

## What metadata are needed to make Institut Laue-Langevin neutron diffraction raw data intelligible for new users?

#### Matthew Blakeley

LADI-III Instrument scientist, Large-Scale Structures group, Institut Laue-Langevin





## Institut Laue-Langevin

## International research centre for neutron science

- ILL high-flux reactor is one of the most intense neutron sources in the world
- Funded by France, Germany and UK (with contributions from other member countries)
- Since 2010, ILL is part of the 'European Photon and Neutron Science Campus' with ESRF, IBS and the EMBL Grenoble outstation (plus PSB, PSCM, UVHCI)







## Institut Laue-Langevin

## Around 50 instruments in use (~40 ILL and 10 CRG)

- Diffraction; powder, singlecrystal, small-angle diffractometers and reflectometers
- Inelastic neutron scattering; triple-axis, time-of-flight and high-resolution spectrometers
- Particle & nuclear physics instruments



#### All raw data has been stored at the ILL since 1973



## ILL Scientific Data Policy (10/2012)

Defines the ownership, curation and access to raw data and metadata collected and stored at the ILL (www.ill.eu/users/ill-data-policy/)

- Raw data and associated metadata obtained from publically funded research are open access after an embargo period of 3 years, in which access is restricted to the experimental team
- ILL will provide means for the capture of metadata items that are not automatically captured by instrument software
- Each dataset will have a DOI and be available via searchable on-line catalogues (*e.g.* data.ill.eu, logs.ill.eu)
- **\*Remark:** ILL is not responsible to fully curate the data e.g. to ensure that software to manipulate the data is available



### ILL Web-based tools

## Necessary to have an 'ILL User Club' account (userclub.ill.eu) to access searchable on-line catalogues

 Click on 'New User' and fill in the necessary details (*e.g.* name, date of birth, email, employer name and address) to create an account

\varTheta 🔿 🔿 The ILL User Club 🗙 🕂		
♦ A https://userclub.ill.eu/cvng/	[] ⊽ C	
業 🔳 🔳 S	ign in	
I ne ILL User Club		
	New user	New user
New User Experimental report	User details Employer details	User details Employer details
	Sex * 🔘 Male 🔘 Female	Enter the employer's name or city ** Reset
	First name *	
	Last name *	Please complete the form below if you cannot find the employer above ** Sector *  O University/School  Caboratory  Company
	User name *	Employer Name *
	Password *	Street *
	Confirm password *	
	Email #	City *
		Postcode
	Birth date * (QQ/MM/yyyy)	Country *
	Telephone	Phone number
	*: required	*: required **: do not use accents





# Left panel provides various choices to search for datasets

- Global search: general search of all proposal information
- Advanced Search: search by specific criteria (*e.g.* cycle number [XXX = 2 digits for year plus cycle in that year e.g.  $152 - 2^{nd}$  cycle of 2015], instrument, proposer, proposal number, DOI etc.,)
- \*Remark: ILL data policy began with the 3<sup>rd</sup> cycle of 2012 (i.e. 123), therefore all data are still under embargo, except one (http://doi.ill.fr/10.5291/ILL-DATA.5-12-250)



#### Top-right panel displays the results of the search

- Proposal number (college-keyword-number)
- Proposal title
- Cycle number
- Instrument

#### College 8 : Structure and Dynamics of Biological Systems

8-01 Macromolecular crystallography
8-02 Biological membranes
8-03 Solution structure of biological macromolecules
8-04 Dynamics of biological macromolecules
8-05 Methodological studies in biology and biophysics

Proposals list	
5 proposal(s) found with the given criterias (sorted by proposal number)	<u> </u>
8-01-416 : Understanding carbohydrate recognition in galectins using neutron and ultra-high resolution X-ray crystallography • Cycle/instrument:152 / LADI	3/
8-01-418 : An understanding of proton couple-electron transfer (PCET)using highly diffracting copper nitrite reductases	3
8-01-427 : Neutron Diffraction Studies on Human Carbonic Anhydrase II in Complex with Methazolamide and/or 4-Methyl Imidazole. • Cycle/instrument:152 / LADI	3
8-01-428 : Understanding key features in transthyretin stability and their implications on therapeutic strategies • Cycle/instrument:152 / LADI	3
8-01-431 : Enzymatic characterization of trypsin inhibition mechanism through cryogenic neutron crystallography • Cycle/instrument:152 / LADI	<i>⋧</i> ∕ ▼



Once a proposal is selected, the middle-right panel displays details of that proposal

• **'Proposal'** gives Title, Abstract\*, Planning and Sample formula\*

Proposal 8-01-427					
Proposal	😫 Links	🙀 Members	Pata ranges	E Data folders	
🔮 You are a	2 You are a manager of this proposal.				
Title Neutron Diffraction Studies on Human Carbonic Anhydrase II in Complex with Methazolamide and/or 4-Methyl Imidazole.					
Abstract					
Planning ▶ Cycle 152 (from 17/06/2015 to 06/08/2015), instrument LADI					
Sample(s)					

\*Not shown due to embargo



• **`Links'** provides a link to the User Club proposal and the DOI (http://doi.ill.fr/10.5291/ ILL-DATA.8-01-427) for the dataset

Proposal 8-01-427				
🕕 Proposal 🛞 Links 🙀 Members 😭 Data ranges 📄 Data folders				
Links to this proposal in others applications.				
User Club proposal information				
Digital Object Identifier: 10.5291/ILL-DATA.8-01-427				

If all metadata have not been extracted automatically *i.e.* where non-standard software and file formats are used, metadata can be retrieved from the proposal and DOI

 Proposal & DOI provide sample description (*e.g.* formula, unit-cell, space group) and experimental parameters (*e.g.* temperature, wavelength range)



### **`Members'** provides details of the experimental team and PI

Proposal 8-01-427				
🕕 Proposal 😣 Links 🙀 Members 😭 Data ranges 📄 Data folders				
Scheduled public availability: Not set Make the proposal public now				
Proposal team				
AGGARWAL Mayank	ORNL, OAK RIDGE	Proposer, Experimenter		
BLAKELEY Matthew	ILL, GRENOBLE	Proposer, Local contact, Experimenter		
FISHER Zoe	EUROPEAN SPALLATION SOURCE, LUND	Main proposer, Labo responsible		
Other authorized people				
There isn't any authorized people who isn't in the proposal team				
S Add another authorized person				



 'Data ranges' should\* provide data and metadata given by the ILL standard software 'NOMAD'

Proposal 8-01-4	Proposal 8-01-427			
Proposal	😫 Links	🙀 Members 📑 Data ranges 📄 Data folders		
There is no indexed data for this proposal.				
		nowever, it may have lies in the Data folders view.		

• **\*Remark:** Data.ill.eu does not currently provide data ranges for instruments which do not use NOMAD. LADI-III will implement NOMAD in the near future



'Data folders' lists various folders which include the raw data (NeXus, 16 bit TIFF, ASCII), log files and processed data/log files

Proposal 8-01-427				
Proposal	😫 Links	🙀 Members	🚰 Data ranges	E Data folders
🔲 152/ladi/ex	p_8-01-427			
🚞 histo				
logfiles				
process	ed			
🚞 rawdata	1			

• **`Rawdata'** folder for LADI-III contains 16-bit TIFF files

Data		
<u>15</u>	52/ladi/exp_8-01-427 / rawdata	
	hCA_BZM_ap2-2_5h25_phi_00.tif18 MB	
9	hCA_BZM_ap2-2_8h_phi_18.tif18 MB	
9	hCA_BZM_ap2-2_8h_phi_28.tif18 MB	
9	hCA_BZM_ap2-2_8h_phi_48.tif18 MB	
_	hCA_BZM_ap2-2_8h_phi_58.tif18 MB	



• **'Processed'** folder for LADI-III contains 3 folders



- `Neutron\_data\_reduction' contains input parameter files (LDM) for LAUEGEN and output MTZ files, input file for LSCALE and wavelength-normalized output MTZ, input and output MTZ files for SCALA/TRUNCATE
- `X-ray\_data\_reduction' contains input parameter files and the raw data for XDS and the output MTZ and log files
- 'Structural\_refinement' contains the starting files (neutron and X-ray MTZ plus PDB) for PHENIX and the output PHENIX refinement directories (PDB and MTZ)



## Metadata required for LADI-III raw data

- Crystallographic parameters
  - Unit-cell, space group, orientation matrix
- Experimental parameters
  - Wavelength range *i.e.* λ-min and λ-max, spindle angle/step, exposure time, temperature
- Detector parameters
  - Detector geometry and size, pixel size







### General issues to be resolved

- Non-standard software and file formats on different instruments
  - NOMAD is not implemented on all instruments, NeXus is not yet a standard file format at ILL
- Curation of data
  - ILL should *ideally* provide remote access to all software required for data analysis



 ILL should however be commended for implementation of the data policy and many of the current issues are being addressed by SI and DPT