

E-Science DataGrid for DLS and ISIS

Brian Matthews,
Leader, Scientific Applications Group,
E-Science Centre,
STFC Rutherford Appleton Laboratory

brian.matthews@stfc.ac.uk

Acknowledgements

A team effort with many people contributing

Especially:, **Damian Flannery**

Also: Shoaib Sufi, Kerstin Kleese-van-Dam, Michael Gleaves, Glen Drinkwater, Louisa Casely-Hayford, Rik Tyer, Ken Shankland, Gordon Brown, Carmine Cioffi, Alun Ashton, Bill Pulford, Rob Allan, Lisa Blanchard, Laurent Lerusse, Roger Downing, Juan Bicarregui, and others...

STFC e-Science Centre

Exploit e-Science technologies throughout STFC's programmes, the research communities they support and the national science and engineering base.

- Especially ISIS (Neutron Spallation Source), DLS (Synchrotron X-Rays), CLF (Lasers), CERN
- Grid, HPC, Data storage, Libraries, Data Management, Visualisation
- R&D programme
- c.80 staff

<http://www.e-science.clrc.ac.uk/>



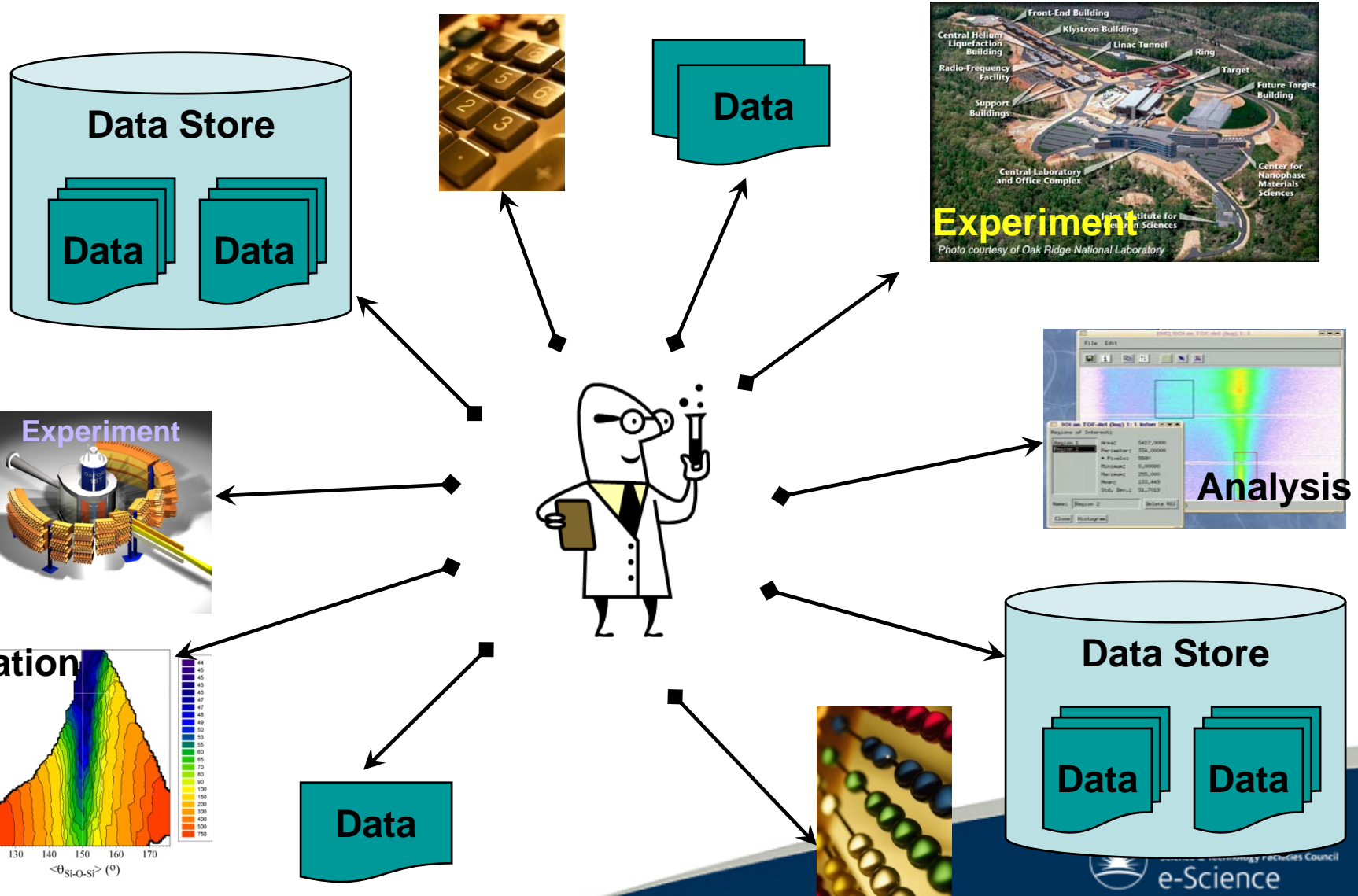
Grid@STFC

STFC (in particular eScience) takes part in many of the Grid initiatives we heard about yesterday

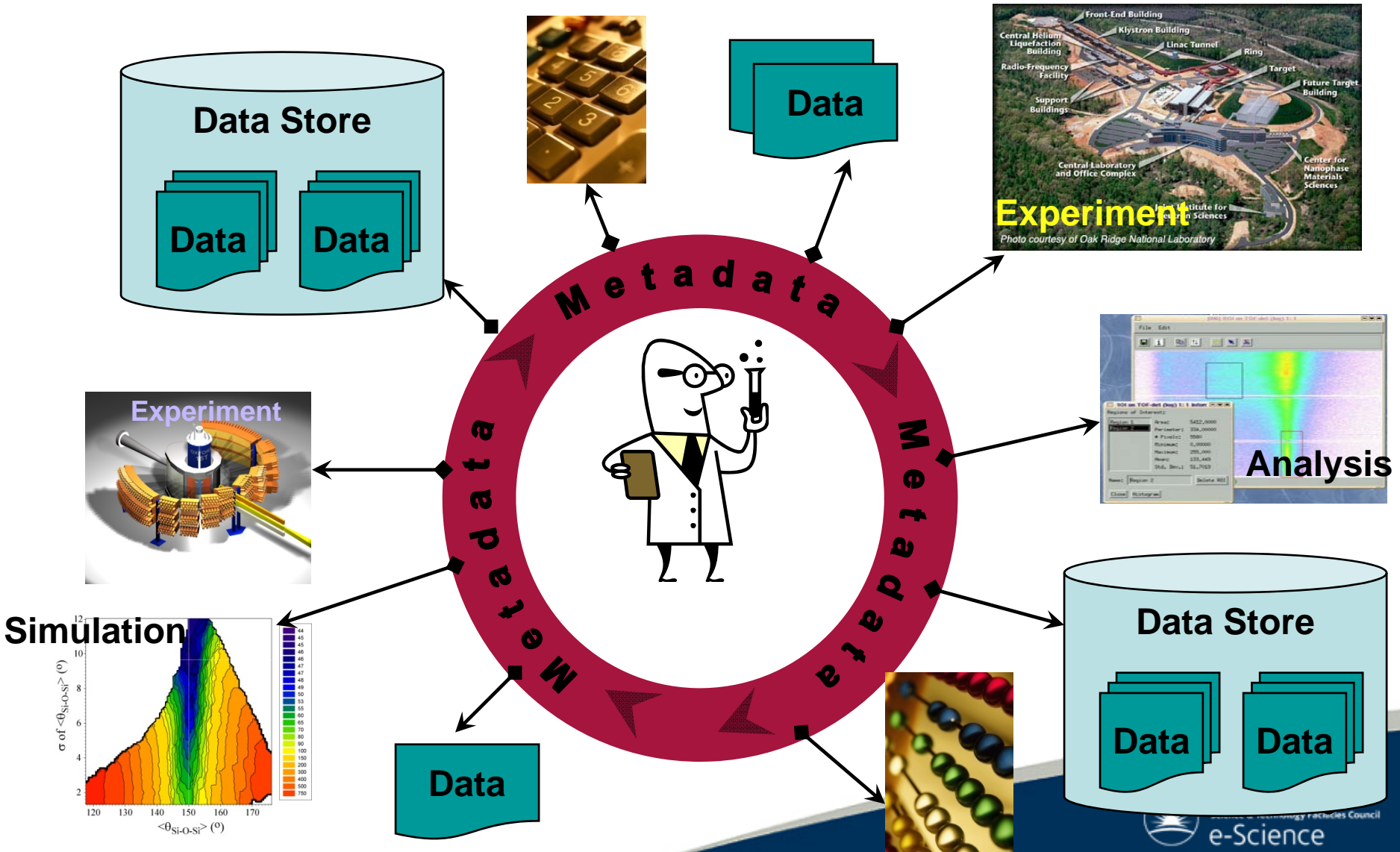
- LHC Tier 1 centre
- EGEE – ROC for UK and Ireland
- National Grid Service
- EGI
- Many Research Projects
 - Grid Portals
 - XtremOS – Grid OS
 - VOs and Security
 - ShibGrid



Complexity of large-scale research



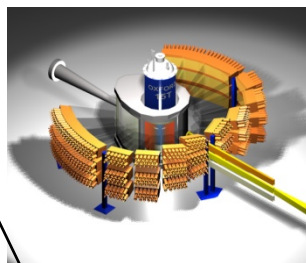
Integration via Metadata



Integrated e-Infrastructure

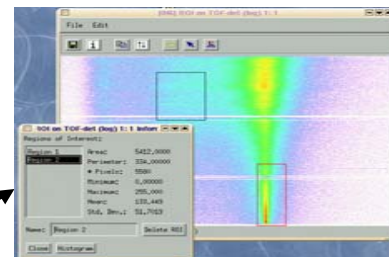


Experiment



Data Acquisition System

Data Analysis



Proposal

All Data and Metadata Capture is automated.

Publication



Information

Data Portal expanded results

Crystal Structure: Copper : Palladium: complexes
 Investigator: SRS, CLRC Daresbury Laboratory
 PI Prof E Sims
 Date: 21/04/1999

Study Reference P10002
 Investigator: Prof E Sims
 SRS, CLRC
 Daresbury
 Laboratory Principle
 Investigator
 Position: Professor

Study Information
 Funding: EPSRC, RCUK, unknown, Funding Unknown
 Body: To study the structure of Copper and Palladium co-ordination complexes at a temperature of 150K.
 Type of Study

Data Manager
 Institution: SRS, CLRC Daresbury Laboratory
 Contact: Dr Cui Tang
 University of Hull
 Instrument Used: SRS Station 9.6, BRUKER AXS SMART 9k

View All Metadata

Collection	Parent Collection	Owner	Created	Updated	Size	Bytes Transfer	File Size	File Type
CLRC	CLRC	CLRC	2001-03-28 12:42:14	2001-03-28 12:42:14	0	0	1024	application/octet-stream
CLRC	CLRC	CLRC	2001-03-28 12:42:14	2001-03-28 12:42:14	0	0	1024	application/octet-stream
CLRC	CLRC	CLRC	2001-03-28 12:42:14	2001-03-28 12:42:14	0	0	1024	application/octet-stream
CLRC	CLRC	CLRC	2001-03-28 12:42:14	2001-03-28 12:42:14	0	0	1024	application/octet-stream
CLRC	CLRC	CLRC	2001-03-28 12:42:14	2001-03-28 12:42:14	0	0	1024	application/octet-stream
CLRC	CLRC	CLRC	2001-03-28 12:42:14	2001-03-28 12:42:14	0	0	1024	application/octet-stream
CLRC	CLRC	CLRC	2001-03-28 12:42:14	2001-03-28 12:42:14	0	0	1024	application/octet-stream
CLRC	CLRC	CLRC	2001-03-28 12:42:14	2001-03-28 12:42:14	0	0	1024	application/octet-stream
CLRC	CLRC	CLRC	2001-03-28 12:42:14	2001-03-28 12:42:14	0	0	1024	application/octet-stream
CLRC	CLRC	CLRC	2001-03-28 12:42:14	2001-03-28 12:42:14	0	0	1024	application/octet-stream



Proposal System

Metadata Catalogue

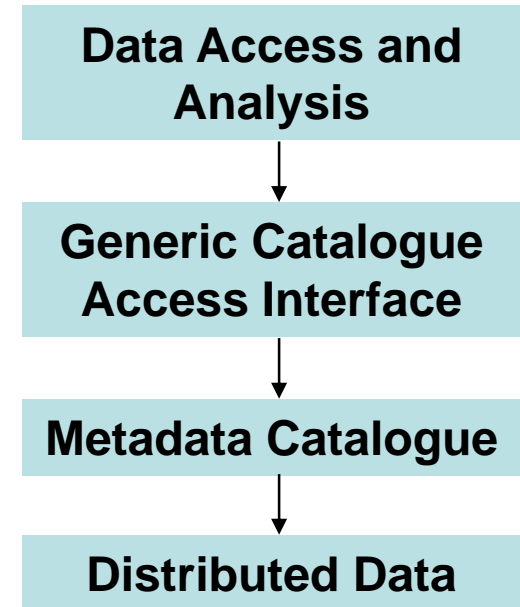
Secure Storage

E-Pubs



The ICAT software suite

- Catalogues all experiment related information
- Metadata gathered via integration with existing IT systems
 - proposal systems
 - data acquisition
- Provides a well defined API for easy embedding into any applications.



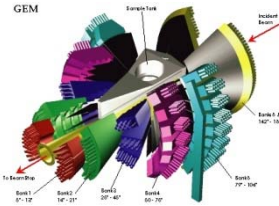
What is ICAT

What is ICAT ?

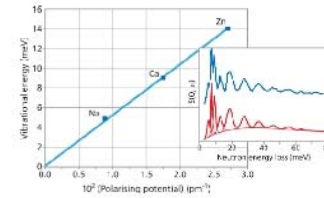
ICAT is a database (with a well defined API) that provides a uniform interface to experimental data and a mechanism to link all aspects of research from proposal through to publication.

- Access data anywhere via the web
- Annotate your data
- Search for data in a meaningful way e.g. taxonomy, Sample, temperature, pressure etc
- Share data with colleagues
- Access data via your own programs (C++, Fortran, Java etc.) via the ICAT API
- Identify potential collaborations
- Utilise integrated e-Science High-Performance Computing and Visualisation resources
- Link to data from your publications
- Etc.

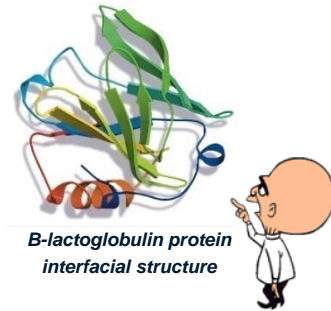
Example ISIS Proposal



GEM – High intensity, high resolution neutron diffractometer



H₂(zeolite) vibrational frequencies vs polarising potential of cations



ICAT

Proposals

Once awarded beamtime, 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

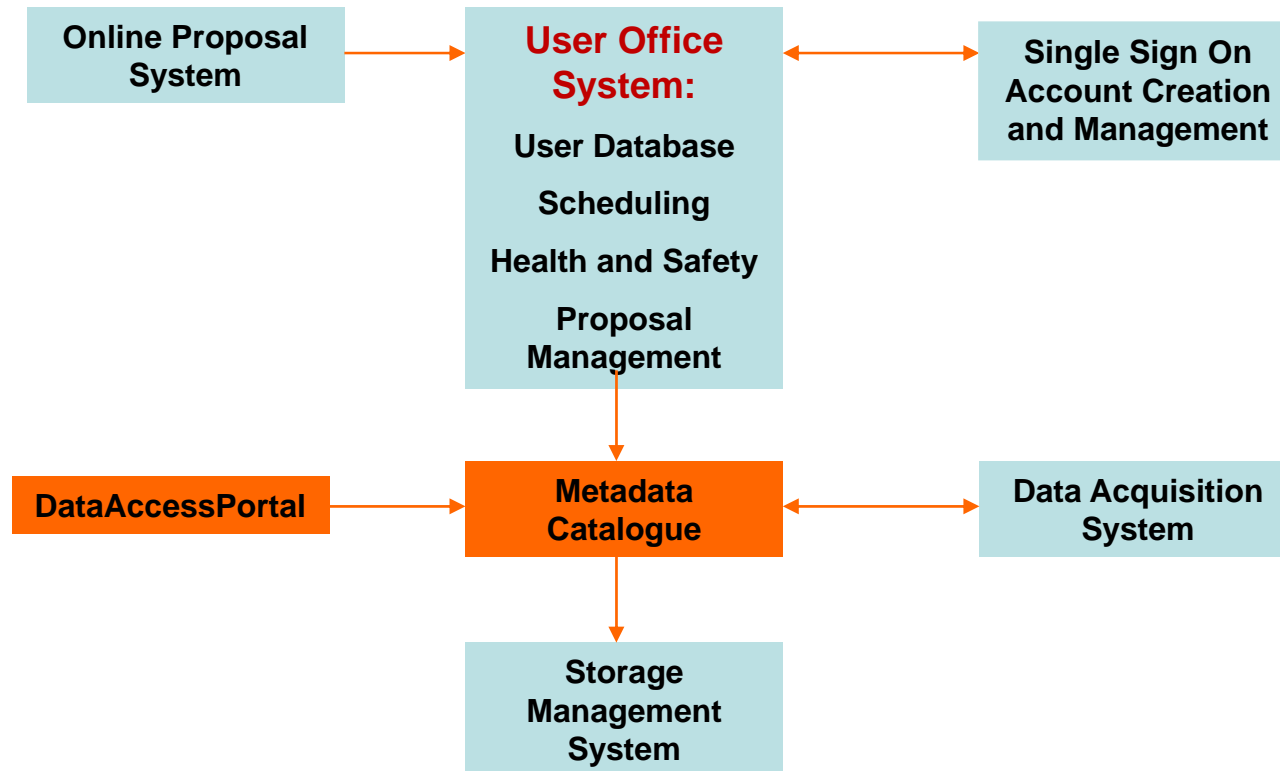
Analysed Data

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

Publication

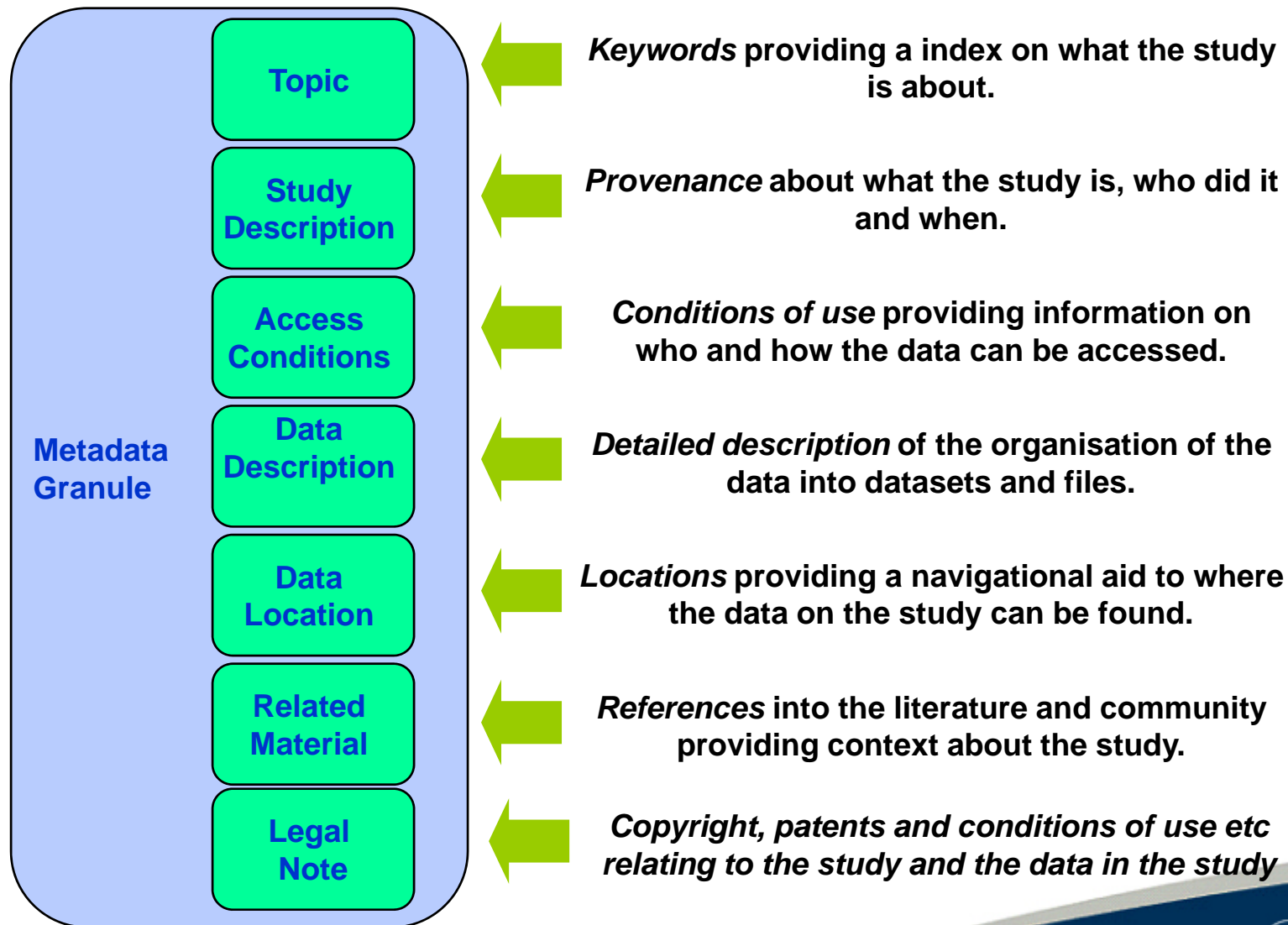
Using ICAT you will also be able to associate publications to your experiment and even reference data from your publications.

Underlying Data Infrastructure

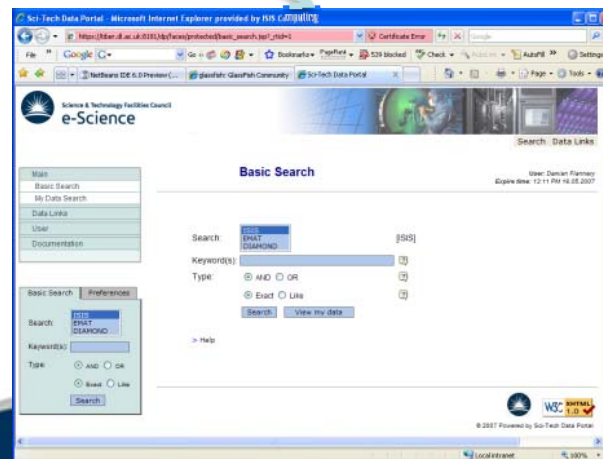
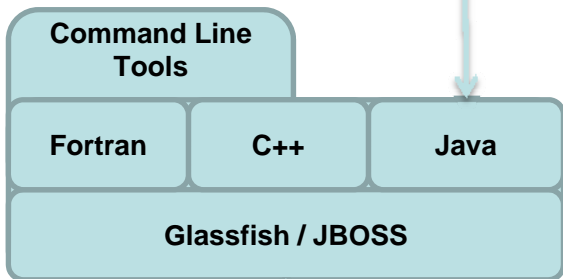
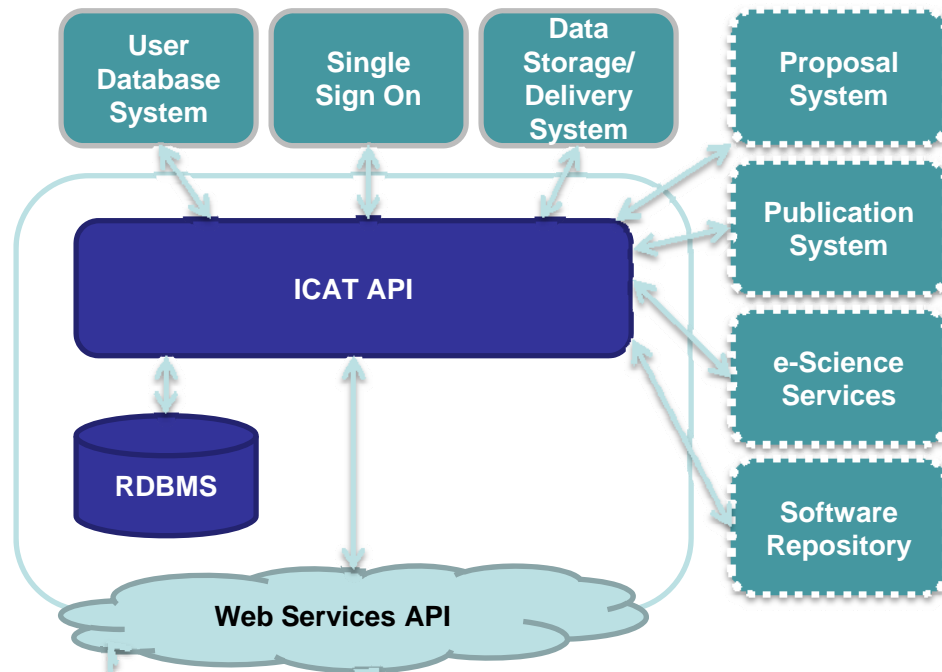


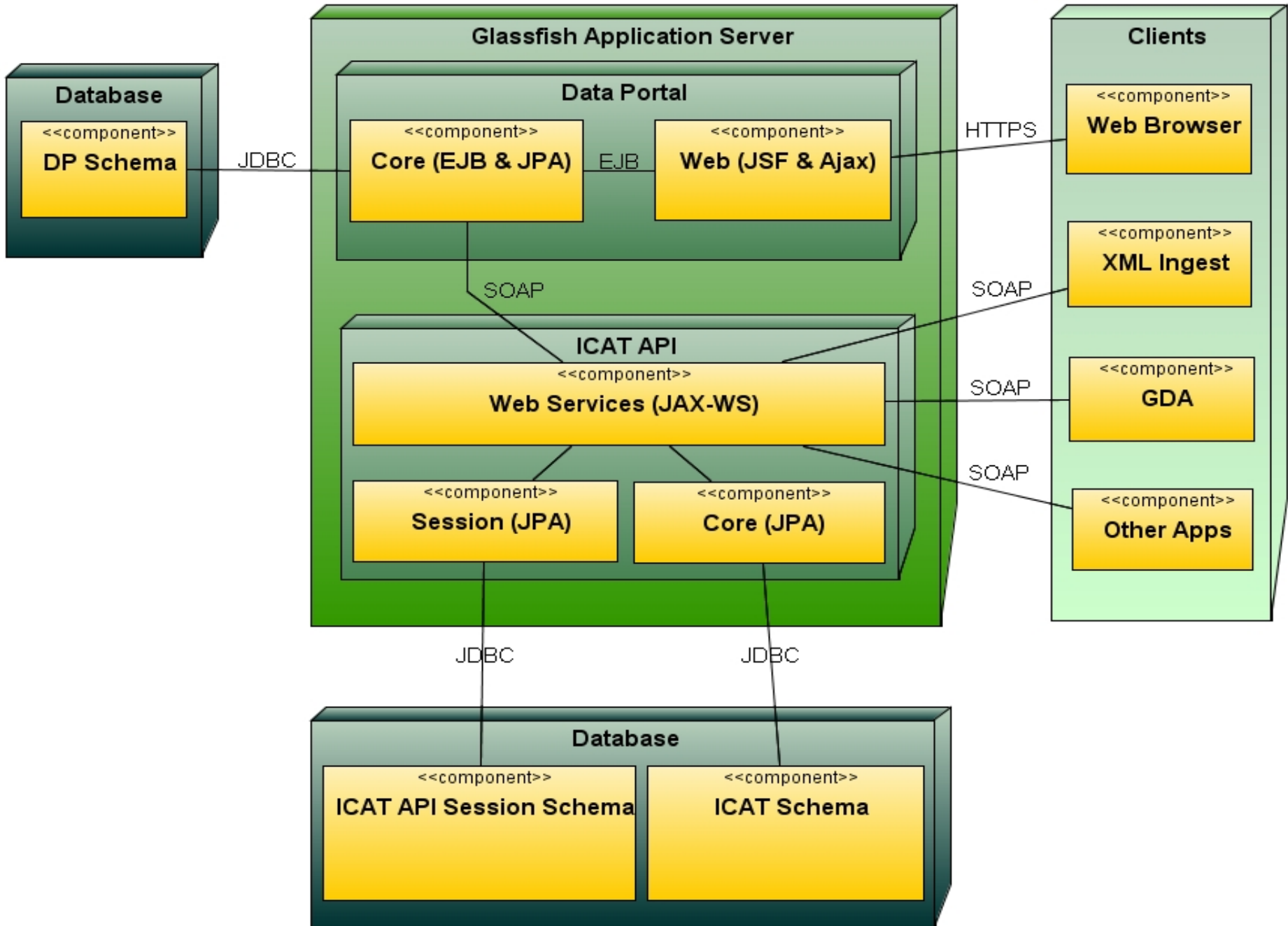
ICAT Software Suite, providing the crucial integration of key functions.

Core Scientific Metadata Model



ICAT Architecture Overview





ICAT API

Service Oriented Architecture

- Services exposed as Web Services
- User required to authenticate in order to obtain Session Token
- Token is used in all subsequent API calls to for authorisation

The API is modular in order to fit the needs of the facilities

- Plugin own user database
- Plugin data delivery system

Characteristics

- Platform independent [*Java*]
- Application Server independent [*EJB3*]
- Database Independent (Almost!) [*JPL*]
- Language independent [*Web Services*]

Internals

- Core functionality implemented as POJOs using JPA
- For deployment EJB3 Session Beans bind the core API, user db and data delivery aspects together
- Services are unit tested using JUNIT
- Services are logged at every interaction point using LOG4J

ICAT API Continued

ICATService

- login: String
- loginLifetime: String
- logout: boolean
- getKeywordsForUser: Collection
- getKeywordsForUserStartW
- getKeywordsForUserMax: C
- getKeywordsForUserType:
- getAllKeywords: Collection
- searchByAdvanced: Collect
- searchByAdvancedPaginati
- searchByKeywords: Collect
- searchByKeywordsAll: Colle
- getMyInvestigations: Collec
- getMyInvestigationsInclude
- getMyInvestigationsInclude
- searchByUserID: Collection
- searchByUserIDPagination:
- searchByUserSurname: Coll
- searchByUserSurnamePagir
- listInstruments: Collection
- listRoles: Collection
- listParameters: Collection
- listInvestigationTypes: Coll
- searchSamplesBySampleNar
- searchDatasetsBySample: C
- listDatasetTypes: Collection
- listDatasetStatus: Collection
- searchByRunNumber: Collection
- searchByRunNumberPagination: Collection
- getDatasetIncludes: Dataset
- getDatasets: Collection
- ingestMetadata: Long[]
- downloadDatafile: String

addSample

Parameters (3)

Parameter Name	Parameter Type
sessionId	java.lang.String
sample	uk.icat3.entity.Sample
investigationId	java.lang.Long

Output

Return type: uk.icat3.entity.Sample

Faults (4)

Parameter Name	Parameter Type
SessionException	uk.icat3.exceptions.SessionException
ValidationException	uk.icat3.exceptions.ValidationException
InsufficientPrivilegesException	uk.icat3.exceptions.InsufficientPrivilegesException
NoSuchObjectFoundException	uk.icat3.exceptions.NoSuchObjectFoundException

Description

Adds a sample to investigation, depending on whether the user has permission to update this Investigation object.

@param sessionId sessionId of the user.
 @param sample {@link Sample} object to be updated
 @param investigationId id of the investigation
 @throws uk.icat3.exceptions.NoSuchObjectFoundException if entity does not exist in database
 @throws uk.icat3.exceptions.InsufficientPrivilegesException if user has insufficient privileges to the object
 @throws uk.icat3.exceptions.ValidationException if the investigation object is invalid
 @throws uk.icat3.exceptions.SessionException if the session id is invalid
 @return sample

DownloadInfo

DownloadInfo

ICAT and Storage Management

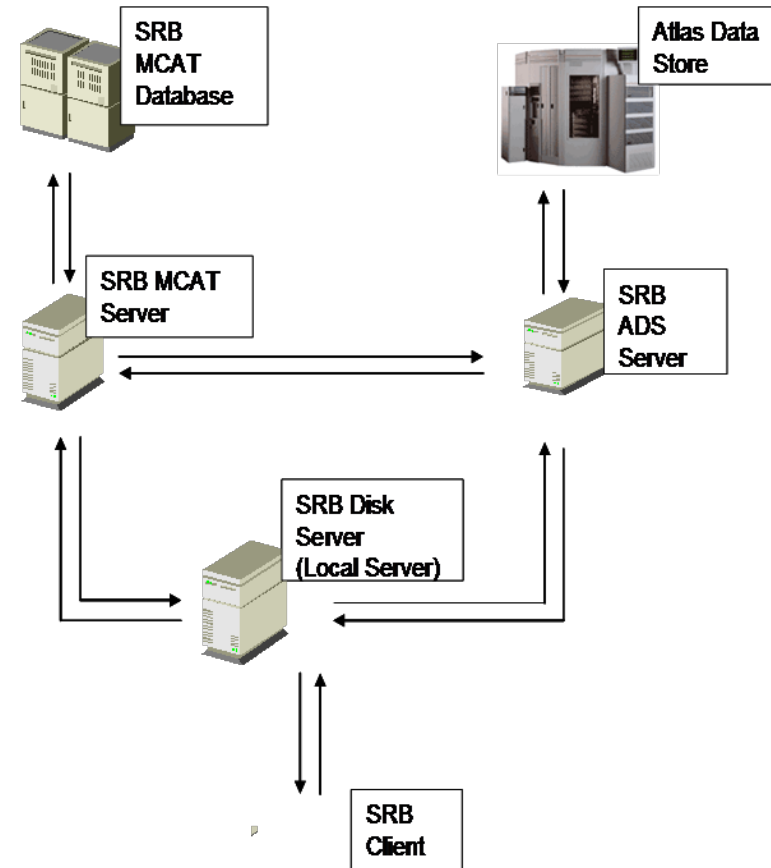
DLS uses the Storage Resource Broker for its Storage Management, this has been integrated with ICAT for data access and delivery.

Main advantage :

- Decoupling physical file location from the logical one.
- Strict Security
- Expandable to many storage systems

ISIS uses their own in house developed data storage access system called Data.ISIS.

Similar to SRB it abstracts from the physical location of the files and delivers the same advantageous in terms of decoupling of logical and physical location of files and security.



ISIS Data Portal

Advanced Search

User: kk44
Expire time: 13:18 PM 05.08.2008

Investigation Search

Keyword(s): ?
 Auto Complete ?

Investigation name: ?
Investigation abstract: ?
Sample: ?
Investigator surname: ?
Datafile name: ?
 Case Sensitive ?

Run Number (To - From): ?
Start Date: DD/MM/YYYY ... ?
End Date: DD/MM/YYYY ... ?
Rb Number: ?
Investigation type: ?
Instrument: ?

> Help

3	0	1 top	experiment	GEM	Prag - 6090	18576	2004
4	0	Bronze helmet scan 5 middle-bottom	experiment	GEM	Prag -	18577- 18579	2004

DLS Data Portal

The screenshot displays the DLS Data Portal interface in Internet Explorer. The browser address bar shows the URL: https://volga.dl.ac.uk:8181/dataportal/faces/protected/basic_search.jsp. The page title is "Data". The user is identified as "User: kk44" with an "Expire time: 13:31 PM 07.08.2008".

The main content area shows a search result for "SrF2 calibration w=-25.3". The search criteria are: Visit Id: 1, Beamline: SXD, and Datasets: Default. The status is "complete" and the type is "experiment_raw". The description states: "These files were processed retrospectively using application 'writeRaw' v1.6" on Tue Apr 25 00:56:38 2006.

Below the search result is a table titled "Default's datafiles" with the following data:

#	Name	File Size (B)	Format	Format Version	Format Type	Create Time
1	SXD01064.RAW	2430976				
2	SXD01256.RAW	6024704				
3	SXD01300.LOG	227				
4	SXD01300.RAW	3169280				
5	new name of 2fe31519-1e08-431f-b6a7-62b30e344aa1		nexus	3.0.0	HDF5	

5 Datafiles found, displaying 5, from 1 to 5. Page 1 / 1

ICAT Usage

- ISIS has 22 neutron and muon instruments which are
 - populating ICAT in real time at an average of 330 datafiles per hour.
 - 3,133,639 files (as of 9 Oct 08) that are indexed by the ISIS ICAT
 - ~4 Tb in terms of data volume..
 - 6.7GB metadata, 33M rows
 - 550+ unit & stress tests
- The new Target Station 2 at ISIS be entered into ICAT in exactly the same way as TS1.
- There are in the region of 800 experiments/investigations performed at ISIS each year.
- Rolling out ICAT 3.3 to DLS
 - DLS likely to be much larger

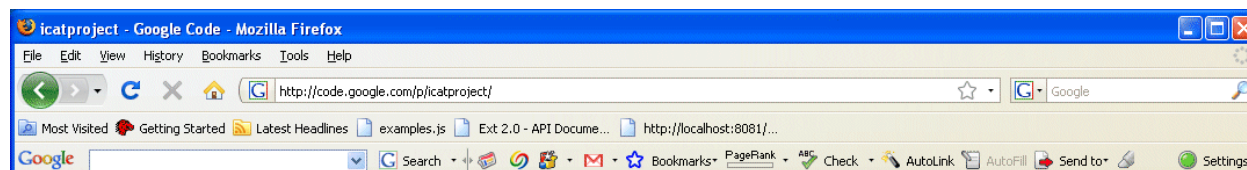


Data Policy

- Data Policy (ISIS)
 - 3 year embargo on data (+1 if requested)
 - Commercial data is never made public
 - Instrument Scientists can access all data from their beamline
 - Calibration data is public
 - Any data that involves IPR (e.g. analysed) is private for perpetuity unless explicitly shared by user
- Automatic Enforcement of policy
 - A research area

CONFIDENTIAL

ICAT for Developers



Google Groups **d.w.flannery@googlemail.com** | My Groups | Favorites | Options

icat Search Projects Search the Web

Home New since last time: [1 message](#)

Welcome to the ICAT Discussion Group!

ICAT is a database (with supporting software) that provides an interface to experimental data and will provide a mechanism to link all aspects of research from proposal through to publication. This is the official discussion forum for ICAT.

Here are some important ICAT resources:

- [ICAT Homepage](#)
- [ICAT Demo site](#)
- [ICAT Blog](#)
- [Developer resources at code.google.com](#)

Feel free to post messages on any ICAT-related topic, but please check the list of resources above first; you may find your question has already been answered.

[\[edit welcome message\]](#)

Discussions All 2 messages [view all](#) + new post

[New members will have their first post moderated](#)
By d.w.flann...@googlemail.com - Oct 8 - 1 author - 0 replies

[Future of ICAT Data Portal](#)
By Laurent Lerusse - 4:15am - 1 author - 0 replies

Members 3 members [view all](#) + invite members

 **d.w.flann...@googlemail.com (you)** Group owner

 **Laurent Lerusse** Manager

 **Tom** Member

- [New BSD License](#)
- [ICAT, Damien, Flannery, STFC, Neutrons, ISIS, Muons, Nexus, X-rays, Diamond, SNS, ANSTO, ScientificData, APIs, Java](#)
- Downloads:** [Show all](#)
[icat3.3.zip](#)
- Wiki Pages:** [Show all](#)
[Database Installation](#)
[API Installation](#)
[Dp Installation](#)
[Main](#)
- [STFC Website](#)
- [ICAT Website](#)
- [ICAT/Data Portal demo site](#)
- [ICAT Blog](#)
- [Project News](#)
- [ICAT Discussion Group](#)



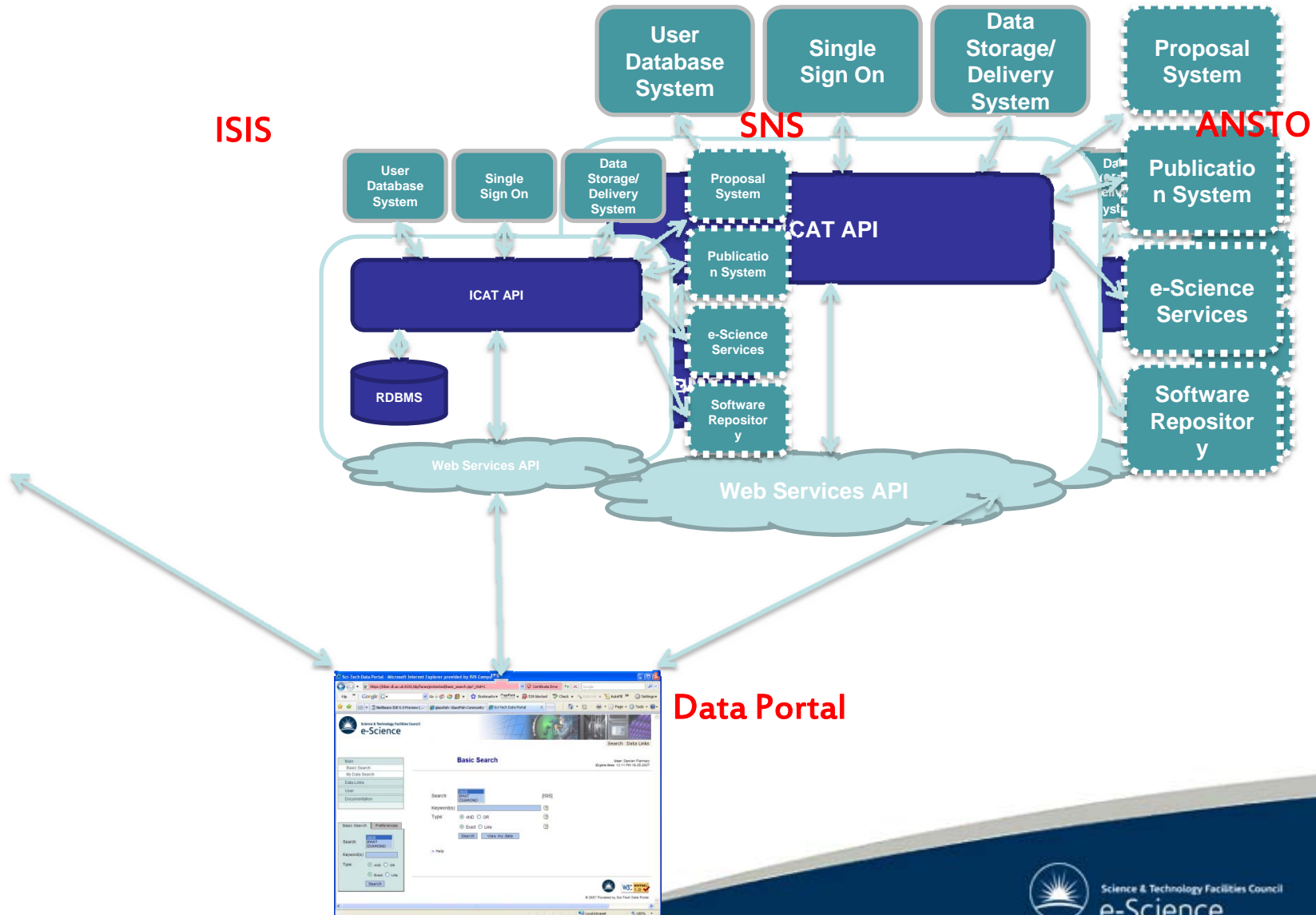
FRIDAY
UNITE
WHAT
ICAT is provided
will provide
research



WHY IS
The stor
moon di
resourc

The schema can be found in the ICAT Api source code under the [BLOG ARCHIVE](#)

Federation



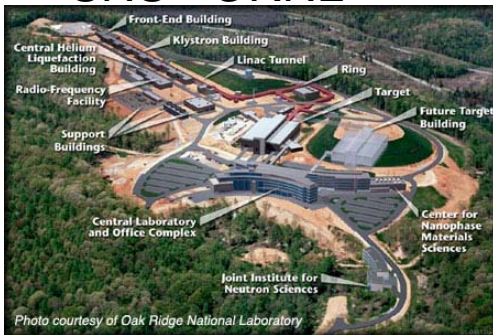
e-Infrastructure – Access to Multiple Facilities



ANSTO - Australia



SNS - ORNL



iCat



ISIS – TS1 + 2



DLS



CLF

eCrystals 'Data Federation'

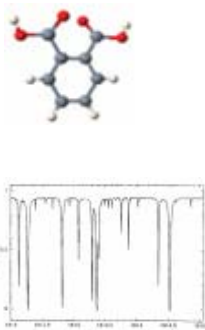


Data discovery,
linking, citation

Presentation services / portals

Data discovery,
linking, citation

Data creation
& capture in
"Smart lab"

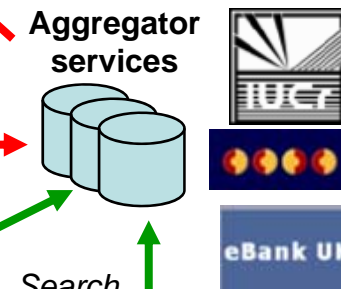
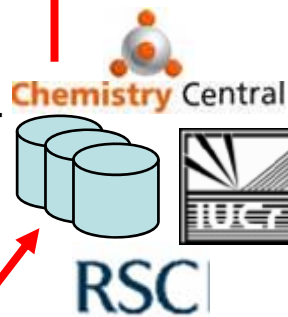


Chemical Database Service

Aggregator
services

Search,
harvest

Publishers: peer-
review journals,
conference
proceedings, etc



Search,
harvest

Publication

Deposit

Laboratory
repository



Data analysis

Institutional
data repositories

Validation

Subject
Repository

Search,
harvest

Deposit

Deposit

Deposit,
Validation

D|C|C



Crystal Structure Report Archive

Curation



Reciprocal Net



Preservation
Deposit

Institution Library &
Information Services

Summary

- Data Infrastructure at ISIS and DLS
- We believe it has some good software characteristics
 - Scalability, Maintainability, Reliability, Availability, Extensibility, Performance, Manageability, Security
- We want to take this forward
- We would like to do it in collaboration with other facilities
- Integrate with other Grid resources

Questions?

brian.matthews@stfc.ac.uk

