The NeXus Data Format for muon Spectroscopy and Neutron or X-ray Scattering

Mark Könnecke

Paul Scherrer Institute
Switzerland

May 9, 2010

The Predicament of the Traveling Scientist

- A different data format wherever she goes

The Predicament of the Traveling Scientist

- A different data format wherever she goes
- Spends lots of time converting formats or writing readers

The Predicament of the Traveling Scientist

- A different data format wherever she goes
- Spends lots of time converting formats or writing readers
- Waits even longer to load data from inefficient data formats

NeXus

The Predicament of the Traveling Scientist

- A different data format wherever she goes
- Spends lots of time converting formats or writing readers
- Waits even longer to load data from inefficient data formats
- DA requires N files in different formats, notes, local knowledge

NeXus

The Predicament of the Traveling Scientist

- A different data format wherever she goes
- Spends lots of time converting formats or writing readers
- Waits even longer to load data from inefficient data formats
- DA requires N files in different formats, notes, local knowledge
- Cannot read her collaborators data

NeXus

The Predicament of the Traveling Scientist

- A different data format wherever she goes
- Spends lots of time converting formats or writing readers
- Waits even longer to load data from inefficient data formats
- DA requires N files in different formats, notes, local knowledge
- Cannot read her collaborators data
- Has to keep extra information in yet another form

NeXus

NeXus Design

- Complete data for typical use
- Extendable, add additional data as you please
- Self describing
- Easy automatic plotting
- Platform independent, public domain, efficient
- Suitable for a wild variety of applications

NeXus History

- Devised from three independent proposals by Jonathan Tischler, APS, Przemek Klosowski, NIST and Mark Koennecke, ISIS, PSI in 94-96
- Improved during various NOBUGS conferences
- NeXus International Advisory Committee, NIAC, since 2003
- Since 2003 yearly meetings of the NIAC
- We already considered many issues!
- Except for one year, we never had money to develop NeXus

NeXus Levels

1. Physical file format and API for accessing files
2. Rules for storing data in files
3. Component and application definitions
4. NeXus Utilities

Physical File Format

- Portable, self describing, extendable, public domain
- Hierarchical data format, NCSA, HDF-4, later HDF-5
- HDF-5:
  - grouping support
  - on the fly compression
  - reading/writing subsets
  - first dimension appendable
  - Public domain C, F77 access library
  - Used by: NASA, Boing, the weathermen, ...
- XML for those who wish to edit their data

NeXus

NeXus API

- NeXus-API hides complex HDF API
- Transparent access to all three supported physical file formats
- ANSI-C implementation
- Bindings: C++, F77, Java, python, IDL, SWIG
- January, 4, 2010: 1311217 files processed at PSI alone

NeXus API Example

```
nxfile = nxs.open('hrpt2008n152088.hdf','r')
nxfile.openpath('/entry1/data1/two_theta')
x = nxfile.getdata()
nxfile.openpath('/entry1/data1/counts')
y = nxfile.getdata()
nxfile.openpath('/entry1/title')
txt = nxfile.getdata()
nxfile.close()

plot(x,y)
xlabel('two theta')
ylabel('counts')
title(txt)
show()
```

NeXus Objects

- Files
- Groups identified by name and a classname beginning with NX
- Scientific data sets
- Attributes
- Links

Rules for Storing Data in NeXus Files

- NeXus files have a hierarchy
- NXentry
  - NXuser
  - NXsample
  - NXmonitor
  - NXdata
  - NXinstrument
    - NXmonochromator
    - NXdetector
    - ...

Why the Hierarchy??

- Supports self description and allows short names in components

Why the Hierarchy??

- Supports self description and allows short names in components
- Name, classname pair allows for multiple components of the same type

Why the Hierarchy??

- Supports self description and allows short names in components
- Name, classname pair allows for multiple components of the same type
- NXentry allows for multiple datasets in the same file

Why the Hierarchy??

- Supports self description and allows short names in components
- Name, classname pair allows for multiple components of the same type
- NXentry allows for multiple datasets in the same file
- NXdata supports automatic plotting

Why the Hierarchy??

- Supports self description and allows short names in components
- Name, classname pair allows for multiple components of the same type
- NXentry allows for multiple datasets in the same file
- NXdata supports automatic plotting
- Take care once when writing, use n times

Storing Single Data Items

- Units have to specified
- Locating axis, by example
- (Proposed) Taking care of scaled data

NeXus Component and Application Definitions

- Component definitions: dictionaries of allowed field names for the various NeXus groups
- Application Definitions
  - Define what has to be in a NeXus file for a certain application
  - Defines standards
  - Another view: Contract between file producers and users about what has to be in a NeXus file for a well defined purpose
  - Validation by NXvalidate
- Written in NeXus Definition Language, NXDL

Available NeXus Application Definitions

- NXarchive
- NXreftof
- NXtas
- NXtomophase
- NXxnb
- NXtomoproc
- NXmonopd
- NXsas
- NXtofraw
- NXxeuler
- NXxrot
- NXtofsingle
- NXrefscan
- NXscan
- NXtomo
- NXxkappa
- NXiqproc

Application Definition Process

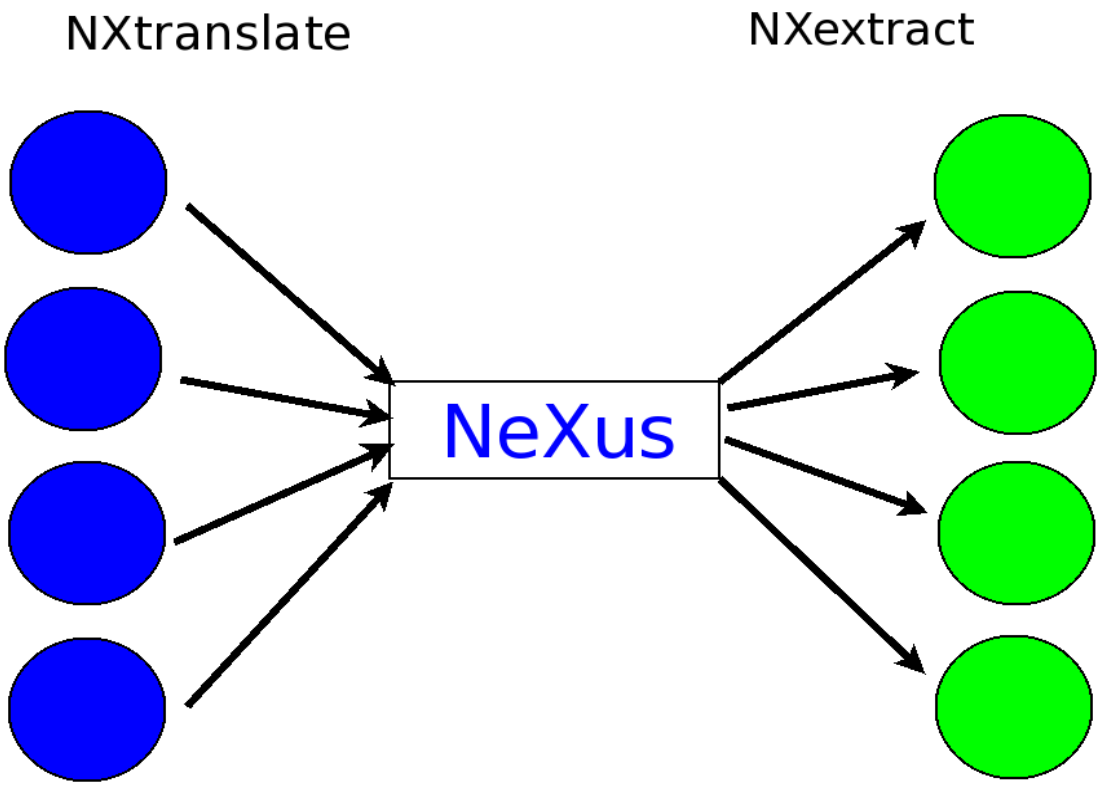
1. Construct an application definition with advice from the NIAC
2. Cure for a year; data should be produced in the new format in this time
3. After curation and review: this is the standard for this application type.

- No promises, but the NIAC may do it for you
  - Description of experiment
  - Minimum set of data items necessary form common use
  - Example data

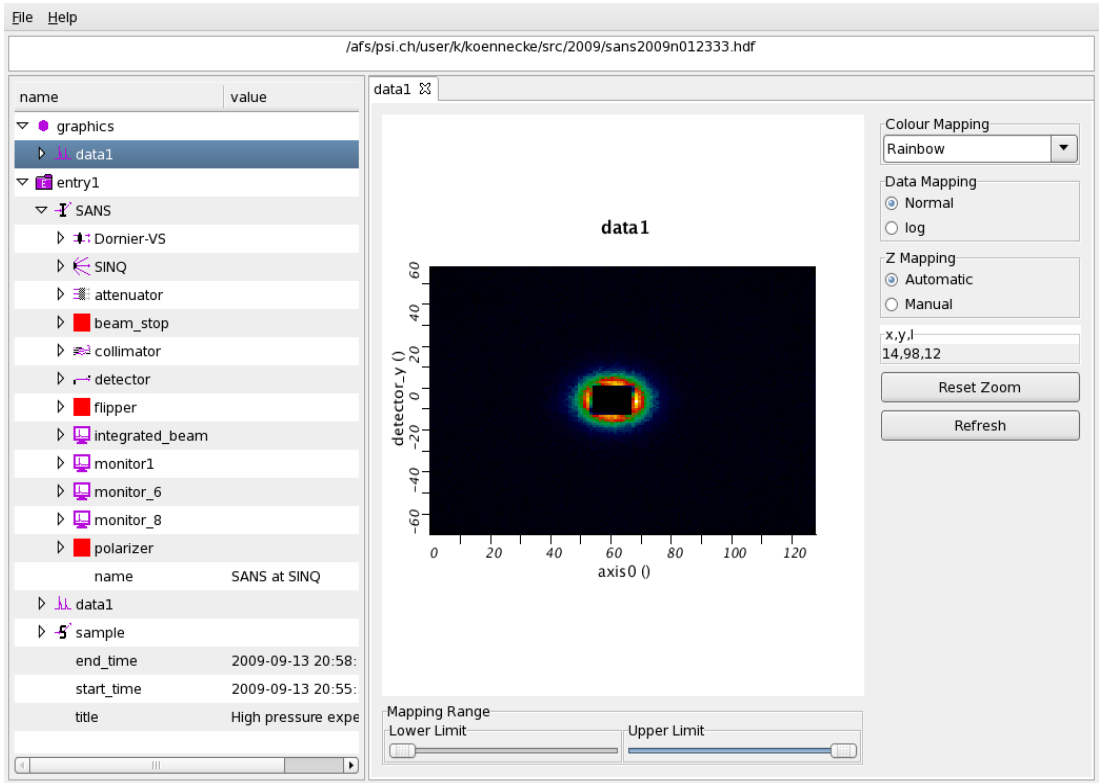
NeXus Tools

nxbrowse	CLI NeXus browser
nxtree	prints NeXus tree
NXmeta	dumps all NeXus meta data
nxtranslate	transforms into NeXus
nxvalid	validates NeXus files
nxextract	converts from neXus to ASCII and binary
nxplot	plots any NeXus file

NeXus Conversions



NXplot


File Help
/afs/psi.ch/user/k/koennecke/src/2009/sans2009n012333.hdf
name
value
graphics
data1
entry1
SANS
Dornier-VS
SINQ
attenuator
beam_stop
collimator
detector
flipper
integrated_beam
monitor1
monitor_6
monitor_8
polarizer
name
SANS at SINQ
data1
sample
end_time
2009-09-13 20:58:
start_time
2009-09-13 20:55:
title
High pressure expe
data1
data1
detector_y ()
axis0 ()
Colour Mapping
Rainbow
Data Mapping
Normal
log
Z Mapping
Automatic
Manual
x,y,I
14,98,12
Reset Zoom
Refresh
Mapping Range
Lower Limit
Upper Limit

Other Systems Using NeXus

- DANSE
- DAVE
- FABLE (ESRF)
- ISAW
- LAMP
- openGenie
- ICAT
- Mantid
- openGDA
- All HDF tools

Data Format Challenges

Challenge 1 in science you are supposed to do new, non standard, things. These of course cannot be easily cast into a standard.

Data Format Challenges

Challenge 1 in science you are supposed to do new, non standard, things. These of course cannot be easily cast into a standard.

Challenge 2 in order to establish a standard a lot of people need to agree

NeXus

Data Format Challenges

Challenge 1 in science you are supposed to do new, non standard, things. These of course cannot be easily cast into a standard.

Challenge 2 in order to establish a standard a lot of people need to agree

Challenge 3 a standard requires scarce scientific programming resources for adoption

Data Format Chances

Chance 1 By using a discoverable data format like NeXus, XML, HDF-5, people can at least figure out what is in the data file.

Data Format Chances

Chance 1 By using a discoverable data format like NeXus, XML, HDF-5, people can at least figure out what is in the data file.

Chance 2 Using predefined names from a dictionary gives meaning to the data in a file.

Data Format Chances

Chance 1 By using a discoverable data format like NeXus, XML, HDF-5, people can at least figure out what is in the data file.

Chance 2 Using predefined names from a dictionary gives meaning to the data in a file.

Chance 3 Using a shared API reduces learning costs and increases application stability.

Data Format Chances

Chance 1 By using a discoverable data format like NeXus, XML, HDF-5, people can at least figure out what is in the data file.

Chance 2 Using predefined names from a dictionary gives meaning to the data in a file.

Chance 3 Using a shared API reduces learning costs and increases application stability.

Chance 4 With NeXus, HDF-5 plus professional programming techniques a DA application can read any file which contains the required data.

Data Format Chances

Chance 1 By using a discoverable data format like NeXus, XML, HDF-5, people can at least figure out what is in the data file.

Chance 2 Using predefined names from a dictionary gives meaning to the data in a file.

Chance 3 Using a shared API reduces learning costs and increases application stability.

Chance 4 With NeXus, HDF-5 plus professional programming techniques a DA application can read any file which contains the required data.

Chance 5 Storing as much data as possible increases the likelihood that the needed data is actually on file, even for unforeseen uses.

Data Format Chances

Chance 1 By using a discoverable data format like NeXus, XML, HDF-5, people can at least figure out what is in the data file.

Chance 2 Using predefined names from a dictionary gives meaning to the data in a file.

Chance 3 Using a shared API reduces learning costs and increases application stability.

Chance 4 With NeXus, HDF-5 plus professional programming techniques a DA application can read any file which contains the required data.

Chance 5 Storing as much data as possible increases the likelihood that the needed data is actually on file, even for unforeseen uses.

Chance 6 In many experiments not the technique but the sample is important: then an application definition simplifies life.

NeXus

Who commits to NeXus?

What You Can Do with NeXus

1. Store and archive data from a wild variety of instruments
2. Store processed data
3. Store a complete workflow from raw data to publication ready data in several NXentries in one file
4. Store a set of related experiments in one file
5. Define strict and validatable standards

Aims of this Workshop

- Disseminate detailed Information about NeXus and NAPI
- Add synchrotron specific data fileds and base classes to NeXus
- Create a synchrotron wishlist for the NIAC
- Create application definitions for synchrotron specific instrumentation
- Review existing application definitions
- http://www.nexusformat.org
- http://lns00.psi.ch/nexus2010