NeXus Recapitulation and Developments

Mark Könnecke

NeXus International Advisory Committee

September 10, 2010

PAUL SCHERRER INSTITUT

PSI

The Predicament of the Traveling Scientist

- A different data format wherever she goes

PAUL SCHERRER INSTITUT

PSI

The Predicament of the Traveling Scientist

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

PAUL SCHERRER INSTITUT

PSI

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

PAUL SCHERRER INSTITUT

PSI

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

PAUL SCHERRER INSTITUT

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

PAUL SCHERRER INSTITUT

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

PAUL SCHERRER INSTITUT

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

PAUL SCHERRER INSTITUT

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

PAUL SCHERRER INSTITUT

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

PAUL SCHERRER INSTITUT

PSI

NeXus API

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

PAUL SCHERRER INSTITUT

NeXus Objects

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

PAUL SCHERRER INSTITUT

PSI

entry:NXentry
- sample:NXsample
- instrument:NXinstrument
  - source:NXsource
  - velocity_selector:NXvelocity_selector
  - detector:NXdetector
    - data[xsize,ysize], signal=1 (1)
- control:NXmonitor
  - data
- data:NXdata
  - link to (1)

PAUL SCHERRER INSTITUT

Why the Hierarchy??

- Supports self description and allows short names in components

PAUL SCHERRER INSTITUT

Why the Hierarchy??

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

PAUL SCHERRER INSTITUT

PSI

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

PAUL SCHERRER INSTITUT

PSI

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

PAUL SCHERRER INSTITUT

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

PAUL SCHERRER INSTITUT

PSI

- Come in all shapes and sizes
- Captured by rules:
  - Store all varied parameters as arrays of length NP at the appropriate place in the NeXus hierarchy
  - For multi detectors, NP, number of scan points is always the first dimension
  - In NXdata: create links to counts and varied variables

PAUL SCHERRER INSTITUT

Scan Example 1: rotating sample

entry:NXentry
- sample:NXsample
  - rotation_angle[NP], axis=1 (1)
- instrument:NXinstrument
  - detector:NXdetector
    - data[NP],signal=1 (2)
- control:NXmonitor
  - data[NP]
- data:NXdata
  - link to (1)
  - link to (2)

PAUL SCHERRER INSTITUT

Scan Example 2: complex scan in Q

entry:NXentry
- sample:NXsample
  - rotation_angle[NP], axis=1 (1)
  - phi[NP], axis=1 (2)
  - chi[NP], axis=1 (3)
  - h[NP], axis=1 (4), primary=1
  - k[NP], axis=1 (5)
  - l[NP], axis=1 (6)
- instrument:NXinstrument
  - detector:NXdetector
    - data[NP],signal=1 (7)
    - polar_angle[NP],signal=1 (8)
- data:NXdata
  - link to (1)
  - link to (2)
  - link to (...)
  - link to (8)

PAUL SCHERRER INSTITUT

Scan Example 3: sample rotation, area detector

entry:NXentry
- sample:NXsample
  - rotation_angle[NP], axis=1 (1)
- instrument:NXinstrument
  - detector:NXdetector
    - data[NP,xsize,ysize],signal=1 (2)
- control:NXmonitor
  - data[NP]
- data:NXdata
  - link to (1)
  - link to (2)

PAUL SCHERRER INSTITUT

Storing Single Data Items

- Units have to specified
- Locating axis, by example
- NEW: Taking care of scaled data

PAUL SCHERRER INSTITUT

PSI

Coordinate Systems

- McStas Coordinate System
- Angle based coordinates
- NXgeomtry for enginnering coordinates and describing shapes

PAUL SCHERRER INSTITUT

PSI

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

PAUL SCHERRER INSTITUT PSI

NEW: Available NeXus Application Definitions

- NXarchive
- NXreftof
- NXtas
- NXtomophase
- NXxnb
- NXtomoproc
- NXindirectof
- NXsastof
- NXtofsingle
- NXmonopd
- NXsas
- NXtofraw
- NXxeuler
- NXxrot
- NXtofsingle
- NXiqproc
- NXsqom
- NXxas
- NXrefscan
- NXscan
- NXtomo
- NXxkappa
- NXiqproc
- NXdirectof
- NXlauetof
- NXtofraw
- NXxasproc

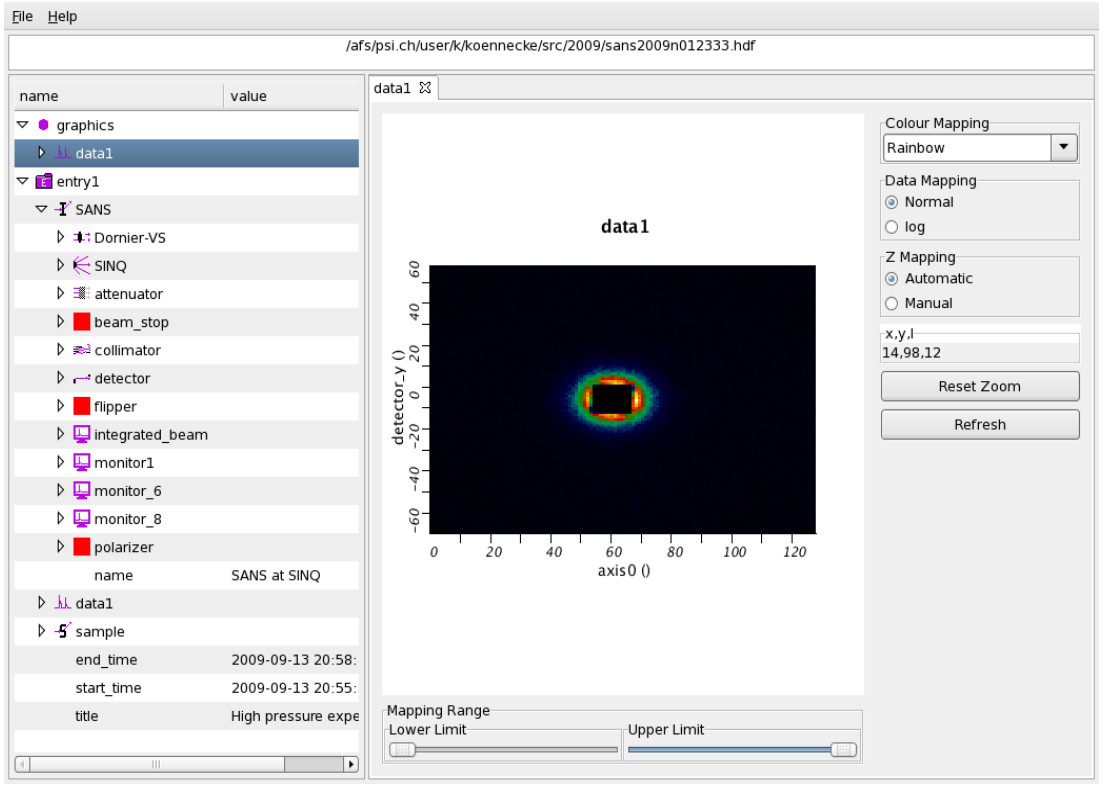
PAUL SCHERRER INSTITUT

NeXus Tools

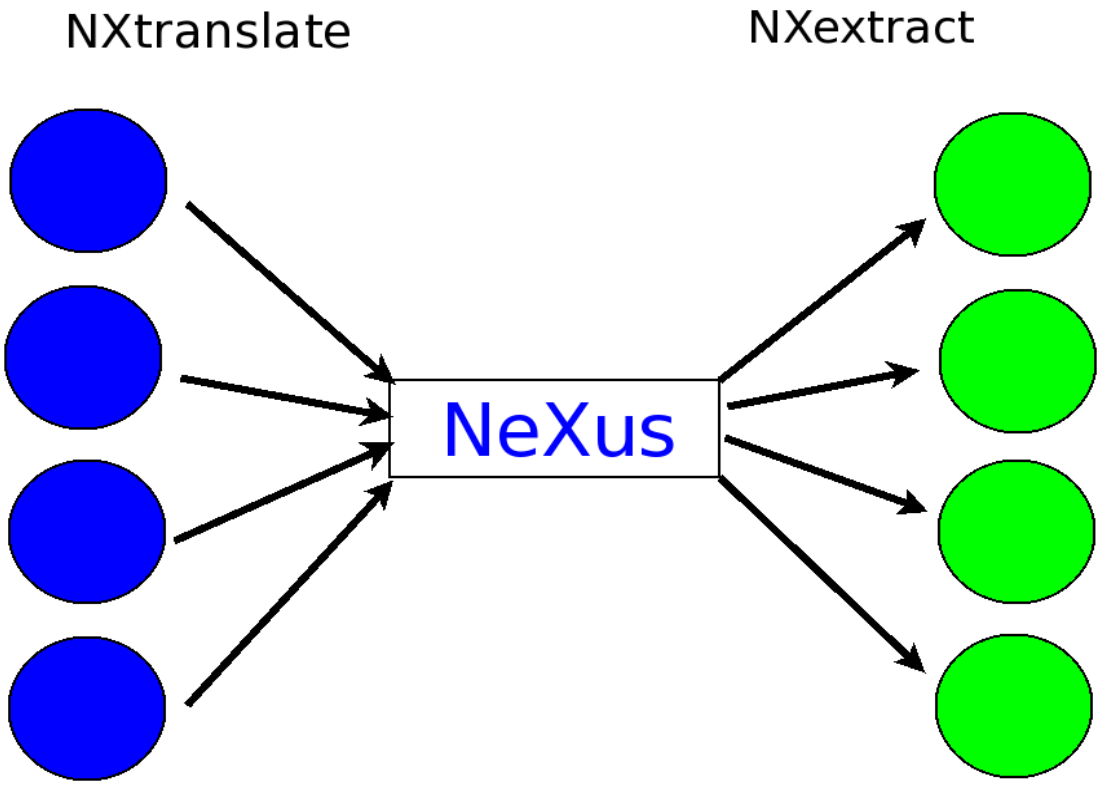
nxbrowse	CLI NeXus browser
nxtree	prints NeXus tree
NXmeta	dumps all NeXus meta data
nxtranslate	transforms into NeXus
nxvalidate	NEW: validates NeXus files
nxextract	converts from neXus to ASCII and binary
nxplot	NEW: plots any NeXus file

PAUL SCHERRER INSTITUT

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

NeXus Conversions



PAUL SCHERRER INSTITUT

Other Systems Using NeXus

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

PAUL SCHERRER INSTITUT

PSI

Who commits to NeXus?

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

PAUL SCHERRER INSTITUT

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

PAUL SCHERRER INSTITUT

PSI

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.

PAUL SCHERRER INSTITUT

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.

PAUL SCHERRER INSTITUT

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.

PAUL SCHERRER INSTITUT

PSI

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.

PAUL SCHERRER INSTITUT

PSI

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.

PAUL SCHERRER INSTITUT

PSI

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.

PAUL SCHERRER INSTITUT

PSI

- PanData
- Workshop: HDF5 as hyperspectral data format, January, ESRF
- NeXus workshop at PSI, May
- Upcoming: Data formats for HDR, DESY, end of october

PAUL SCHERRER INSTITUT

- European initiative for SSO, a shared data file catalog, DA etc
- PanData needs a shared data format in order to make the catalog fly
- Works with NeXus
- Prompted us to have a project plan which we actually executed by now!
- 5.5MM money

PAUL SCHERRER INSTITUT

PSI

HDF5 as hyperspectral data format

- NeXus well received
- Some confusion over a HDF-5 bug in 1.8
- Herbert Berstein conceded that imageCIF is mostly dead
- Demand: make it possible to fully map imageCIF to NeXus
- Missing in NeXus to do full CIF mapping:
  - Scaled data
  - CIF axis specification more accurate

PAUL SCHERRER INSTITUT

- Workshop at PSI
- NeXus seen as HDF-5 with NeXus structures, no interest in API
- Requests:
  - NXsubentry
  - Scaled data
  - Simplified hierarchy for experts
  - Indicate image dimensions
  - Higher level API

PAUL SCHERRER INSTITUT

Topics for this NIAC Meeting

1. Constitution change: NIAC – Tech
2. NXsubentry
3. Scaled data
4. Coordinate systems
5. Next project plan
6. Simplified hierarchy, NXmeasurement

PAUL SCHERRER INSTITUT

- Observation: the NIAC in its current form is inefficient, most of the time most people do not understand or are interested int the ongoing discussion
- Suggestion: Divide into two entities:
  - Full NIAC: votes officers, ratifies project plans, decides general directions
  - Technical subcommittee: decides technical and implementation details. Members selected on merit (contributed work) and approved by full NIAC

PAUL SCHERRER INSTITUT

entry:NXentry
    sas,NXsubentry
        sample:NXsample

        instrument:NXinstrument
            source:NXsource
            velocity_selector:NXvelocity_selector
            detector:NXdetector
                data[xsize,ysize], signal=1 (1)
        control:NXmonitor
            data
        data:NXdata
            link to (1)

PAUL SCHERRER INSTITUT

NXsubentry Rationale

- Multi-method instrument
  - Especially synchrotrons have instruments which combine multiple techniques in one experiment
  - Current NeXus would require seprate NXentries for each technique
  - This becomes unnatural with the additional requirement to store multiple experiments in the same file
  - Combining multiple application definitions in one NXentry would cause name collisions
  - The synchrotron people are aware that NXsubentry requires a lot of links but are willing to do the work
- Add application definition NXsubentries to existing files

PAUL SCHERRER INSTITUT

Scaled Data

- NeXus stronly suggests storing physical values
- But allow scaled data for performance or other reasons
- Implement through additional data attributes

PAUL SCHERRER INSTITUT

Scaled Data Attributes

- transform: This is the indicator that a transformation of the Vraw data is necessary. Transform can have one the following values:
  - offset: Vtrue = Vraw + offset
  - scaling: Vtrue = Vraw * scaling
  - scaling_offset: both an offset and scaling is applied. Vtrue = Vraw*scaling + offset
  - sqrt_scaled: Vtrue = (Vraw/scaling)*(Vraw/scaling)
  - logarithmic_scaled: Vtrue = (Vraw/scaling)**10
  - polynomial: Vtrue = p1 + p2*Vraw + p3*Vraw*Vraw + p4*Vraw*Vraw*Vraw ....
- direction: a komma separated list of length ndim which specifies for each dimension if it is increasing or decreasing
- precedence: a komma separated list of length ndim which gives the rank order in which array indexes change with respect to other indexes.
- coefficients, offset, scaling: parameters for calculating Vtrue

PAUL SCHERRER INSTITUT

entry,NXentry
- measurement,NXmeasurement
  - positions:NXpositioner
    - om
    - two_theta
  - scalars:NXscalar
    - title
    - wavelength
  - images:NXimagedata
    - detector1
    - mca5

PAUL SCHERRER INSTITUT

Ideas for Project Plan 2012

- Refinement of application definitions with communities
- OO base classes?
- Higher level NeXus-APIs?

PAUL SCHERRER INSTITUT

PSI