2596.759889	3.6061
5	183630	540.112	0	3.60708	23.0761	2860.884288	3.60708	23.0761	2860.884288	3.6070
6	183631	572.536	0	3.60826	28.0806	3188.954596	3.60826	28.0806	3188.954596	3.6082
7	183632	621.92	0	3.61087	36.9774	3914.601631	3.61087	36.9774	3914.601631	3.6108
8	183633	631.616	0	3.61282	41.9819	4456.751715	3.61282	41.9819	4456.751715	3.6128
9	183634	617.984	0	3.61324	43.094	4573.522502	3.61324	43.094	4573.522502	3.6132
10	183635	481.523	0	3.60961	38.3676	3564.289269	3.60961	38.3676	3564.289269	3.6096
11						-1000000			-1000000	
12						-1000000			-1000000	
13						-1000000			-1000000	

Restful APIs

File System

CSV

JSON

XML

Python

HDF lib

Nexus lib

HDF files

Nexus files

HDFView 2.9

Recent Files: local\Dropbox\National-Lab-Liaison-Office\HPP\Projects\Project-1-Accelerator\ISIS\Technical\Data\Pawley-Single-Peak\183626_183635_single_peak_fit_bank_1.hdf

TableView - l_max - fit_details - C:\Applications\Apache Software Foundation\Apache2\2nd...

	0	1	2	3	4
0	18.359553	11.092115	5		
1	18.682916	11.079747	5		
2	19.245849	11.233288	5		
3	18.962958	11.163406	5		
4	18.962964	10.637275	5		
5	21.495496	10.903928	3		

TableView - l_max - fit_results - C:\Applications\Apache Software Foundation\Apache2\2nd...

	0	1	2	3	4
0	48.327091	45.754711	26.500089	24.389825	18.580612
1	44.585853	43.039207	25.164182	22.287389	17.137351
2	44.812271	45.081344	26.146888	22.679500	15.146616
3	44.826503	44.908959	26.647193	26.031783	14.905953
4	48.435089	53.209453	25.103054	29.503213	16.055263
5	44.813581	55.963764	19.819719	29.626274	17.873386
6	44.833437	62.129045	13.188522	39.845481	22.247453
7	42.299530	54.298080	33.751052	51.546260	18.534646
8	38.777549	51.831304	5.6702628	44.201193	22.596108
9	38.331096	55.729583	21.886848	44.627943	25.981116

TableView - gauss_width - fit_results - C:\Applications\Apache Software Foundation\Apache2\2nd...

	0	1	2	3	4
0	17.122560	30.817520	14.707698	14.099	
1	18.544603	32.114364	17.576900	16.631	
2	18.341392	30.491664	15.440402	15.587	
3	19.245304	34.463821	21.131681	15.542	

gauss_width (720, 551)
64-bit floating-point, 10 x 5

Log Info Metadata

MVC: Model, View, Controller



Data Catalogue and Tools

PaN-data ODI– an Open Data Infrastructure for European Photon and Neutron laboratories

Federated data catalogues supporting cross-facility, cross-discipline interaction at the scale of atoms and molecules

- Unification of data management policies
- Shared protocols for exchange of user information
- Common scientific data formats
- Interoperation of data analysis software
- Data Provenance WP: Linking Data and Publications
- Digital Preservation: supporting the long-term preservation of the research outputs



ICAT and CSMD

- The Core Scientific Meta-Data Model (CSMD) is a study-data oriented model which has been developed at STFC since 2004.
 - It captures high level information about scientific studies and the data that they produce throughout a **facility's scientific workflow**.
 - It is a key aspect of the ICAT, a software suite designed to manage the **cataloguing and (continuous) access** to facilities data.
- **Investigation**
 - **Investigator**
 - **Topic and Keyword**
 - **Publication**
 - **Sample**
 - **SampleParameter**
 - **Dataset**
 - **DatasetParameter**
 - **Datafile**
 - **DatafileParameter**
 - **Parameter**

ICAT Schema

[Application](#), [Datafile](#), [DatafileFormat](#), [DatafileParameter](#), [Dataset](#), [DatasetParameter](#), [DatasetType](#), [Facility](#), [FacilityCycle](#), [Group](#), [InputDatafile](#), [InputDataset](#), [Instrument](#), [InstrumentScientist](#), [Investigation](#), [InvestigationParameter](#), [InvestigationType](#), [InvestigationUser](#), [Job](#), [Keyword](#), [NotificationRequest](#), [OutputDatafile](#), [OutputDataset](#), [ParameterType](#), [PermissibleStringValue](#), [Publication](#), [RelatedDatafile](#), [Rule](#), [Sample](#), [SampleParameter](#), [SampleType](#), [Shift](#), [Study](#), [StudyInvestigation](#), [User](#), [UserGroup](#)

<http://www.icatproject.org/mvn/site/icat/4.2.5/icat.core/schema.html>

Nexus and CSMD

2.3. NeXus class definitions » 2.3.1.2. NeXus Class Specifications »

Application Definitions

A description of each NeXus application definition is given. NeXus application definitions define the *minimum* set of terms that *must* be used in an instance of that class. Consider the application definitions as a *contract* between a data provider (such as the beam line control system) and a data consumer (such as a data analysis program for a scientific technique) that describes the information is certain to be available in a data file.

- [NXarchive](#)
- [NXdirecttof](#)
- [NXfluxo](#)
- [NXindirecttof](#)
- [NXiqproc](#)
- [NXlauetof](#)
- [NXmonopd](#)
- [NXrefscan](#)
- [NXreftof](#)
- [NXsas](#)
- [NXsastof](#)
- [NXscan](#)
- [NXspe](#)
- [NXsqom](#)
- [NXtas](#)
- [NXtofnpd](#)
- [NXtofraw](#)
- [NXtofsingle](#)
- [NXtomop](#)
- [NXtomophas](#)
- [NXtomoproc](#)
- [NXxas](#)

[NeXus: Manual 3.1 documentation](#) » [2. NeXus: Reference Documentation](#) » [2.3. NeXus class definitions](#) » [2.3.1.2. NeXus Class Specifications](#) » [Application Definitions](#) »

NXsas

version	category	extends	groups cited
1.0b	application	NXobject	NXcollimator , NXdata , NXdetector , NXentry , NXgeometry , NXinstrument , NXmonitor , NXmonochromator , NXsample , NXshape , NXsource

This is an application definition for raw data (not processed or reduced data) from a 2-D small angle scattering instrument collected with a monochromatic beam and an area detector. It is meant to be suitable both for neutron SANS and X-ray SAXS data.

It covers all raw data from all SAS techniques: SAS, WSAS, grazing incidence, GISAS

symbol list:

No symbol table.

NXDL source:

<http://svn.nexusformat.org/definitions/trunk/applications/NXsas.nxd1.xml>

svnid:

Id: NXsas.nxd.xml 1165 2012-09-27 18:12:40Z Pete Jemian

Structure of NXsas

```

1 NXgas (application definition, version 1.0b)
2 (overlays IXentry)
3 NXentry
4 @entry
5 definition:IX_CHAR
6 end_time:IX_DATE_TIME
7 start_time:IX_DATE_TIME
8 title:IX_CHAR
9 data:IXdata
10 data --> /IXentry/NXInstrument/NXdetector/data
11 instrument:IXInstrument
12 name:IX_CHAR
13 collimator:IXcollimator
14 geometry:IXgeometry
15 shape:IXshape
16 size:IX_CHAR
17 size:IX_FLOAT

```

NeXus

[Previous topic](#)

NXxraylens

Next topic

NXarchive

This Page

[Show Source](#)[previous](#) | [next](#) | [index](#)

NeXus

Table Of Contents

NXsas

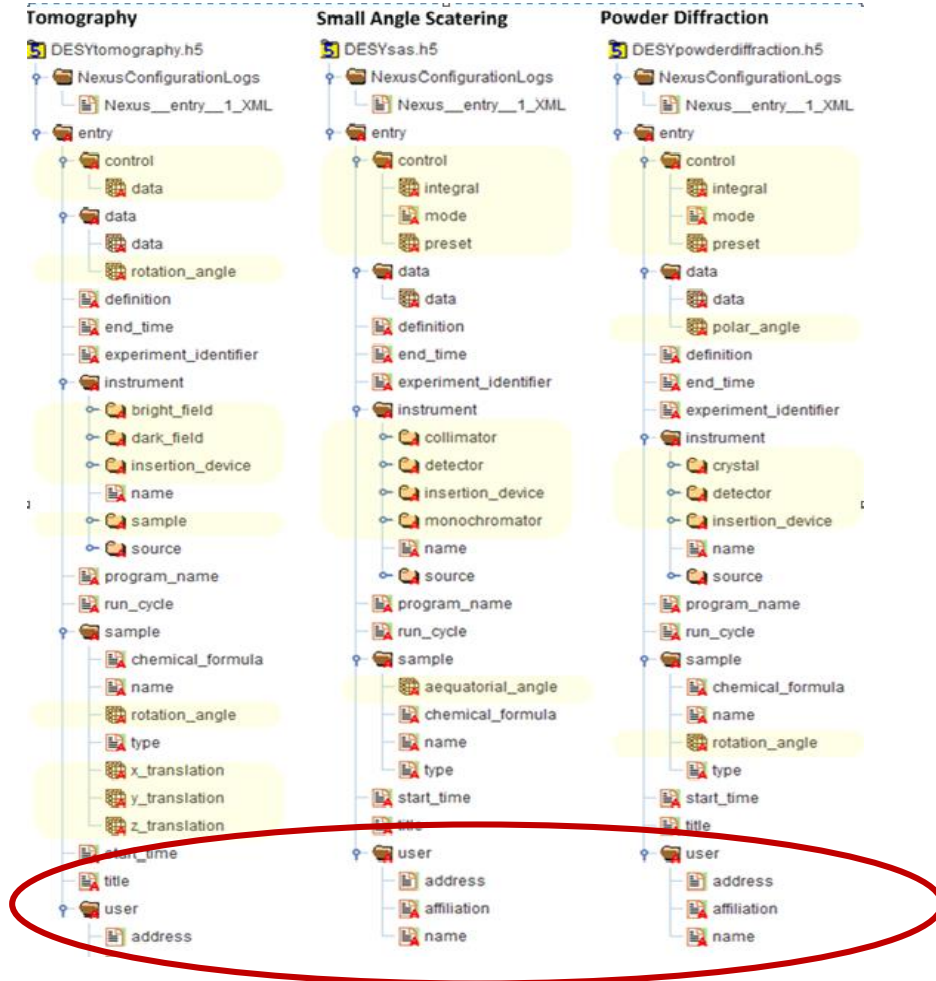
- [illegible]

[Previous topic](#)

NXreftof

Next topic

NXsastof

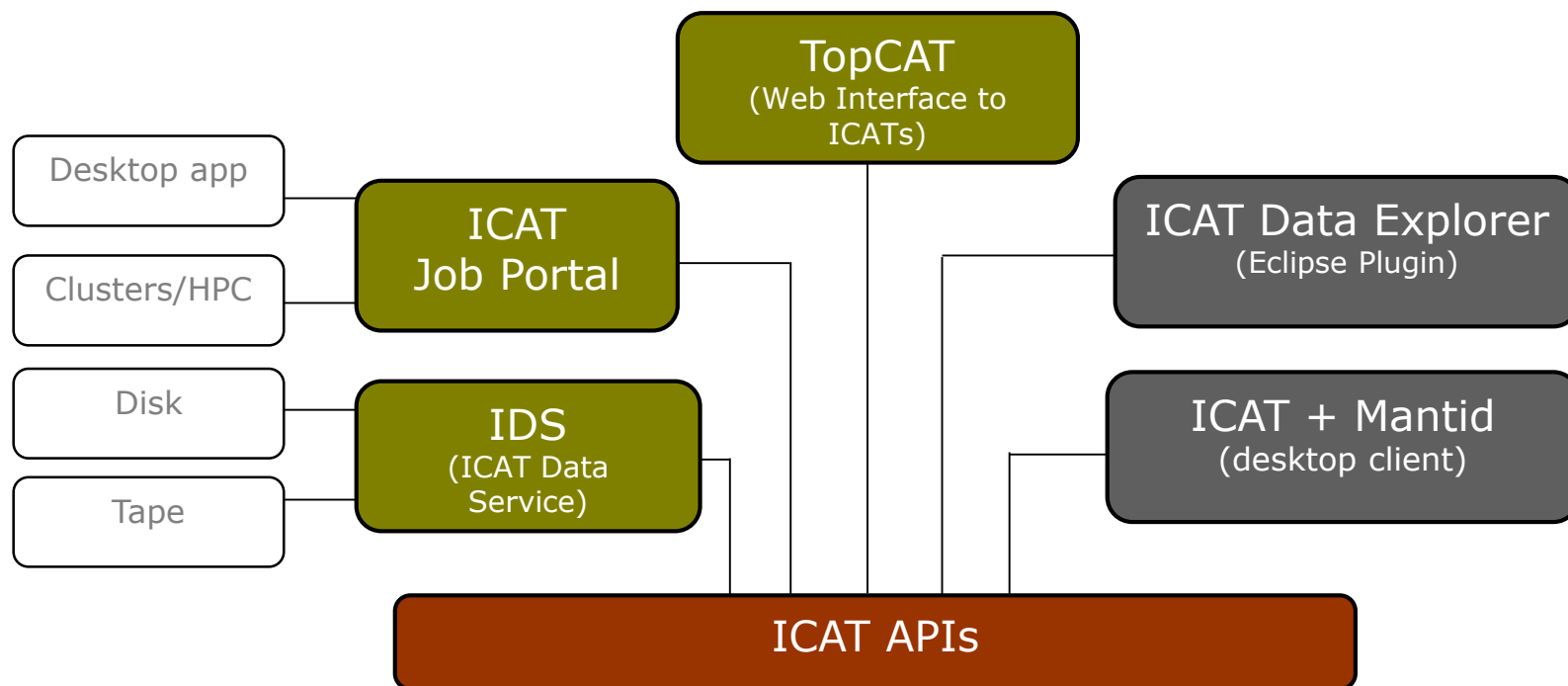


Nexus Application Profile for SAS

<http://download.nexusformat.org/>

PaNdata-ODI deliverable

ICAT Tool Suite and Clients



<http://www.mantidproject.org/>
<http://www.dawnsci.org/>
<https://code.google.com/p/icat-job-portal/>

Ontology for Facility Science

Facilities, instruments, and techniques
(applications: cataloguing, searching, and linking)

Diffraction

Neutron Diffraction/Elastic Neutron Scattering

- Powder Diffraction
- Single Crystal Diffraction

X-ray Diffraction

- Grazing Incidence Diffraction
- Powder Diffraction
- Resonant Diffraction
- Small Angle Diffraction
- Single Crystal Diffraction
- Soft Diffraction
- Surface Diffraction

Other

- Coherent Diffraction Imaging
- Diffraction Imaging (Topography)
- Enhanced Diffraction Imaging

Diffusive - Diffusive MRI

Imaging

Holography

Microscopy

- X-Ray Photoemission Microscopy
- X-Ray Scanning Microscopy
- Scanning Transmission X-Ray Microscopy
- Tomographic Microscopy With CRLs

Tomography

- Fluorescence Tomography

The screenshot displays the Protégé ontology editor interface. The main window shows the 'Top Concepts Hierarchy View' for the 'facilities' ontology. The hierarchy is as follows:

- facilities
 - DESY
 - FLASH
 - BL2
 - BL3
 - PG1
 - PG2
 - THz_BL
 - PETRA_III
 - Diamond
 - ESRF
 - ILL
 - ISIS
 - PSI
 - SOLEIL
 - ANTARES
 - CASSIOPEE
 - CRISTAL
 - DEIMOS
 - DESIRS
 - DIFFABS
 - DISCO
 - GALAXIES
 - HERMES
 - ILES
 - LUCIA
 - MARS
 - METROLOGIE
 - NANOSCOPIUM

On the right, the 'SKOS Usage: BL2' panel shows the following information:

- Show: ☒ this ☒ different
- Found 3 uses of BL2
 - BL2
 - BL2 broader FLASH
 - BL2 Type Concept
 - Individual: BL2

At the bottom, there are two panels for SKOS assertions:

- SKOS Object Property Assertion: SKOS related assertion, SKOS in scheme assertion, SKOS top concept of assertion, SKOS broad match assertion, SKOS narrow match assertion, SKOS related match assertion, SKOS exact match assertion.
- SKOS Data Property: SKOS notation, Other property assertions.

The status bar at the bottom indicates: 'No Reasoner set. Select a reasoner from the Reasoner menu' and 'Show Inferences' is checked.



(Open) Data Access

DOI Data Access Process

Paper → DataCite → STFC Page → TopCAT

Physical Review B 84, 075219 (2011)

Thickness-dependent magnetic properties of oxygen-deficient EuO

M. Barbagallo,^{1,*} T. Stollenwerk,² J. Kroha,² N.-J. Steinke,¹ N. D. M. Hine,^{1,3} J. F. K. Cooper,¹ C. H. W. Barnes,^{1,3} A. Ionescu,¹ P. M. D. S. Monteiro,¹ J.-Y. Kim,¹ K. R. A. Ziebeck,¹ C. J. Kinane,⁴ R. M. Dalgliesh,⁴ T. R. Charlton,⁴ and S. Langridge⁴

¹Cambridge Laboratory, Physics Department, University of Cambridge, Cambridge CB3 0HE, United Kingdom
²Physikalisches Institut und Bethe Center for Theoretical Physics, Universität Bonn, D-53115 Bonn, Germany
³Thomas Young Centre, Department of Materials and Department of Physics, Imperial College London, Exhibition Road SW7 2AZ, United Kingdom
⁴ISIS, Harwell Science and Innovation, 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^{1,2} and the conduction electrons are nearly 100% spin polarized,^{3,4} making EuO a candidate for efficient spin filtering.^{5,6} 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 μ_B from the intrinsic value of 7 μ_B . This is due to the enhanced, conduction-electron-mediated Ruderman-Kittel-Kasuya-Yosida (RKKY) coupling between the Eu 4f spins.^{8,9} In thin films and interfaces, these additional factors, such as surface-induced modification of crystalline environment and of the band structure,¹⁰ as well as magnetic proximity effects,¹¹⁻¹³ These interface effects have been studied experimentally mainly in 3d systems, but itinerant ferromagnets¹⁰ or transition metal oxides,^{14,15} interfaces of the 4f compound EuO have only been studied theoretically.^{16,17}

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_{0.96}, 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_{1-x}.

Metadata Search beta

DataCite

Filter

allocator

datacentre

prefix

resourceType

contributor

creator

publicationYear

publisher

language

refQuality

has_metadata

No active filters. 31 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 magnetic moment of EuO in spin

doi:10.5286/ISIS.E.24066298

Steinke, N. J.

BL STFC - Science and Technology Facilities Council

RB1010380: Interaction of the conduction electrons with the magnetic moment of EuO in spin

doi:10.5286/ISIS.E.24079772

Meesman, F. P. S.

BL STFC - Science and Technology Facilities Council

GBS 20.7GHz slant path radio propagation measurement at 40GHz in the United Kingdom

doi:10.5286/ISIS.E.24079772

Science and Technology Facilities Council

GBS 20.7GHz slant path radio propagation measurement at 40GHz in the United Kingdom

doi:10.5286/ISIS.E.24079772

Science and Technology Facilities Council

GBS 20.7GHz slant path radio propagation measurement at 40GHz in the United Kingdom

doi:10.5286/ISIS.E.24079772

Science and Technology Facilities Council

ITALSAT radio propagation measurement at 40GHz in the United Kingdom [version 1.0]

doi:10.5286/ISIS.E.24079772

Science and Technology Facilities Council

ITALSAT radio propagation measurement at 200GHz in the United Kingdom [version 1.0]

doi:10.5286/ISIS.E.24079772

Science and Technology Facilities Council

Science & Technology Facilities Council

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. In this light we strongly believe that the study of the tunnelling of single electrons in quantum devices such as NiFe, Co and Y, and with substrate of EuO is influenced by and influences the adjacent layers.

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_CPreout.bit

GEM56174_CPrevent.bit

GEM56174_CPrestat.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)

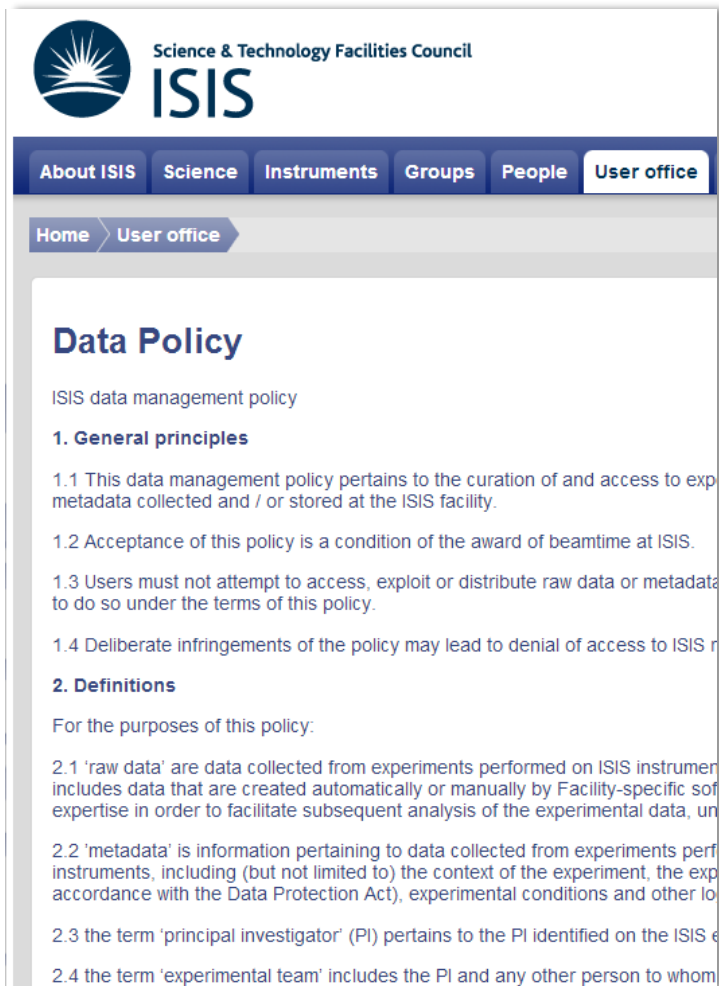
PoTiO3 pos 4 (id: CAL_GEM_2011-10-30T19:03:23)

Username

Password

Login Cancel

Data Access and Open Access



The screenshot shows the ISIS website header with the Science & Technology Facilities Council logo and the ISIS name. A navigation bar includes links for About ISIS, Science, Instruments, Groups, People, and User office. Below this, a breadcrumb trail shows Home > User office. The main content area is titled 'Data Policy' and contains the following text:

ISIS data management policy

1. General principles

1.1 This data management policy pertains to the curation of and access to experimental data and metadata collected and / or stored at the ISIS facility.

1.2 Acceptance of this policy is a condition of the award of beamtime at ISIS.

1.3 Users must not attempt to access, exploit or distribute raw data or metadata to do so under the terms of this policy.

1.4 Deliberate infringements of the policy may lead to denial of access to ISIS resources.

2. Definitions

For the purposes of this policy:

2.1 'raw data' are data collected from experiments performed on ISIS instruments, including data that are created automatically or manually by Facility-specific software, and are stored in a format that allows for subsequent analysis of the experimental data, under the supervision of ISIS experts in order to facilitate subsequent analysis of the experimental data, under the supervision of ISIS experts.

2.2 'metadata' is information pertaining to data collected from experiments performed on ISIS instruments, including (but not limited to) the context of the experiment, the experimental conditions, the experimental results, the experimental conditions and other information relevant to the experiment.

2.3 the term 'principal investigator' (PI) pertains to the PI identified on the ISIS experiment proposal.

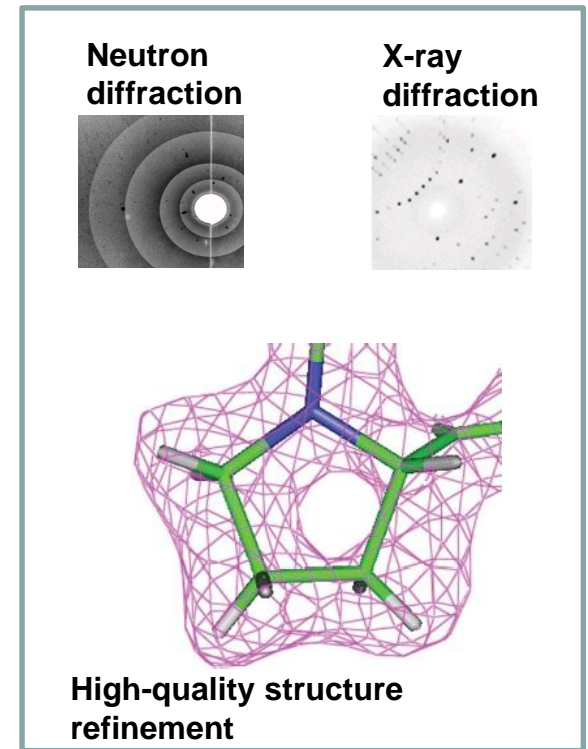
2.4 the term 'experimental team' includes the PI and any other person to whom access to the data is granted.

- Access to the on-line catalogue will be restricted to those who **register** with STFC/ISIS as users of the on-line catalogue.
- Access to raw data and the associated metadata obtained from an experiment is restricted to the experimental team for a period of **three years** after the end of the experiment. Thereafter, it will become publicly accessible.
- The term 'long-term' means a minimum of **ten years**.

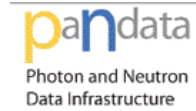
Outlooks

Developments that will influence how the data is managed

- Facilities offer complementary experimental techniques for a single beamline (e.g. tomography+diffraction)
- Users increasingly use multiple facilities leading to the need for multi-stream data fusion and processing
- Computational needs of experiments
- The rise of data intensive experiments and computation
 - Real time data processing for live experiments
 - Streaming data processing



Acknowledgement



- STFC (Scientific Computing and ISIS)
 - Brian Matthews, Steve Fisher, Alistair Mills, Kevin Philips, Anthony Wilson, Tom Griffin, Holly Zhen, Juan Bicarregui, Martin Turner, Ronald Fowler
- Diamond Light Source
 - Mark Basham, Alun Aston, Kaz Wanelik, Robert Atwood
- Manchester University
 - Philip Withers, Peter Lee
- And many others who have contributed to the development of ICAT, CSMD, and the data infrastructure...





Managing Research Data for Diverse Scientific Experiments

Erica Yang

erica.yang@stfc.ac.uk



Science & Technology Facilities Council
Rutherford Appleton Laboratory

