The NeXus format,
one more step towards homogeneity

> NeXus is a good candidate to harmonize data formats

Stéphane Poirier,
SOLEIL Computing division – "control, order and acquisition" group

What is the NeXus format, what are its advantages?
The NeXus format was not invented for the SOLEIL synchrotron. It was created in about 1996 in research institutes that were using neutron sources to analyze matter. Laboratories needed a format for recording larger volumes and a greater variety of data, for storing both experimental data and data describing the context of the experiment. They based it on an existing format developed in U.S laboratories, the Hierarchical Data Format (HDF), which is extremely powerful but rather complex to use. The NeXus format is an approach that makes HDF simpler to use and also specifies the contents of the files. The most important added value of NeXus is that it does not impose constraints in terms of number and volume of datasets that can be stored in the file, unlike most of the formats used until now.

Why is the NeXus format of interest to SOLEIL?
I arrived at SOLEIL in 2004 and my mission was to define a way of recording the data: what format and with which tools. I spent my first few months searching the "ideal" format and I then came across NeXus. Neutron sources and synchrotron radiation allow research which is relatively similar, so the specifications I found with the NeXus format were applicable without modification. Today, more and more of the measurements recorded on the beamlines represent large volumes of information that will always grow in size. Thus, the classic text formats that people used previously have proved unsuitable. On the other hand, one of the constraints imposed on the site was to use the same format on all beamlines. What was needed was a generic format to meet these diverse needs. There has also been an effort at the European level to try to harmonize data formats and NeXus is a good candidate. But it is not alone.
Now, at SOLEIL, in collaboration with an Australian Institute (ANSTO), we are trying to go beyond NeXus and free ourselves completely from the data format.
Having standardized the format of data files produced by the beamlines allows SOLEIL to have a "library" of its experimental data, a library the content of which can be explored computationally using, for example, the Web application TWIST. SOLEIL thus possesses the scientific memory of experiments that have taken place on its beamlines.

Managing on a large-scale basis, our experimental data is new in the synchrotron world and puts us in a very favorable position for different European projects (PANDATA, PNI...) related to this issue. Today, we are taking advantage of this position trying to convince IT personnel of partner institutes (ESRF, DESY, etc.) to join us in the "CommonDataModel" approach, an ambitious project that aims to unify access to scientific data for the data analysis software. So stay tuned…

Have you encountered problems in getting NeXus accepted at SOLEIL?
It took time convincing SOLEIL scientists, so it was quite a slow process. The biggest barrier is that in order to use an HDF file, you need software that can read it. The development of data reduction applications able to read NeXus files started one year ago, three years after the production of our first Nexus file. In the meantime, so that people can work, we have developed a "conversion tool" that allows NeXus files content to be read by existing analysis applications. For the SWING beamline, we developed an application called FOXTROT which can analyze experiments that have been stored in NeXus files. This application is heavily used on that beamline and we are starting to deploy it on other beamlines.

→ **For further information** www.synchrotron-soleil.fr/Soleil/ToutesActualites/2009/TWIST

→ **Contact :** stephane.poirier@synchrotron-soleil.fr

IN BRIEF

SOLEIL INVOLVED IN THE PARIS-SACLAY-INNOVATION NETWORK

SOLEIL is one of 18 participants forming part of the Paris-Saclay-Innovation network set up for research development and the creation and development of innovative technology companies in order to give the business community access to experimental facilities, research skills and the scientific and technical know how of the laboratories on the Saclay campus and in the Yvette valley. The network runs a bilingual website presenting its offer (www.paris-Saclay-innovation.fr) and had a stand at the Orly TechInnov on February, 11th, 2011.

ICM HAS CHOSEN THE SAME NEW DATA STORAGE SOLUTION AS SOLEIL

The Brain and Spinal Cord Disorders Institute (ICM) has just adopted the innovative scientific data storage solution Active Circle, already installed at SOLEIL.
A visit of SOLEIL, followed by discussions with experts in the Computing Division at SOLEIL, helped ICM decide on the Active Circle solution, proposed by a French company with its headquarters in Jouy-en-Josas, in close proximity to the Saclay campus.