

# Gestion des données scientifiques à l'Institut Laue Langevin.

IMAG – Journée archivage

19th of Nov 2019

Jean-François Perrin (ILL)

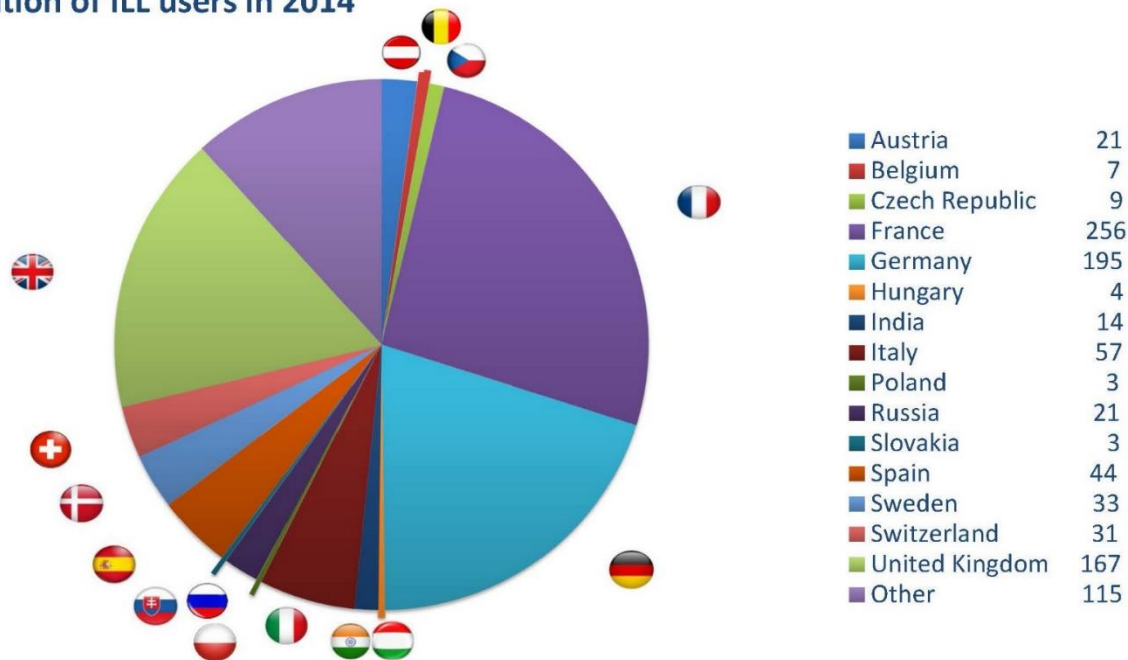
# Short introduction to the ILL

# Users

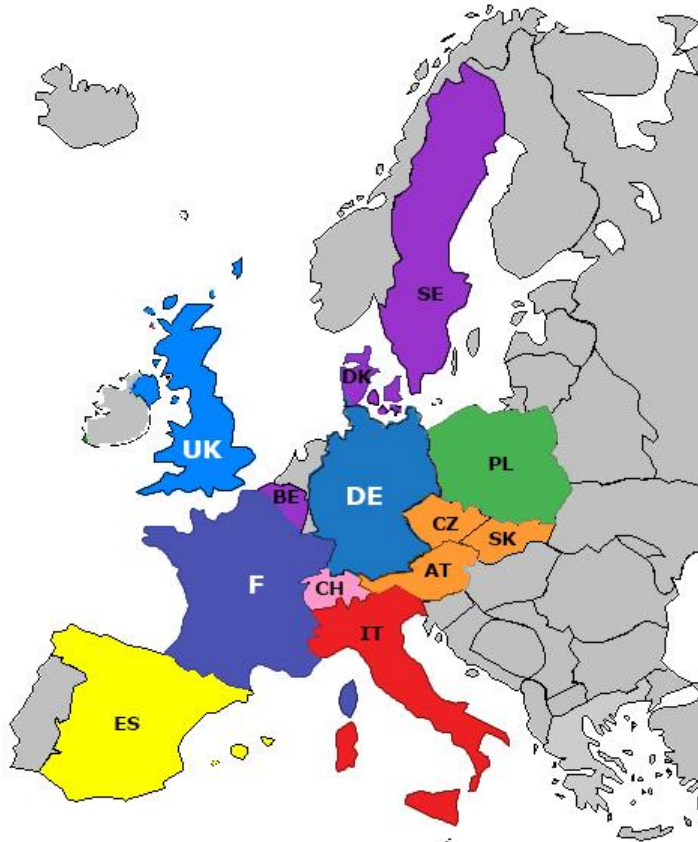
An active community of 12 000 scientists from 28 countries

1400 invited experimenters / year

National affiliation of ILL users in 2014



# ILL member countries and their financial participation



Germany : 25 %

UK : 25 %

France : 25 %

Spain

Italy

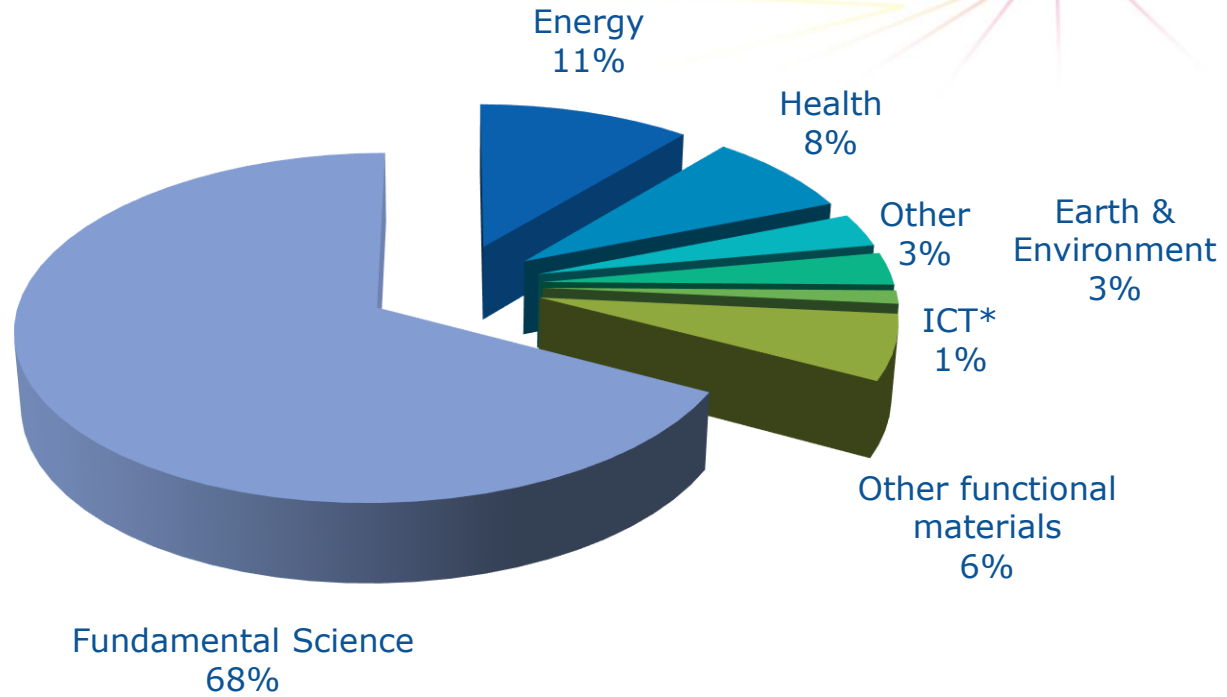
Switzerland

Poland

CENI (Central European Neutron Initiative,  
Austria, Czech Republic, Slovakia)

TRANSNI (TRANSnational Neutron Initiative,  
Belgium, Denmark, Sweden)

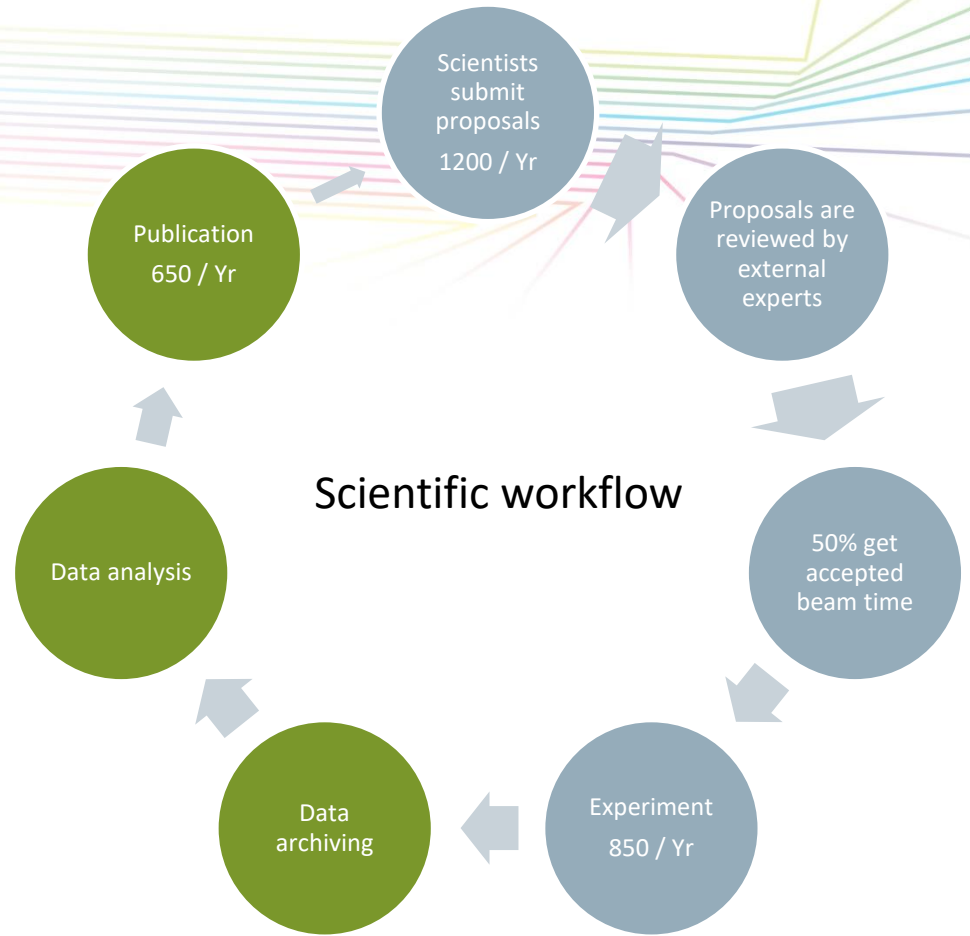
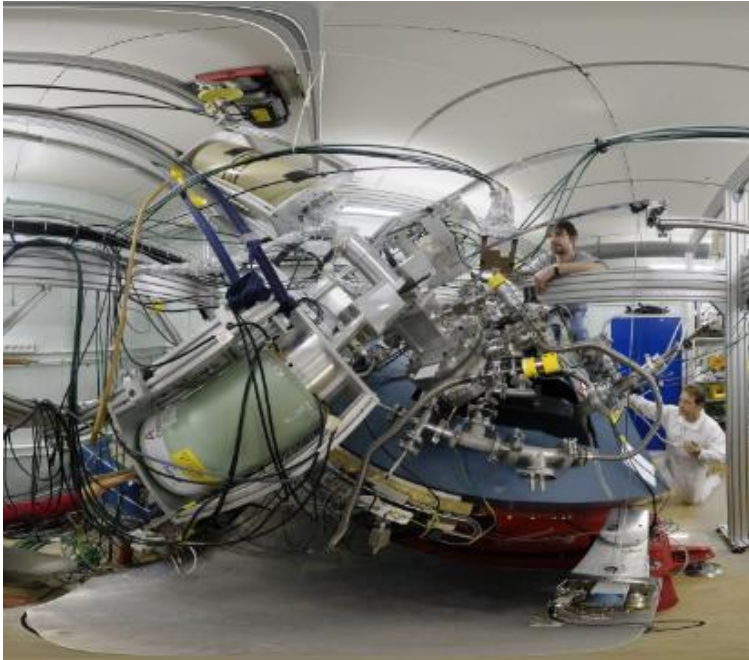
# Societal impact

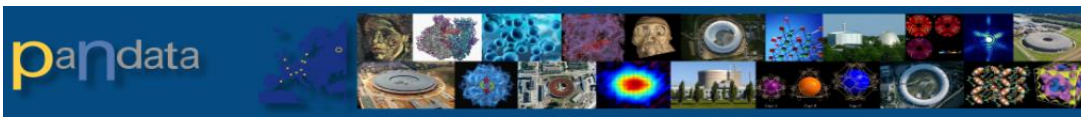




# Science

28 instruments + 10 CRG





# We are not alone

Snapshots of unique X-Ray & Neutron scattering users, over 18 months (2013 – 2014).

	ALBA	BERII	BESSYII	DESY	DLS	ELETTRA	ESRF	MLZ	ILL	ISIS	LCLS	SINQ	SLS	SOLEIL	SNS	neutron	photon	all
ALBA	1303	5	43	90	274	128	356	8	83	36	2	23	124	161	7	122	679	1303
BER II	5	237	27	28	20	0	42	66	104	50	0	75	22	7	37	162	86	237
BESSY II	43	27	838	128	96	95	143	16	31	16	28	29	119	100	11	76	418	838
DESY	90	28	128	3680	396	255	901	110	167	92	151	82	326	246	63	343	1579	3680
DLS	274	20	96	396	10445	297	1606	82	485	763	70	144	559	526	124	1136	2598	10445
ELETTRA	128	0	95	255	297	3422	480	21	99	41	68	14	218	379	12	149	1171	3422
ESRF	356	42	143	901	1606	480	10786	203	731	356	102	203	899	1390	155	1165	4242	10786
MLZ	8	66	16	110	82	21	203	1430	409	167	3	222	52	46	158	601	353	1430
ILL	83	104	31	167	485	99	731	409	4138	606	3	384	130	239	316	1252	1304	4138
ISIS	36	50	16	92	763	41	356	167	606	3406	9	236	101	84	267	891	1052	3406
LCLS	2	0	28	151	70	68	102	3	3	9	1123	1	79	44	6	17	329	1123
SINQ	23	75	29	82	144	14	203	222	384	236	1	1424	250	65	185	614	501	1424
SLS	124	22	119	326	559	218	899	52	130	101	79	250	3981	366	64	365	1637	3981
SOLEIL	161	7	100	246	526	379	1390	46	239	84	44	65	366	5134	40	349	2145	5134
SNS	7	37	11	63	124	12	155	158	316	267	6	185	64	40	3723	581	327	3723

# Experimental data management



Co-funded by the European Union :

PaNData-Europe Grant Agreement No 261537

PaNData-ODI Grant Agreement No RI-283556

INSTITUT MAX VON LAUE - PAUL LANGEVIN





# Open Data/Science for a facility?

- Data are the real/factual production.
- Knowledge (peer-reviewed articles, conference contributions, thematic courses, software ...) is the main output.

Openness is a tool for increasing our **impact**.

# Data Policy

Based on the PaNData framework



## Open data & how to protect and credit our users?

- The facility shall act as a custodian for the data.
- All raw data will be curated in a well-defined format with a unique ID (DOI).
- Metadata is captured automatically and resides either within the raw data files, and/or in an associated on-line catalogue.
- Users can release or give access to their data at any time, by default access to raw data and the associated metadata is restricted to the experimental team for a maximum period of 3 years. Thereafter, it will become publicly accessible.
- The embargo period can be extended on requests to the direction.
- Publication based on data must acknowledge the source of the data and cite its unique identifier (CC-BY licence).

<https://www.ill.eu/DataPolicy>

# Accessible: DP

## Scientific data policy

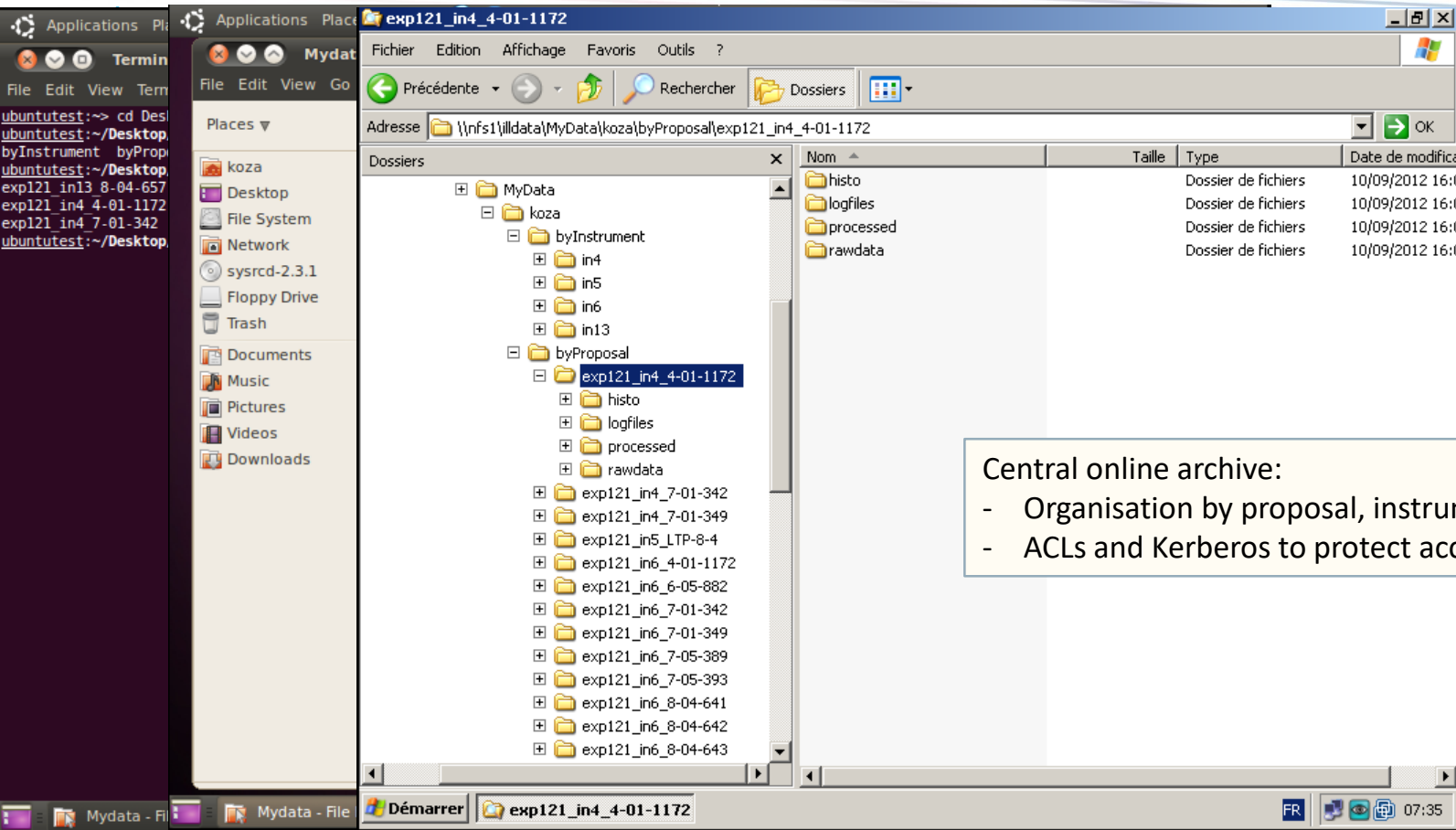
- Open Data policy based on PaNData-EU work.
  - Published in Sept 2011
  - In place since Oct 2012 (i.e. 1<sup>st</sup> experiment under DP)
  - 2013 complete set of Data Management Services available for users: search, access, annotate, archive, identify, publish.
  - Revised In July 2017
    - Include per default CRG instruments
    - More practical : "PID" replaced by "DOI"
    - All type of scientific data (Raw, reduced and results stored on ILL infrastructure)
    - Completely rewritten by a scientist to simplify wording and make it more practical and understandable by users.

<https://www.ill.eu/DataPolicy>

Identification of the proposal  
on the instrument.

Setting username to ortizh

# Archive: ACLs & user experience improvements



The image shows a Linux desktop environment with a terminal window on the left and a file manager window on the right. The terminal window displays a series of commands and their outputs, showing the user navigating through the file system. The file manager window shows a directory tree with a hierarchical structure. The address bar indicates the current path is `\\nfs1\\illdata\\MyData\\koza\\byProposal\\exp121_in4_4-01-1172`. The left pane shows a tree view with folders like `koza`, `MyData`, `koza`, `byInstrument`, `in4`, `in5`, `in6`, `in13`, `byProposal`, and `exp121_in4_4-01-1172`. The right pane shows a list of files and folders within the selected directory, including `histo`, `logfiles`, `processed`, and `rawdata`.

Applications Place

Termin

Mydat

File Edit View Tern

ubuntutest:~> cd Des

ubuntutest:~/Desktop

byInstrument byProp

ubuntutest:~/Desktop

exp121\_in13 8-04-657

exp121\_in4 4-01-1172

exp121\_in4 7-01-342

ubuntutest:~/Desktop

Places ▾

koza

Desktop

File System

Network

sysrctd-2.3.1

Floppy Drive

Trash

Documents

Music

Pictures

Videos

Downloads

Fichier Edition Affichage Favoris Outils ?

Précédente

Rechercher

Dossiers

Adresse \\nfs1\\illdata\\MyData\\koza\\byProposal\\exp121\_in4\_4-01-1172

Dossiers

Nom

Taille

Type

Date de modifica

MyData

koza

byInstrument

in4

in5

in6

in13

byProposal

exp121\_in4\_4-01-1172

histo

logfiles

processed

rawdata

exp121\_in4\_7-01-342

exp121\_in4\_7-01-349

exp121\_in5\_LTP-8-4

exp121\_in6\_4-01-1172

exp121\_in6\_6-05-882

exp121\_in6\_7-01-342

exp121\_in6\_7-01-349

exp121\_in6\_7-05-389

exp121\_in6\_7-05-393

exp121\_in6\_8-04-641

exp121\_in6\_8-04-642

exp121\_in6\_8-04-643

histo

logfiles

processed

rawdata

Dossier de fichiers

Dossier de fichiers

Dossier de fichiers

Dossier de fichiers

10/09/2012 16:0

10/09/2012 16:0

10/09/2012 16:0

10/09/2012 16:0

Central online archive:

- Organisation by proposal, instrument, dates.
- ACLs and Kerberos to protect accesses.

Démarrer

exp121\_in4\_4-01-1172

FR

07:35

ANGEVIN

NEUTRONS FOR SOCIETY



# Accessible: Data Portal

Global Search

Proposals list

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

- Provide access to data, meta-data, logs, DOIs landing page, ...
- Tools for managing data authZ

- 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.

Global Search

in5 quasielastic

☐ Include experiments without data

Search

Advanced Search

My role in the experiment (?)

Proposal number (?)

Proposals list

18 proposal(s) found with the given search filter

8-04-684 : The influence of cofactor binding on thermal fluctuations of alcohol dehydrogenase studied by using quasielastic neutron scattering

- 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: visi

6-02-512 : Dynamics

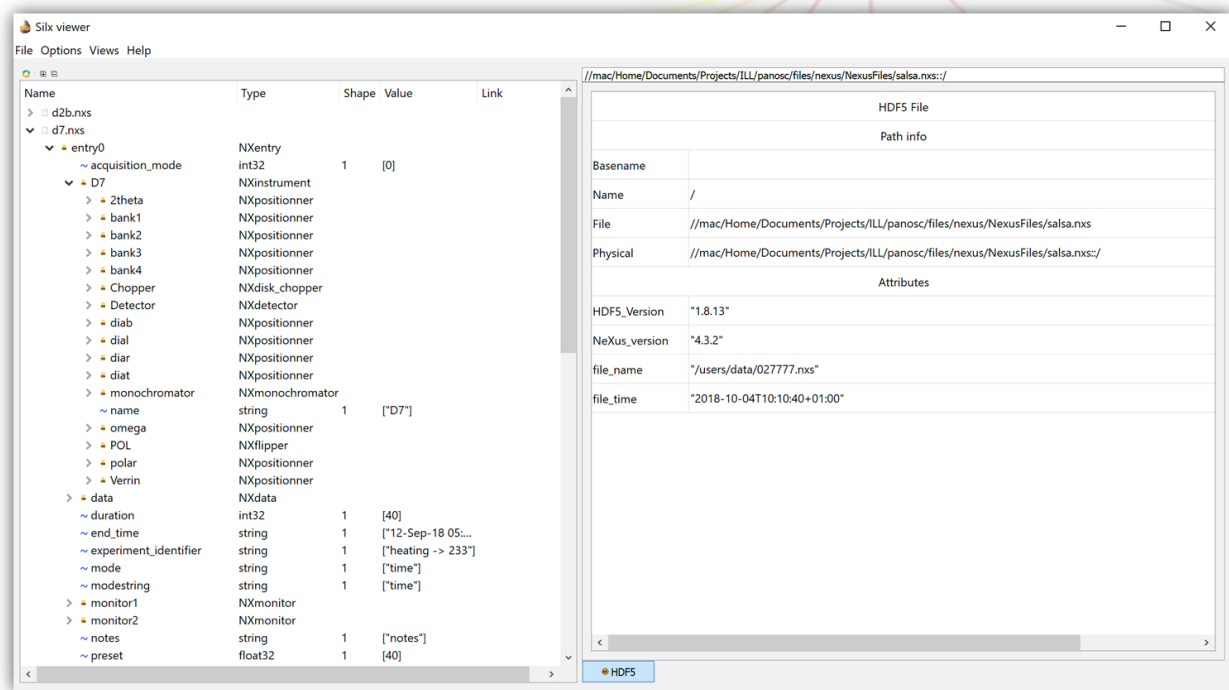
File Name	Size	Action
105641.nxs	2999	Download files (one by one)
105642.nxs	2993	Download a zip with all files
105643.nxs	7798	Select all files
105644.nxs	5805	
105645.nxs	12 MB	
105646.nxs	12 MB	
105647.nxs	12 MB	
105648.nxs	13 MB	

Index all available information: Proposal, experimental report, data file annotation, publications, ...

# Interoperable

## File formats

- ROOT for Nuclear Physics/HEP experiments.
- NeXus for all the other instruments.



# Findable

# DOI

4056 DOIs  
registered  
since 2012

Collaboration  
with DataCite  
- CNRS INIST

## 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](#). Thank you for your understanding.

## Title

Chitosan/gelatin enzymatically cross-linked hydrogels: Composition and temperature effects on the gel's molecular structure.

## Abstract

Hydrogels from biopolymers have been attracting increasing interest in biomedicine, but the lack of structural understanding hinders the development of hydrogel scaffolds obtained from the blend of tilapia (fish) gelatin and chitosan. The hydrogels are cross-linked by the microbial transglutaminase (MTG) enzyme. The hydrogelization will be studied: (a) Chemical gelation, enzymatic reaction (b) Physical-chemical gelation, enzymatic reaction done in presence of the MTG. The hydrogel is more than the added properties of its components and it also requires a particular set of mechanical properties, which are related to its structure at the nanoscopic level. We propose to use SANS to achieve an understanding of the structure of the networks at the nano-scale level, both after and during the gelation. The experiments will give us precious insight into hydrogels architecture and properties allowing us to better correlate bulk and nano-properties in order to allow a better design of the final hydrogels.

## Experimental Report

Download Experimental Report

## Download Data

The data is currently only available to download if you are a member of the proposal team.

Download Data

## Data Citation

The recommended format for citation is:

da Silva, MA  
gel's m

This data has been cited by 1 articles.

**Exploring the Kinetics of Gelation and Final Architecture of Enzymatically Cross-Linked Chitosan/Gelatin Gels**

Marcelo A. da Silva, Franziska Bode, Isabelle Grillo, and Cécile A. Dreiss (2015).

DOI

Data has been collected

[doi:10.1021/acs.biomac.5b00205](https://doi.org/10.1021/acs.biomac.5b00205)

## Metadata

Linking data and people  
through  
ORCID/ResearcherID

## Authors

da Silva, MA ([ORCID](#), [ResearcherID](#))

DREISS Cecile

## Proposals

20131 (19-02-2013 - 09-04-2013)

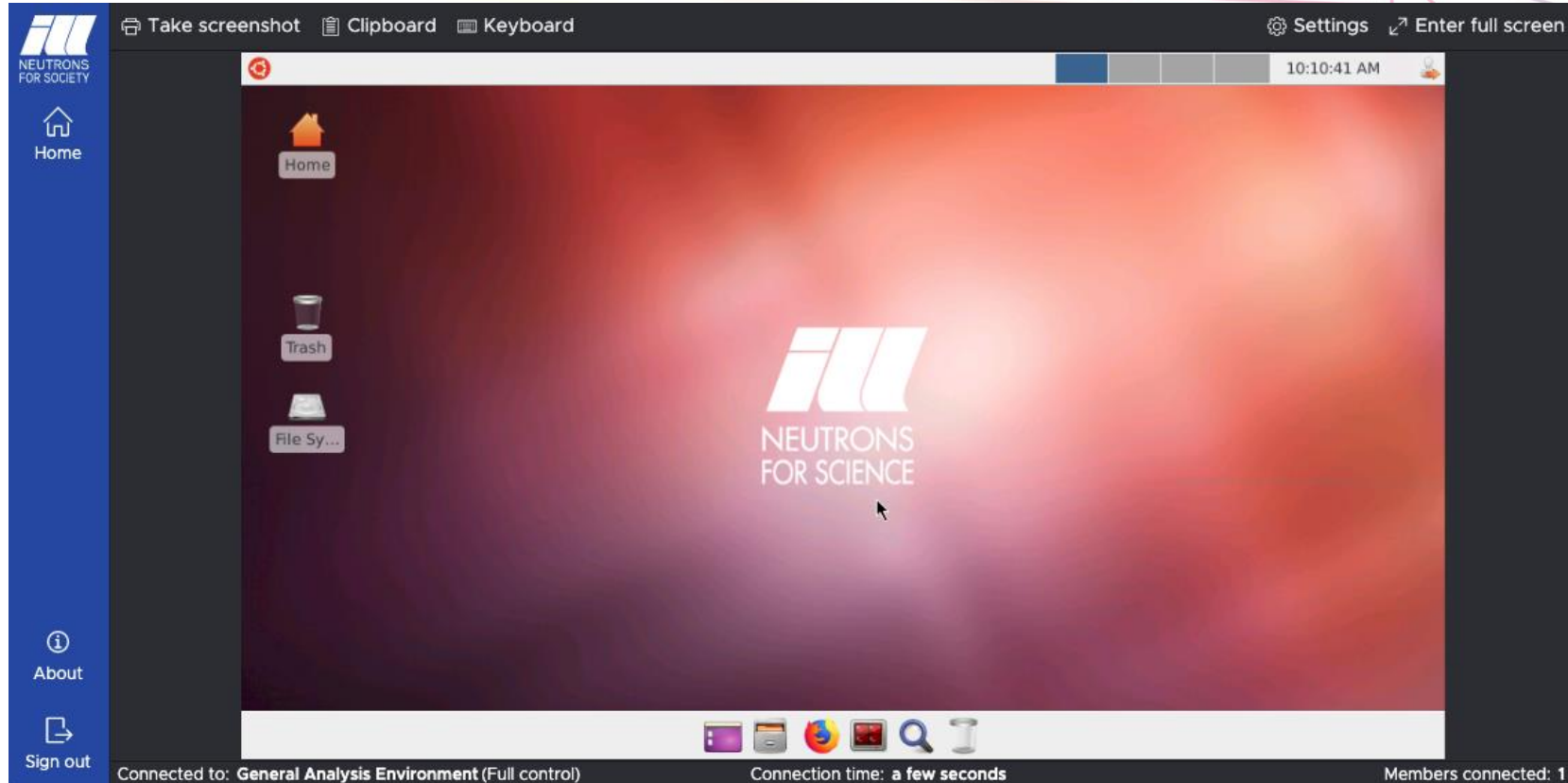
Proposal number  
9-11-1654

## Experiment Parameters

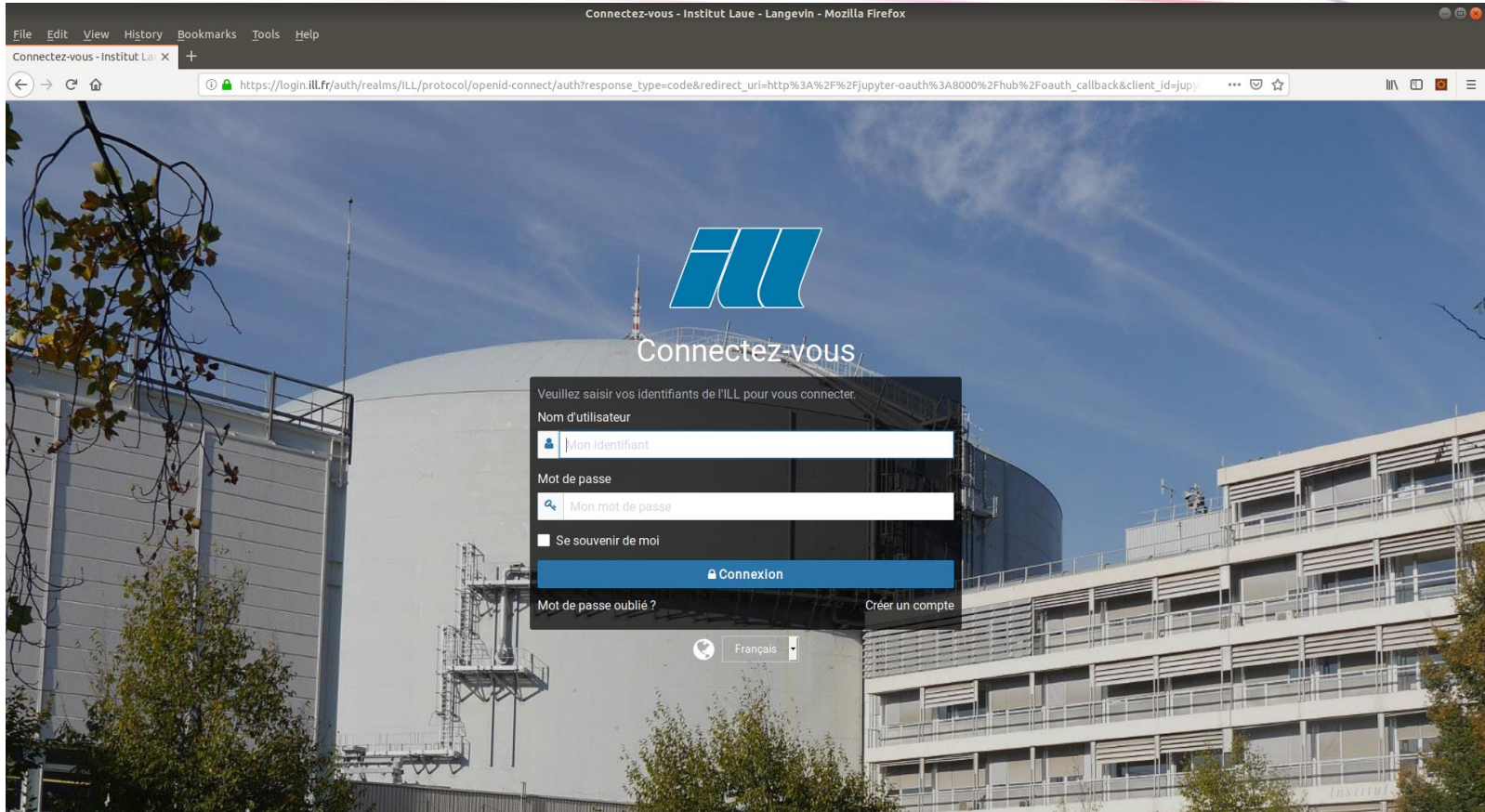
This data is not yet public

## Parameters

# (Re)Usable: DAaS Visa



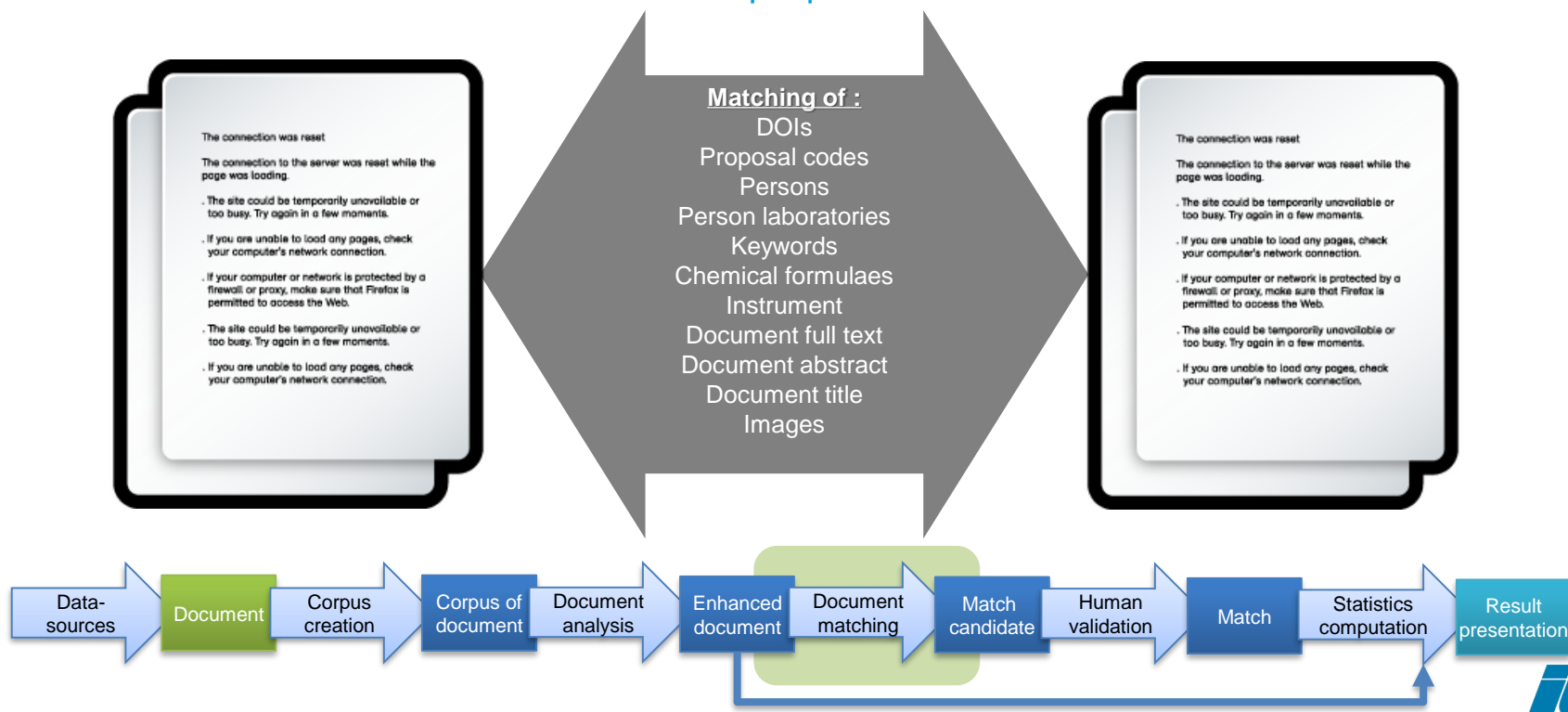
# (Re)Usable: DAaaS Jupyter



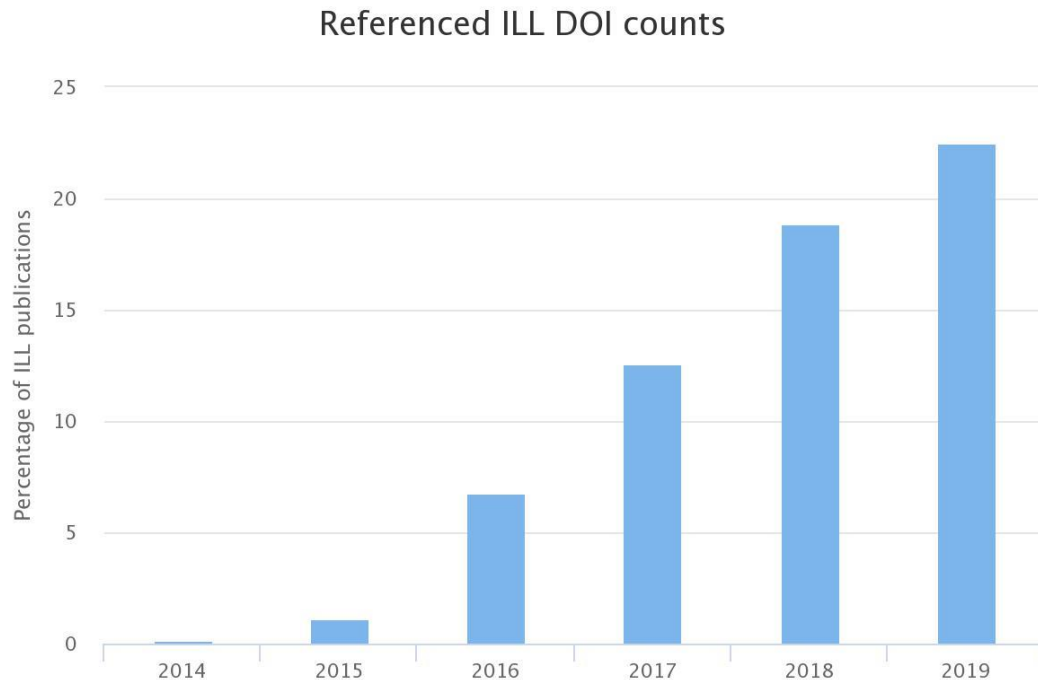


# Are we successful? PUMA-BI

A tool to find and rank similarities between proposal documents and scientific articles



# PUMA-BI: statistics examples



# Some results (Oct 2019)

- 1325 Data sets openly accessible
- 70 Access authorisation added after the experiment (articles reviewers, new collaborations, ...).
- 184 Full Open Data Sets downloaded by non proposal members.
- 7.5 M data files downloaded since 2015

Year	Download by proposal members	by NON proposal members
2015	80925	157
2016	708387	20434
2017	552841	5339
2018	2654162	53098
2019	3410658	12868

Contact: [data@ill.eu](mailto:data@ill.eu)

Data portal: <https://data.ill.eu>

Policy: <https://www.ill.eu/DataPolicy>

Software catalogue: <https://software.pan-data.eu>

PaNData Collaboration: <http://pan-data.eu>

PaNOSC: <https://panosc.eu>



71, rue des Martyrs - 38000 Grenoble - [www.ill.eu](http://www.ill.eu)

NEUTRONS  
FOR SOCIETY®

