



## WP17

# Metadata Management and Data Continuum



# WP17 Agenda

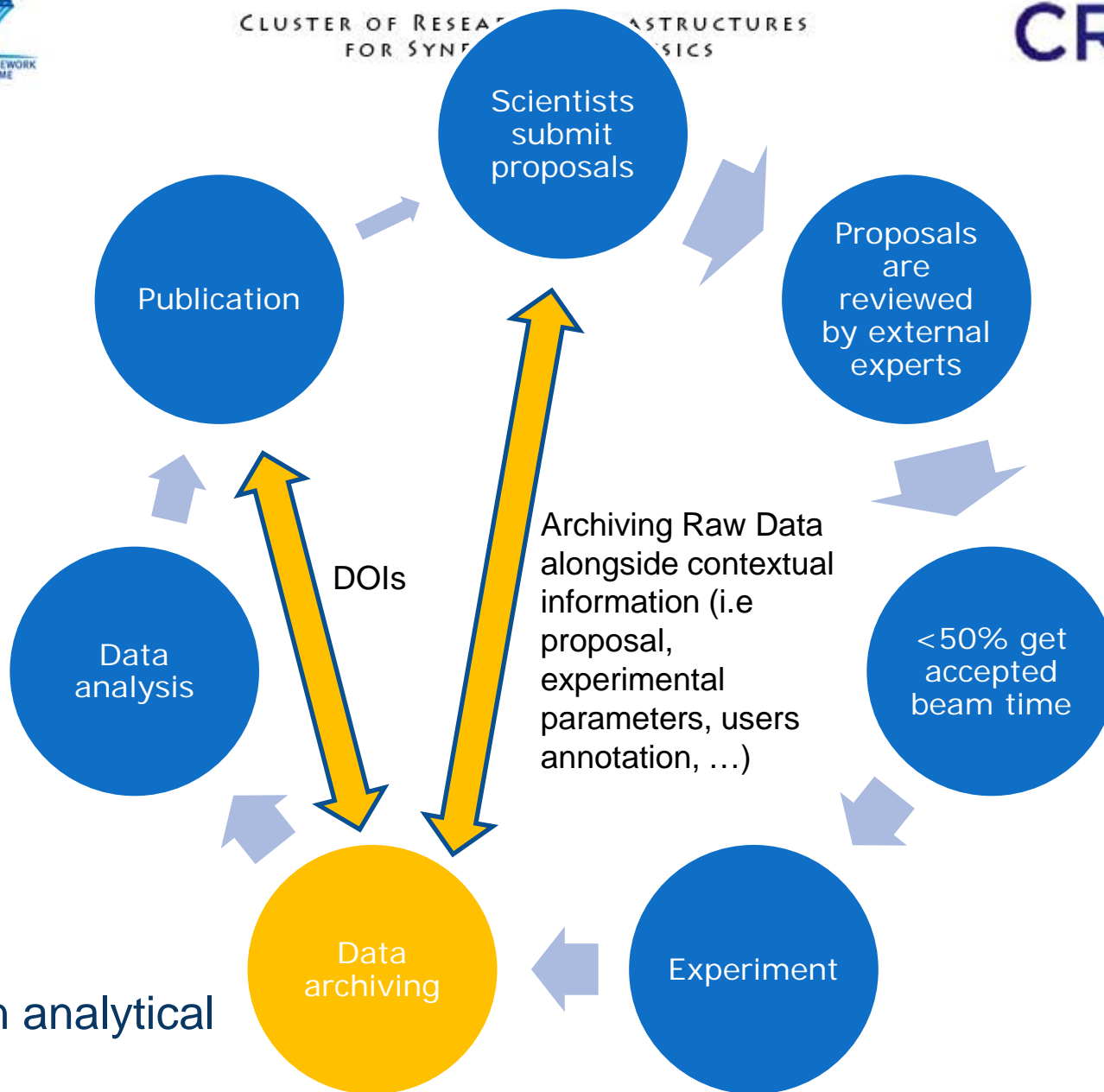
- Data continuum at ILL (Jean-François Perrin)
- DOIs and Data Mining at ESRF (Julien Savoyet)
- OAIS archiving at CERN (Jake Calum Cowton)
- Data and Software preservation - Pilot for LHC (Tibor Simko)



# Data Continuum

Aim for better sustainable links, traceability  
and transparency for the whole chain of  
experimental science.





## Scientific Workflow in analytical facilities

# Data Catalogues 1/2

**Data Toolbox** Jean-Francois Perrin (Staff) Logout

Your tools to manage your ILL Data

**Global Search**  
in5 quasielastic  
 Include experiments without data  
Search

**Advanced Search**  
My role in the experiment (?)  
Proposal number (?)  
Proposal title  
Abstract  
Cycle  
Instrument  
Proposer (?)  
Attached documents (?)  
Data files metadata (?)  
 Include experiments without data  
Search

**Proposals list**  
Proposal 8-04-684 Popout  
Proposal Members Data ranges Data folders  
131/in5/exp\_8-04-684  
histo  
logfiles  
rawdata

**Data**  
131/in5/exp\_8-04-684 / rawdata

File Name	Size
105636.nxs	3785 KB
105637.nxs	3914 KB
105638.nxs	3016 KB
105639.nxs	2991 KB
105640.nxs	2676
105641.nxs	2999
105642.nxs	2993
105643.nxs	7798
105644.nxs	5805
105645.nxs	12 MB
105646.nxs	12 MB
105647.nxs	12 MB
105648.nxs	13 MB

View files (one by one)  
Download files (one by one)  
Download a zip with all files  
Select all files

- Data toolbox : local (ILL only) catalogue
- Tailored to ILL needs
  - User management of data access authorization.
  - Users could decide to publish (open access) their data, before the end of the embargo period.
  - Dissemination of access rights.
  - Linked to DOIs.
  - Linked to experimental logs.
  - Linked to user annotation tool.
  - Linked with proposal system.
  - Download of data.
  - Full text search



# Data Catalogues 2/2

- ICAT has been chosen by the Photon and Neutron community.
- ICAT is a federated catalogue.
  - From a single interface/portal users can search and access data located at different facilities
  - Stores information at the data file level

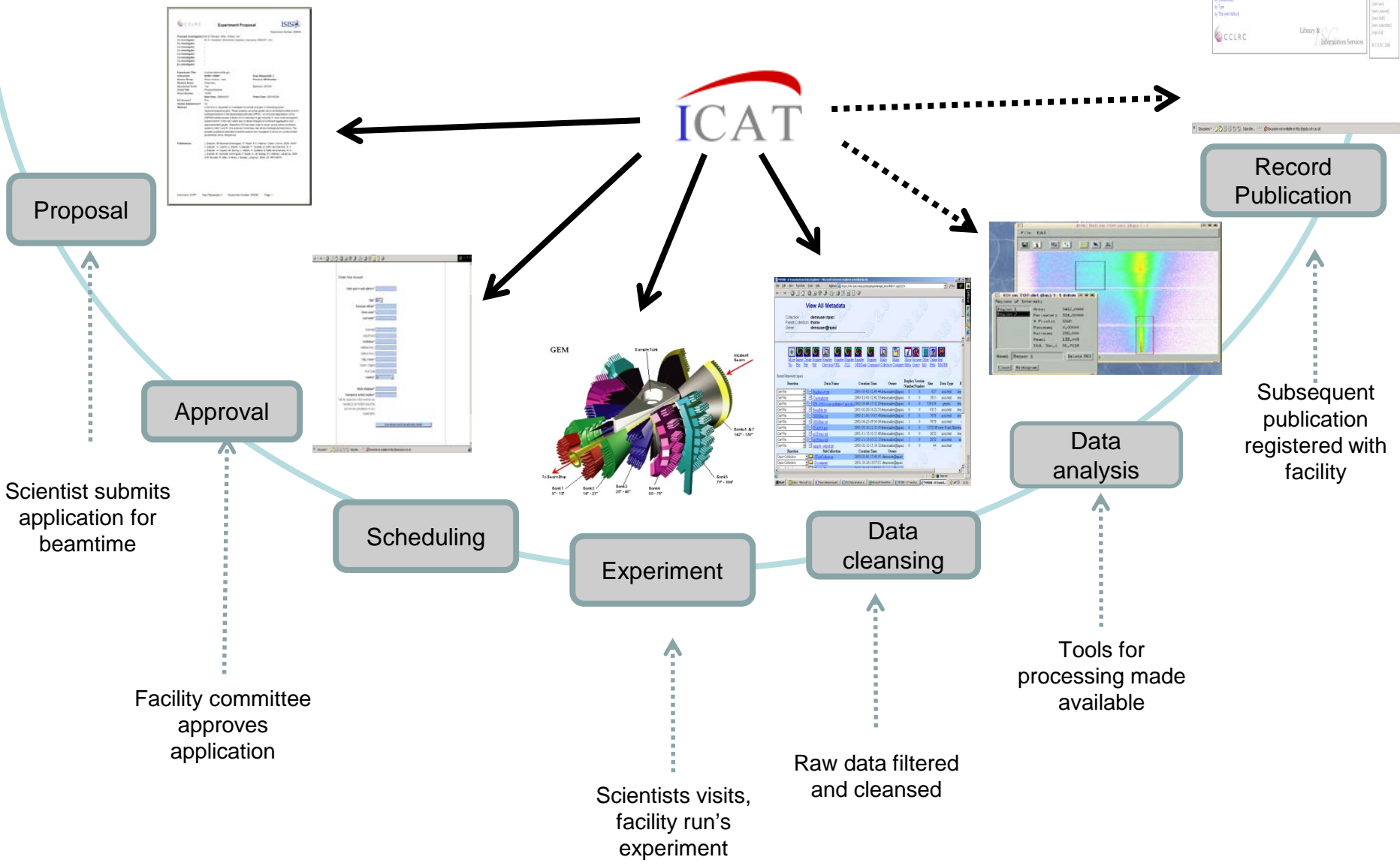


CLUSTER OF RESEARCH INFRASTRUCTURES  
FOR SYNERGIES IN PHYSICS

CR



ICAT





# ILL ICAT Data Ingestion (EDIS)

- Development of ingestion restful web services
  - Flexibility - could be call from any system
    - Script checking archive file systems
    - Experimental (Instrument control) log system
  - Simplicity – Hide the complexity of authorization mechanism by grabbing information from the proposal system



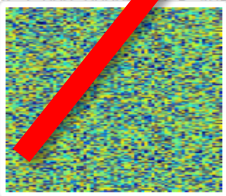


Logs

Family: All families Date: Now Period

Cycle 2014-1 Instrument D17

2014-04-30 13:55:44	6	5.000	000050	0	3.164e+07	1.055e+07	3.097e+0
2014-04-30 13:55:44	7	6.000	000051	0	3.330e+07	1.110e+07	3.260e+0
2014-04-30 13:55:45	8	7.000	000052	0	3.497e+07	1.166e+07	3.423e+0
2014-04-30 13:55:45	9	8.000	000053	0	3.663e+07	1.221e+07	3.586e+0
2014-04-30 13:55:45	10	9.000	000054	0	3.830e+07	1.277e+07	3.749e+0
2014-04-30 13:55:45	11	10.000	000055	0	3.996e+07	1.332e+07	3.912e+0
2014-04-30 13:55:45							Optimum not found
2014-04-30 13:57:29							For { $\$i = 0$ to 2 step 1}
2014-04-30 13:57:29							$\$i = 0$
2014-04-30 13:57:29							Numer: 000056 Time: 0 s Detector: 6.163e+07 (
2014-04-30 13:57:30							Step san Numer Time Detector Rate Det Monitor1
2014-04-30 13:57:30	1	0.000	000057	0	4.329e+07	1.135e+07	4.238e+0
2014-04-30 13:57:30	2	1.000	000058	0	4.496e+07	1.199e+07	4.401e+0
2014-04-30 13:57:30	3	2.000	000059	0	4.662e+07	1.263e+07	4.564e+0
2014-04-30 13:57:30	4	3.000	000060	0	4.829e+07	1.327e+07	4.727e+0
2014-04-30 13:57:30	5	4.000	000061	0	4.996e+07	1.391e+07	4.890e+0
2014-04-30 13:57:31	6	5.000	000062	0	5.162e+07	1.455e+07	5.053e+0
2014-04-30 13:57:31	7	6.000	000063	0	5.329e+07	1.519e+07	5.216e+0
2014-04-30 13:57:31	8	7.000	000064	0	5.496e+07	1.583e+07	5.379e+0
2014-04-30 13:57:31	9	8.000	000065	0	5.662e+07	1.647e+07	5.542e+0
2014-04-30 13:57:31	10	9.000	000066	0	5.829e+07	1.711e+07	5.705e+0
2014-04-30 13:57:31	11	10.000	000067	0	5.996e+07	1.775e+07	5.868e+0
2014-04-30 13:57:31							Detector: 6.161e+07 (
2014-04-30 13:57:32							Step san Numer Time Detector Rate Det Monitor1
2014-04-30 13:57:32							TestTestTestTestTestTestTestTestTestTestTe
2014-04-30 13:57:32	1	0.000	000069	0	6.327e+07	2.109e+07	6.194e+0



NEUTRONS FOR JOINTS Institut Laue-Langevin

My Data My Downloads Search Browse All Data

Download

- LL-DEV
  - D10
  - D20
  - D3
  - INS
    - Investigation for proposal 4-01 1230 on instrument INS (Inv. Id:4-01 1230 - INS & Visit Id:62974)
    - Investigation for proposal 4-01 1324 on instrument INS (Inv. Id:4-01 1324 - INS & Visit Id:62953)
      - 110322.nxs [1.2 MB]
      - 110323.nxs [1.2 MB]
      - 110324.nxs [1.2 MB]
      - 110325.nxs [1.2 MB]
      - 110326.nxs [1.2 MB]

Datafile Window

File Name	File Location	File Size	Format	Format Version	Format Type	Create Time
Dataset Name: test: 4-01 1324 - INS - Ba3NbFe3Si2O14 (20 Items)						
110384.nxs	133/in5/exp_4...	1.2 MB	nx	N/A	application/x-hdf	2013-07-25 17...
110322.nxs	133/in5/exp_4...	1.2 MB	nx	N/A	application/x-hdf	2013-07-25 14...
110385.nxs	133/in5/exp_4...	1.2 MB	nx	N/A	application/x-hdf	2013-07-25 17...
110323.nxs	133/in5/exp_4...	1.2 MB	nx	N/A	application/x-hdf	2013-07-25 14...
110387.nxs	133/in5/exp_4...	1.2 MB	nx	N/A	application/x-hdf	2013-07-25 17...
110324.nxs	133/in5/exp_4...	1.2 MB	nx	N/A	application/x-hdf	2013-07-25 14...
110386.nxs	133/in5/exp_4...	1.2 MB	nx	N/A	application/x-hdf	2013-07-25 17...
110325.nxs	133/in5/exp_4...	1.2 MB	nx	N/A	application/x-hdf	2013-07-25 14...
110388.nxs	133/in5/exp_4...	1.2 MB	nx	N/A	application/x-hdf	2013-07-25 17...
110326.nxs	133/in5/exp_4...	1.2 MB	nx	N/A	application/x-hdf	2013-07-25 14...
110389.nxs	133/in5/exp_4...	1.2 MB	nx	N/A	application/x-hdf	2013-07-25 17...
110327.nxs	133/in5/exp_4...	1.2 MB	nx	N/A	application/x-hdf	2013-07-25 14...

Page 1 of 26

Investigation Data

Page 1 of 1

Title: Investigation for proposal 4-01 1324 on instrument INS  
 Investigation No: 4-01 1324 - INS  
 Visit Id: 62953  
 Start Date: 2013-07-25 00:00  
 End Date: 2013-08-01 00:00



# Search engine

- Index available information (Proposal, experimental report, data file annotation, ...) – currently only publications are still missing.
- Implement user access authorization at the index level.
- Provide web services for querying/fetching results.
- Based on SolR/Lucene.
- Integrated in the different tools.

The screenshot displays the CRISP search engine interface. On the left, there is a 'Global Search' sidebar with a search input field containing 'in5 quasielastic', an unchecked checkbox for 'Include experiments without data', and a 'Search' button. Below it is an 'Advanced Search' section with dropdown menus for 'My role in the experiment (?)' and 'Proposal number (?)'. The main content area, titled 'Proposals list', shows a search result for '18 proposal(s) found with the given search filter'. The first result is '8-04-684 : The influence of cofactor binding on thermal fluctuations of alcohol dehydrogenase studied by using quasielastic neutron scattering'. The title and abstract are highlighted in red, and the words 'quasielastic' and 'IN5' are also highlighted in red. The abstract text reads: 'Title: quasielastic neutron scattering. Abstract: related to the collective domain motions. In this proposal we propose to investigate using the IN5. Data title: quasielastic neutron scattering. Proposal file: visible. Figure 3: Experimental IN5 data: The ADH spectra (red symbols) show two narrow quasielastic'. Below this is another result: '6-02-512 : Dynamics of water inside hydrophobic nanotube'.



# DOIs



DOI > 10.5291/ILL-

6-02-512 : Dynamics of water inside hydrophobic nanotube

- **Proposal file** : time for cryostat cooling. The experiments on the TOF (IN5/IN6) will be performed at four
- **Report file** : Area: This proposal is continuation of: 6-02-486 IN5 0 Instrument Req. Days 1 All. Days 19/11/2012

5-12-246 : Spiral spin liquid in highly frustrated diamond lattice antiferromagnet

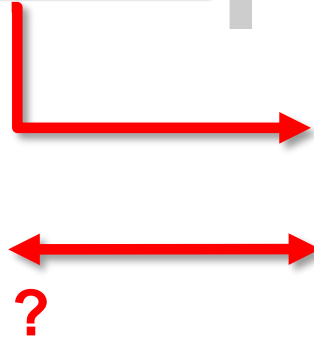
**Proposal 6-02-512** Popout

Proposal Members Data ranges Data folders

**Title**  
Dynamics of water inside hydrophobic nanotube

**Abstract**  
Liquids confined in nanotube environment possess anomalous static and dynamic properties which are of

**DOI (Digital Object Identifier)**  
10.5291/ILL-DATA.6-02-512



**Please note**  
The full details of this dataset is not yet available to the public as it still under its embargo period. As such there are only a few details publicly exposed. To find out more about how the ILL governs the release of data, please go [here](#). If you would like to be kept informed when it is due to be released into the public domain, please fill in the form on the right hand side of this message. Thank you for your understanding.

**Metadata**

DOI  
doi:10.5291/ILL-DATA.6-02-512

**Authors**  
BRIGANTI GIUSEPPE, DE LUCA Francesco, MACCARINI MARCO, PARMENTIER Alexandra, SEYDEL Tilo

**Publisher**  
Institut Laue-Langevin

**Publication year**  
2013

**Cycle(s)**  
2012c

**Proposal number**  
6-02-512

**Date of experiment**  
20-11-2012

**Title**  
Dynamics of water inside hydrophobic nanotube

**Abstract**  
This data is not yet public

**Download**  
This data is not currently available to download

**Data citation**  
The recommended format for citing this dataset in a research publication is in the following format: [author], [date], [title], [publisher], [doi]

**Instrument**  
This data is not yet public

**Experiment parameters**  
This data is not yet public

**Sample parameters**  
This data is not yet public

*(Faint text from a document, likely the abstract or introduction of the paper, partially obscured by the interface elements.)*

**Institut Laue-Langevin**  
Access to data is governed by the **ILL data policy**

Copyright © 2013

ILL has partnered with **DataCite** as the registration agency for data persistent identifiers.



# Data Citation Index: Rewarding

- OK, large effort have been done for opening, preserving, linking and giving access to data.
- What about rewarding for data producers (scientists and facilities)
  - Need to get access (harvest) to publication citing data DOIs.
  - Need for measures of data citation, use and reuse
  - Need for new metrics based on the use an reuse of experimental data (i.e. H-index for data)

# Data Citation Index



- Autumn 2014: agreement between DataCite and Thomson Reuters.
- Through DataCite our repositories will be integrated in the Data Citation Index.



# Where do we stand ?

ILL Reactor restart scheduled for July 2014

- ✓ ICAT ingestion ready
  - ✓ Data toolbox ready with search engine
  - ✓ DOI minting as soon as experiment done
  - ✓ Global search engine integrated
- ILL (CERN, ESRF, ...) data will be available inside the data citation index (Autumn 2014)