

# Experimental Data Storage Management in NeXus Format at Synchrotron SOLEIL

Alain BUTEAU

On behalf of the SOLEIL Computing division



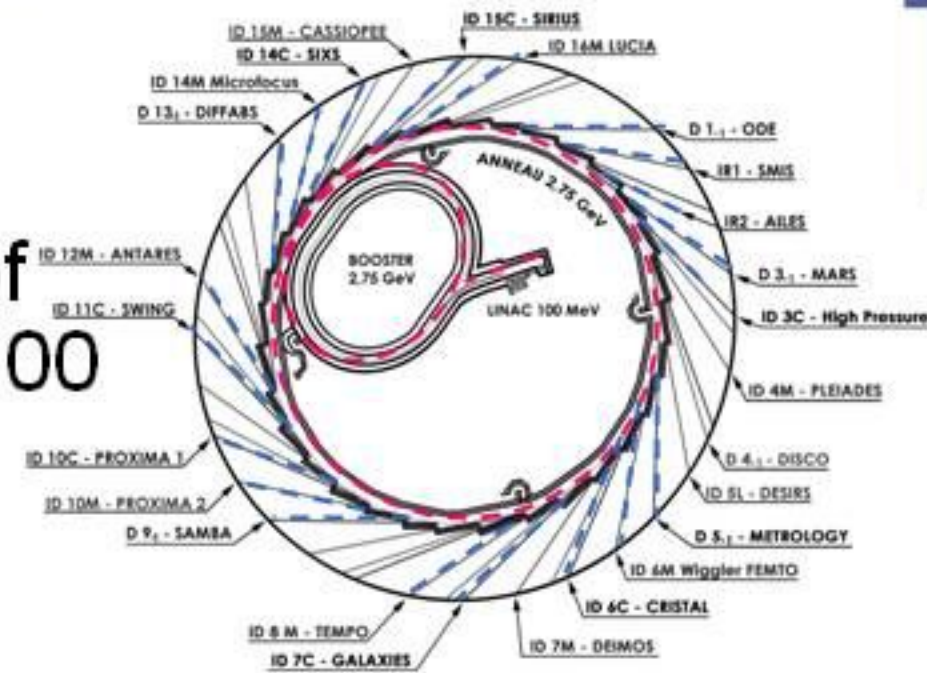
*Synchrotron SOLEIL, Saint Aubin, France,  
<http://www.synchrotron-soleil.fr>*

- A few reminders on experimental data files production at Soleil
- Design guidelines and solutions
  - On the Controls and Data Acquisitions software side
  - On the DataStorage infrastructure side
- Successes and difficulties
- Conclusion and next steps

- **A few reminders on experimental data files production at Soleil**
- *Design guidelines and solutions*
  - *On the Controls and Data Acquisitions software side*
  - *On the DataStorage infrastructure side*
- *Successes and difficulties*
- *Conclusion and next steps*

- ❑ About 20 beamlines in operation
  - 14 of them opened to external users
- ❑ For each beamline, the daily volume of files ranges from a few Mbytes up to 100 Gbytes

- This volume is increasing with:
  - *2 dimensional detectors used instead of punctual detectors*
  - *Continuous (i.e without motor stops) scans*
  - *2D pixel detectors (XPAD, PILATUS)*



See WEP054

- ❑ A great diversity of scientific applications:
  - *physics, chemistry, new materials, environmental science, biology, ...*
  - Which means diversity of detectors, acquisition process, data volumes, data lifetime policies

**nature**

[www.nature.com/nature](http://www.nature.com/nature)

Vol 461 | Issue no. 7261 | 10 September 2009

# Data's shameful neglect

Research cannot flourish if data are not preserved and made accessible. All concerned must act accordingly.

**nature**news

[nature news home](#) [news archive](#) [specials](#) [opinion](#) [features](#) [news blog](#) [events](#)

**Specials**

[See all specials](#)

## Data Sharing

Sharing data is good. But sharing your own data? That can get complicated. As two research communities who held meetings in May on the issue report their proposals to promote data sharing in biology, a special issue of *Nature* examines the cultural and technical hurdles that can get in the way of good intentions.

- [EDITORIAL](#)
- [FEATURE](#)
- [OPINION](#)
- [ELSEWHERE IN NATURE](#)



SOLEIL Computing division has the mission to store and retrieve our “data files patrimony” during the 30 years of SOLEIL operation



Challenge No 1 : Which information must be stored for a given experiment to be able to retrieve the data and re-process it in the future ?

Challenge No 2 : On which physical support should we store data files on a long term basis ?

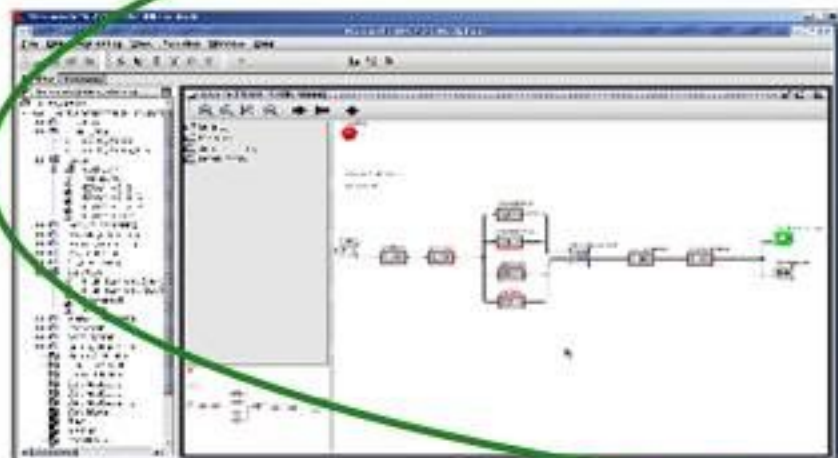
- *A few reminders on experimental data files production at Soleil*
- **Design guidelines and solutions**
  - *On the Controls and Data Acquisitions software side*
  - *On the DataStorage infrastructure side*
- *Successes and difficulties*
- *Conclusion and next steps*

- Decouple 3 activities and competences
  - Data acquisition
    - ✓ *The **data acquisition specialist** should take care of all issues regarding the interfacing of the detection system.*
  - Data collection processes
    - ✓ *The **beamline scientist** should focus on the definition of the experimental process needed to get a good measurement.*
  - Data storage
    - ✓ *The **data storage specialist** should take care of the details regarding the proper organization of all data describing the measurement.*

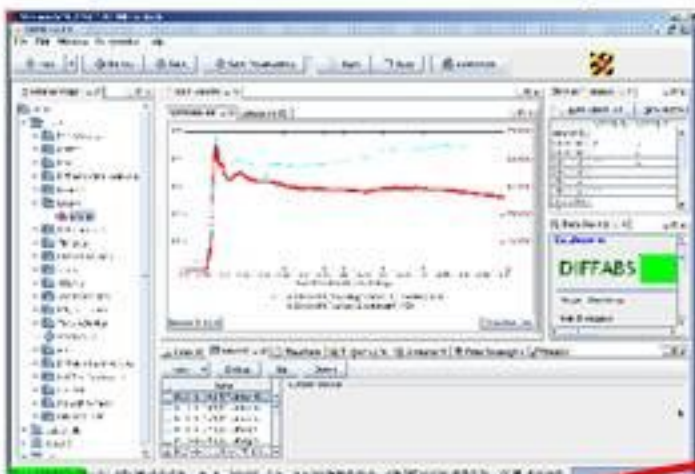


# Storage of the Experimental Data: Data Recording software decomposition

Sequencing application



Scanning application



**Beamline scientist**

**Tango bus**

**Detector device**

*Interfaces detection system*

**Data Acquisition specialist**

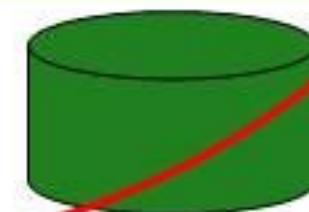
**DataRecorder device**

*Writes NeXus files*

**Data Storage specialist**

*Writes metadata information in an index database*

**DataStorage infrastructure**



- Keep close to our scientific community

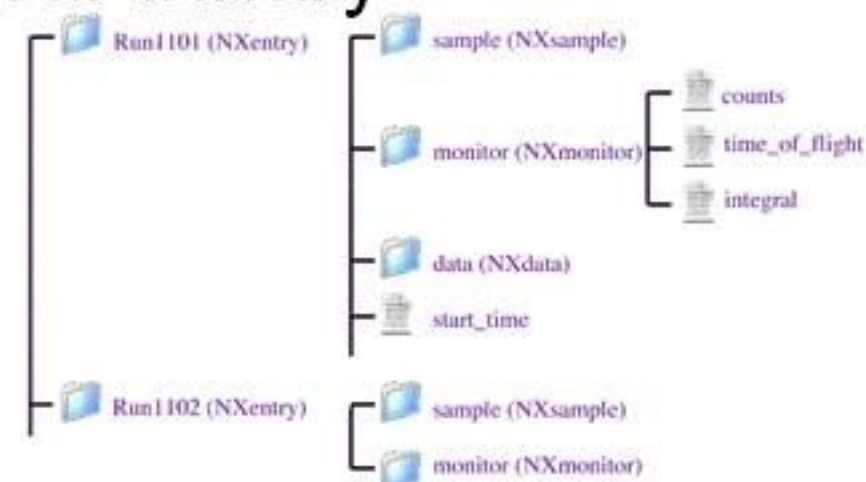
- To benefit from others institutes data reduction and analysis software developments

- NeXus was a quite obvious choice

- ✓ It is not linked to a particular scientific community
- ✓ It simplifies HDF usage
- ✓ It exists since almost 15 years in Neutrons sources and X Rays facilities
  - Even is not largely diffused !

- NeXus is an self-describing hierarchical binary format

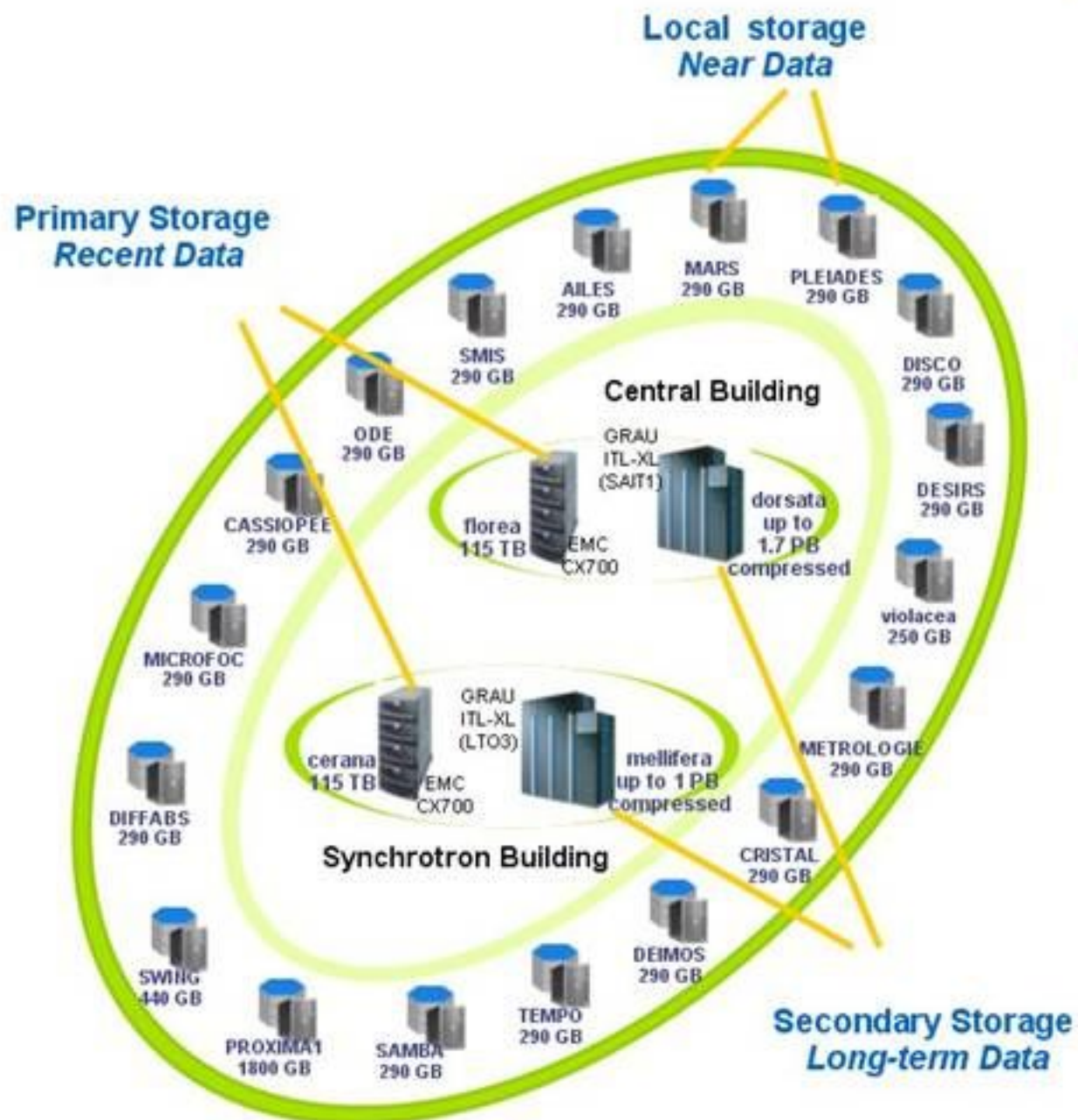
- Based on HDF format
- Standardization of metadata information



- *A few reminders on experimental data files production at Soleil*
- ***Design guidelines and solutions***
  - *On the Controls and Data Acquisitions software side*
  - ***On the DataStorage infrastructure side***
- *Successes and difficulties*
- *Conclusion and next steps*

- Easily manage the lifecycle of data files
  - Which is different according to
    - ✓ The beamline
    - ✓ The user doing the experiment
- Provide built in mechanisms for “high reliability” and “high security”
- Be independent from hardware solutions
  - Storage vendors often offer solutions based on proprietary hardware.
  - Keep the future extensions of our system open to new technological progress in the mass data storage industry.

We selected the **Active Circle** solution for the beamlines data files storage



- Each beamline has a local “ActiveCircle cell”
  - DataRecorder device writes NeXus files on this local cell
  - **High availability**
- Data files are then :
  - transparently migrated on the network of ActiveCircle cells
    - ✓ Which may be massive storage disks
    - ✓ Or DLT tapes
    - ✓ Or other support in the future ...
    - ✓ **High security and extensibility**
  - According to different criteria (*beamline, users, file age*)
    - ✓ **Integrated Lifecycle Management**

- *A few reminders on experimental data files production at Soleil*
- ***Design guidelines and solutions***
  - *On the Controls and Data Acquisitions software side*
  - *On the DataStorage infrastructure side*
- ***Successes and difficulties***
- *Conclusion and next steps*

- NeXus is today used on the majority of the beamlines
  - And NeXus usage is quickly growing

Beamline	NeXus files	Volume(GB)	Average Size (MB)
Cristal	7 733	7	0,9
Diffabs	10 340	61	5,9
Ode	45 542	11	0,3
Swing	15 116	169	11,2
Samba	91 679	205	2,2
Deimos	72	3	40,4
Proxima1	1 331	1	1,1
Cassiopee	1 402	0	0,1
Tempo	704	0	0,1
Antares	375	1	1,9
Lucia	2 784	7	2,2
<b>Total</b>	<b>180 002</b>	<b>465</b>	<b>6,0</b>

*NeXus file production in 2009*

- But it has been a difficult and painful path !!
  - The users did not initially understand our motivations



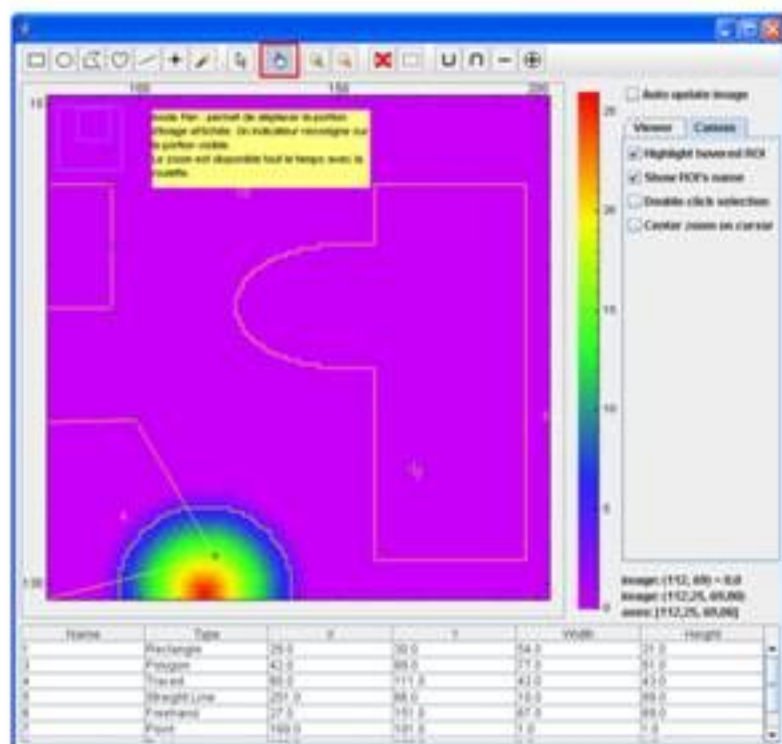
The screenshot shows the 'naturenews' logo at the top. Below it is a navigation menu with links: 'nature news home', 'news archive', 'specials', 'opinion', 'features', and 'news'. Underneath the menu, there is a section titled 'Access' with the text 'This article is part of Nature's premium content.' Below that, it says 'Published online 9 September 2009 | Nature 461, 160-163 (2009) | doi:10.1038/461160a'. At the bottom, there is a red 'New Feature' tag and the article title 'Data sharing: Empty archives' which is circled in green.

- NeXus is not yet a fully defined standard for experimental datasets
  - *We decided to initially use NeXus only as the SOLEIL standard container for our data*
- **ActiveCircle was an innovative choice**
  - But we had to cope with a number of bugs and technical problems during the first year of operation in 2007

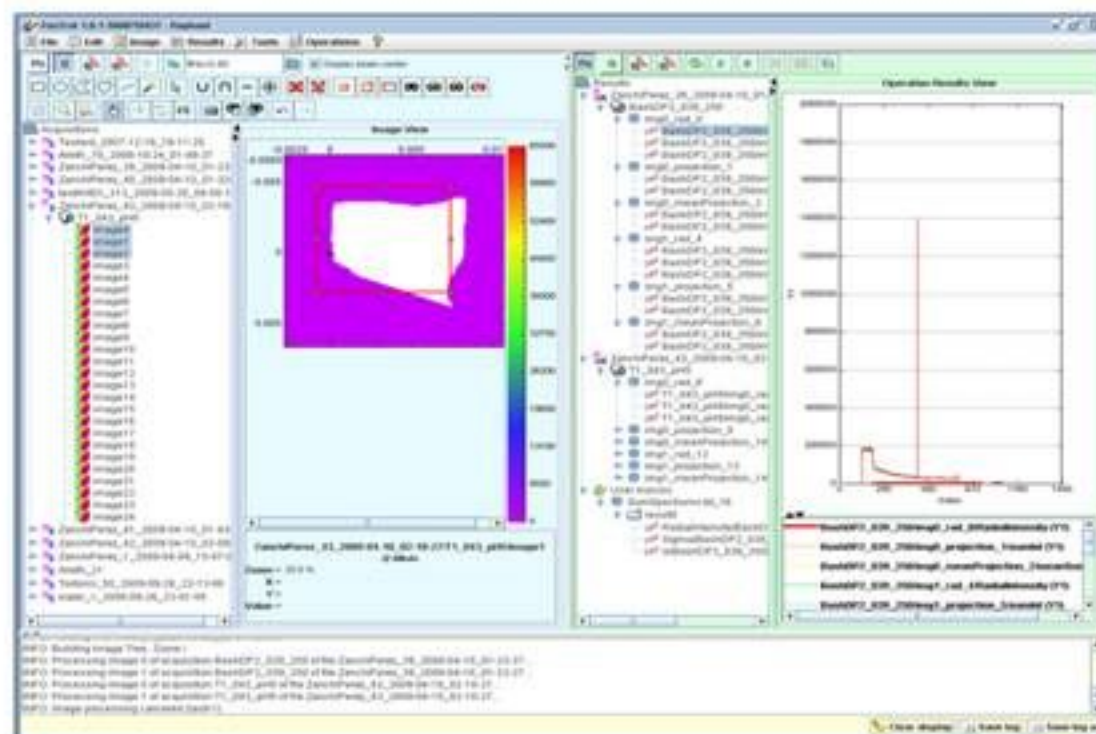


- *A few reminders on experimental data files production at Soleil*
- ***Design guidelines and solutions***
  - *On the Controls and Data Acquisitions software side*
  - *On the DataStorage infrastructure side*
- *Successes and difficulties*
- ***Conclusion and next steps***

- NeXus is becoming a de facto standard at SOLEIL
  - which federates our software development for data reduction, data analysis and data visualization



The COMETE library of data visualization components  
<http://comete.sourceforge.net>



A Small Angle Scattering Data Reduction application fully based on NeXus

- Of course, the most challenging goal is still to be able to transparently exchange data files between institutes.
  - *This would allow us to benefit from the data analysis tools developed in our scientific community.*
- In this respect, we will begin to work before end of year with DESY, ESRF and DIAMOND which have also shown an interest on NeXus.

**See you at ICALEPS 2011**