1 NeXus for Synchrotrons

1.1 NeXus Workshop, May 10-12, 2010, PSI

21 IT personnel and scientists from various synchrotron and neutron source around the world gathered at PSI to attend a workshop on the use of the NeXus data format for synchrotron instrumentation. The purpose of the workshop thus can be summarised:

- Disseminate about information about the NeXus data format and how it works

- Discuss the use NeXus for synchrotron instrumentation

- Create a wishlist for the NeXus International Advisory Committee (NIAC) with changes needed to NeXus in to make it usable for synchrotron instrumentation.

The organisers thank all participants for many open and constructive discussions.

1.2 Workshop Outcomes

There was a general agreement that NeXus has much to offer to the synchrotron community. Beyond an efficient physical file format, HDF-5, NeXus offers a standardized way to store and document data. In addition NeXus provides an API and utilities which simplify the use of NeXus files. However, in order to become useful for synchrotron applications some changes to NeXus are required.

1.2.1 Multi Method Instruments

Synchrotron beamlines often utilise several different detectors and detector types in order to combine multiple techniques in simultaneous measurements. NeXus currently asks for separate NXentry groups to be written for each technique. This is good if one measurement is written to a file. However, there is a second requirement that multiple scans, multiple measurements, possibly a whole log of an experimental session is written to one NeXus file. Then having different techniques in different NXentries will make the files difficult to understand as the relationship between different measurments is lost. Thus, in order to keep the data from these multiple techniques together, it is desirable to have the ability to write it all into a single NXentry in a NeXus. The current NeXus application definitions refer to the same names and paths and so there are many name collisions when trying to satisfy two application definitions in one NXentry in a file. The ability to combine application definitions could be enabled by modifying the application definitions to refer to new and separate groups inside the main NXentry of the NeXus file that refer to the particular application/technique name and which contains all of the data (or links to it) that is relevant to that application/technique. For an example experiment that involves a combination of SAS and Fluorescence, the proposed NeXus structure could look like:

```
entry:NXentry/
```

```
definition = "NXSas, NXFluo"
user:NXuser/
sample:NXsamle/
instrument:NXinstument/
  SASdet:NXdetector/
  fancyname:NXdetector/
  fancyname2:NXdetector/
  ...
SAS:NXsubentry/
  definition = "NXSas"
  instrument:NXinstrument/
    detector:link to SASdet
  data:NXdata/
Fluo:NXsubentry/
  definition = "NXFluo"
  instrument:NXinstrument/
    detector:link to fancyname
    detector2:link to fancyname2
  data:NXdata/
```

In the above NeXus tree, the entire beamline state could be stored in entry/instrument and then any subset of this that is relevant to the SAS or Fluorescence techniques would then be linked within the entry/SAS/instrument and the entry/Fluo/instrument groups as defined by the current application definitions with a minor change in the heirarchy. The advantages of this approach are:

- Only minor changes from current practice.

- The only name collisions to worry about are the names of the applications/techniques themselves.

- Application definitions need not be concerned with the names and paths that other application definitions proscribe.

- The paths for each application remains well defined and an analysis program for either technique can find the relevant data without having to understand the other techniques present in the file. Further, the same analysis programs can read the multi-technique files in the same way (i.e. with the same code) exactly the same as they read single-technique files.

- A user inspecting the data manually can find all the relevant information for a particular analysis in the one group and so doesn't need to understand the entire beamline.

One drawback of this approach is that the beamline staff would have to define many links when configuring the data acquisition software. However, this is necessary work regardless of how the data is saved since the user must be informed of how the different instrument components and detectors relate to the various analyses anyway. In fact, NeXus and the above proposal simplifies this task by clearly documenting in a formal manner where the relevant information can be read.

Some examples of beamlines that would benefit from this proposal include:

- Fluorescence + Absorption + Diffraction (Beamline L, Doris, Beamline P06, PETRA III).

- PX + fluorescence: In PX often a fluorescence-signal is recorded, especially for SAD-/MAD-measurements (PETRA beamlines P11 P13 P14).

- SAXS+fluorescence: fluorescence is often used as a second signal in SAXS (PETRA P03).

- SAXS + ellipsometry: Beamline BW4, DORIS

Summarizing this discussion, the suggestion is to allow NXentry or possibly new NXsubentry groups underneath NXentry. Each of which can adhere to a different application definition. All participants agreed that a good means of handling multi technique instruments in NeXus is essential for the adoption of NeXus at synchrotron facilities. This is a MUST HAVE!

1.2.2 Scaled Data

The data rates possible at synchrotron facilities and the new pixel detectors test current computing technology to their limits. There may not be enough time to scale or convert data on the fly. Thus the acceptance of the Scaled Data proposal from the CIF study is another MUST HAVE. Furthermore it was suggested to rename the linearity attribute to transform, not to be confused with the transform field of the CIF coordinate suggestions, and to add additional transformations, namely for a polynomial data scaling and a conversion from anything to energy using the Bragg equation.

1.2.3 Simplified Hierarchy

In order to satisfy the requirements of the beamline scientist an additional, simplified NeXus hierarchy was proposed:

```
entry:NXentry
   measurement:NXmeasurement
      positions:NXpositioners
      scalars:NXscalar
      images:NXimagedata
```

The meaning is that the NXpositioners groups contains a list of all constants and motor positions, NXscalar arrays of all parameters varied during the scan and NXimageData what has been captured during a scan or measurement. This structure is for the expert, the instrument scientist, who knows his instrument by heart and wishes to be able to plot anything against anything in his instrument. This is not meant to be something which is an archive format. It was pointed out that such a simplified view can be generated by a data analysis application from a full NeXus NXinstrument view. It remains unclear if this is a must have or an optional feature.

1.2.4 Special Techniques

The tomography people among thge participants got together and discussed the NeXus tomography application definitions. This discussion will continue beyond the workshop. An intermediate result is that the sample positions needs more detail then in the current application definition in order to allow for more exact reconstructions.

There was also a review of the SAS definition by some participants. The results were:

- All the SAS techniques should be joined in one application definitions

- The above require more fields to describe both the sample and detector position in more detail

- The position of the beam stop needs to be added

- A few more minor additions.

The NXsas definition will soon be updated to accomodate this feedback.

A couple of neutron people among the participants and discussed how to store S(Q,om) processed data. Agreement was reached to store such data as arrays of Qh, Qk, Ql, En, I. Though data may have some cohesion, in general after conversion the data is on a irregular grid and the data analysis or visualisation application will have to rebin the data anyway in order to generate a representation. This consensus will soon be cast into an application definition.

1.2.5 Additional Suggestions

There are some wishes about documentation: For binary packages it should be stated more clearly for which platform they are and what to do when they do not work. For example NeXus provides a Windows package which works on XP. But people with Vista or Windows 7 may have problems. It should also be clearer how application definitions work with multiple detectors.

There is a wish to store the energy profile in NXmonochromator, NXsource and NXbeam. Suggested name: energy_profile.

Some more fields in NXdetector: x_correction and y-correction. This is for storing the infromation where a pixel truly is. Technique for saying with which technique the detector is used. Live_time is total_counting_time - dead_time).

There is a wish to annotate data with an attribute dimensionality which can have values of: scalars, histogram, image, vertex. This to say that this data is a series of images, histograms etc. This helps a generic visualisation program to interpret the data properly. The string to give is part of an enumeration.

NeXus should be more specific about the encoding for content strings, like in titles, sample names etc. A good suggestion is to specify UTF-8.