NeXus Recapitulation and Developments

Mark Könnecke

NeXus International Advisory Committee

October 8, 2010

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

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
- 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

PSI

NeXus Mission

- Definition of a standard data format
  - Rules
  - Validation tools
- Promotion of NeXus
  - Documentation
  - NeXus API
  - Outreach to the scientific community

PAUL SCHERRER INSTITUT

PSI

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

PSI

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

NeXus Processed Data File Structure

entry:NXentry
- sample:NXsample
- processing_name:NXprocess
  - program
  - version
  - parameters:NXparameter
    - raw_file
- data:NXdata
  - data[nx,ny,nz], signal=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

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

- Units have to specified
- Rules for locating axis

- Polar coordinate system: azimuthal_angle,polar_angle, distance
- McStas Coordinate System
- NXgeomtry for enginnering coordinates and describing shapes

PAUL SCHERRER INSTITUT

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

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

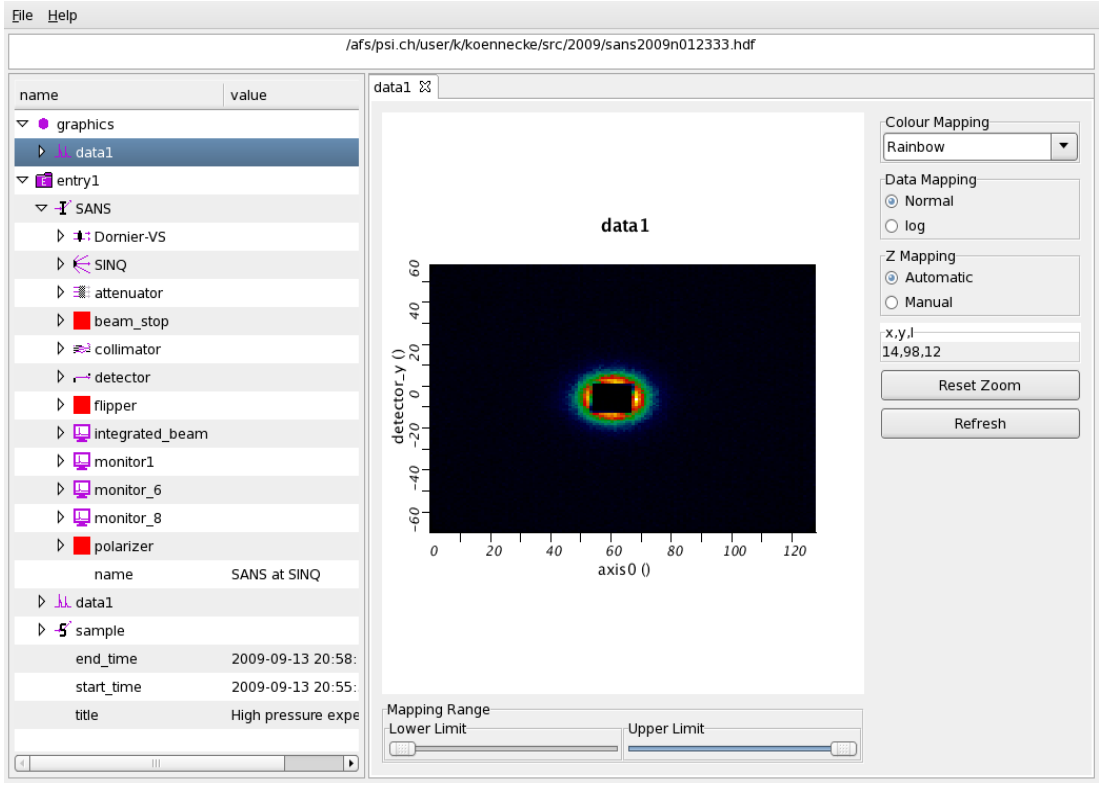
PAUL SCHERRER INSTITUT

NeXus Tools

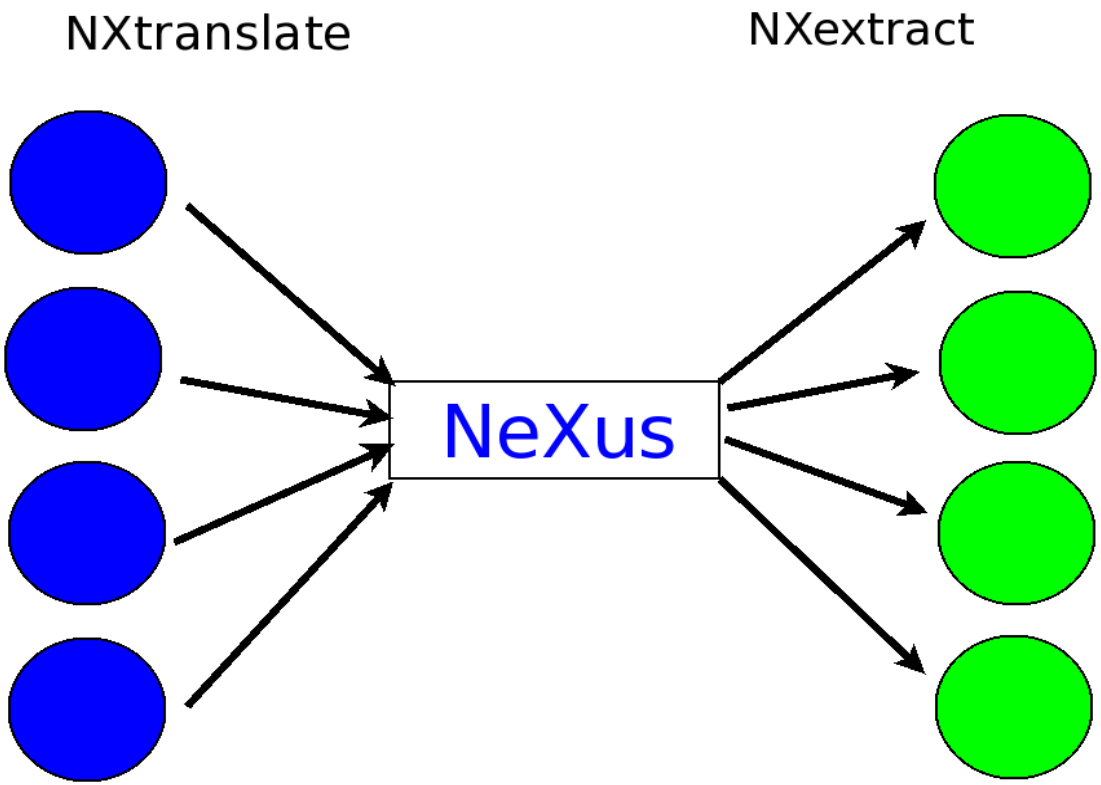
nxbrowse	CLI NeXus browser
nxtree	prints NeXus tree
NXmeta	dumps all NeXus meta data
nxtranslate	transforms into NeXus
nxvalidate	NEW: validates NeXus files
nxextract	converts from neXus to ASCII and binary
nxplot	NEW: plots any NeXus file

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

Who commits to NeXus?

PAUL SCHERRER INSTITUT

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

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.

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 Application Definitions

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

HDF5 as hyperspectral data format

- End of January at ESRF, ca 30 participants
- NeXus well received
- Some confusion over a HDF-5 bug in 1.8
- Demand to map imageCIF fully to NeXus in order to benefit from imageCIF ideas
- Missing in NeXus to do full CIF mapping:
  - Scaled data
  - CIF axis specification more accurate
  - Mapping to database concepts

PAUL SCHERRER INSTITUT

PSI

- Workshop at PSI, 10-12 May
- NeXus seen as HDF-5 with NeXus structures, no interest in API
- Requests:
  - NXsubentry
  - Scaled data
  - Simplified hierarchy for experts
  - Indicate image dimensions
  - Tree based higher level API

PAUL SCHERRER INSTITUT

Threads Versus NeXus

- Planned: NeXus is threadsafe when each thread has its own NXhandle
- A little work needs to be done to arrive there
- BUT: HDF-5 serialises access, no performance gain!
- Parallel HDF5, PHDF, with a different API
- PHDF requires: MPI, MPI-IO, parallel file system
- A new NeXus file driver for PHDF would be required
- Will only be implemented when the community really wants it

PAUL SCHERRER INSTITUT

PSI

Topics for this NIAC Meeting

1. Constitution change: NIAC – Tech, Chairman
2. Choose new officers
3. NXsubentry
4. Scaled data
5. Coordinate systems
6. Next project plan
7. Simplified hierarchy, NXmeasurement
8. Event data
9. muSR NeXus

PAUL SCHERRER INSTITUT

- Observation: we make better progress when working in a smaller group of experts
- Suggestion: Divide NIAC into two entities:
  - Full NIAC: votes officers, ratifies project plans, decides general directions
  - Technical subcommittee: decides technical and implementation details, to be reviewed by full NIAC. Members are selected on merit (contributed work) and approved by full NIAC
- Full NIAC meets only any 2 years: requires extending the terms of officers

PAUL SCHERRER INSTITUT

NXsubentry Rationale

- Suggestion: add NXsubentry group below NXentry with the same structure as NXentry
- Multi-method instrument
  - Especially synchrotrons have instruments which combine multiple techniques in one experiment
  - Current NeXus would require separate 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 willing to do the many links NXsubentry requires
- Add application definition compliant NXsubentries to existing files

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

Scaled Data

- NeXus stronly suggests storing physical values
- Suggestion: allow scaled data for performance or other reasons
- Implement through additional data attributes
- NeXus stores arrays in C storage order
- Suggestion: allow other orders
- Implement through additional data attributes
- NAPI will not implement the transforms

PAUL SCHERRER INSTITUT

- scaling_method: This is the indicator that a transformation of the Vraw data is necessary. Scaling_method can have one the following values:
  - power: Vtrue = p0 + (Vraw/p1)**p2
  - logarithmic: Vtrue = p0 +p1*log(Vraw*p2)
  - polynomial: Vtrue = p0 + p1*Vraw + p2*Vraw*Vraw + p3*Vraw*Vraw*Vraw ....
  - To be expanded?
  - Or give the formula?
- 0, p1, p2, ... Float attributes as parameters
- scaled_units: units after scaling
- Offset, stride

PAUL SCHERRER INSTITUT

- Another look on coordinate systems
- NeXus:
  - Simple, polar coordinates using angles and distances
  - Absolute McStas coordinates using NXgeometry
- imageCIF
  - Arbitrary axis allowed
  - Contains information to build transformation matrices

PAUL SCHERRER INSTITUT

$$T = \begin{pmatrix} 1 & 0 & 0 & x \\ 0 & 1 & 0 & y \\ 0 & 0 & 1 & z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

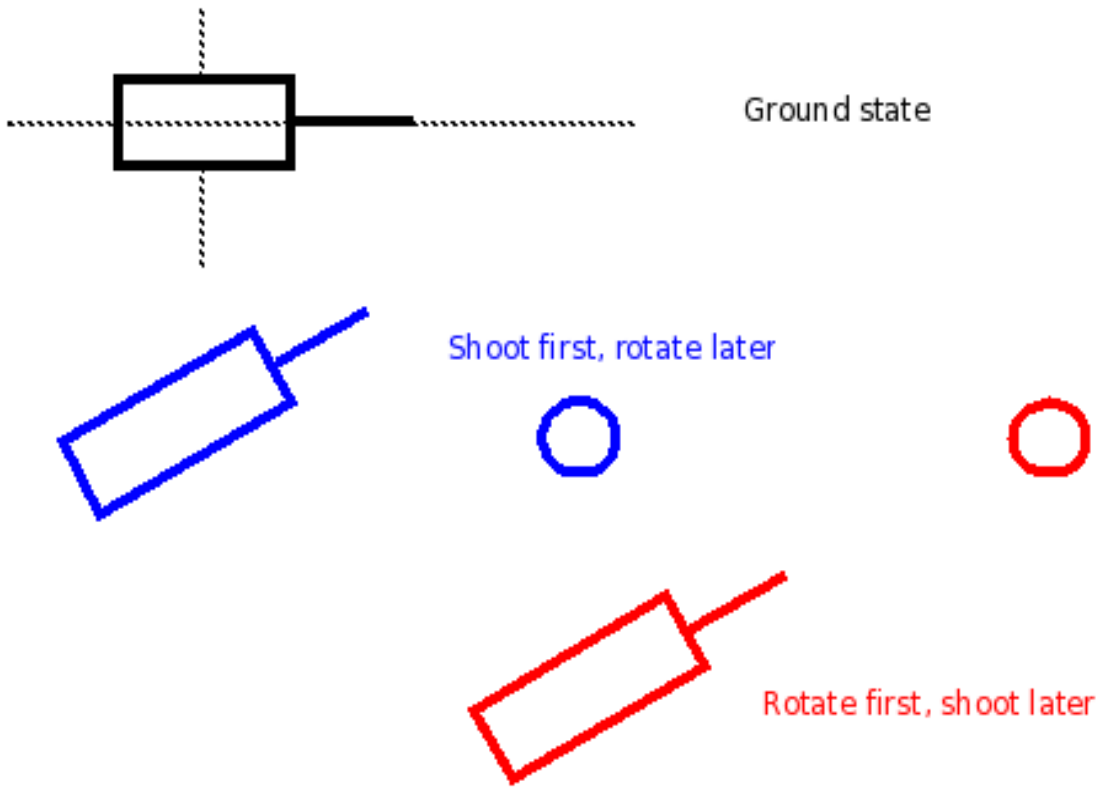
PAUL SCHERRER INSTITUT

$$T = \begin{pmatrix} 1 & 0 & 0 & x \\ 0 & 1 & 0 & y \\ 0 & 0 & 1 & z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$R = \begin{pmatrix} r11 & r12 & r13 & 0 \\ r21 & r22 & r23 & 0 \\ r31 & r32 & r33 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

PAUL SCHERRER INSTITUT

Combining Transformations



PAUL SCHERRER INSTITUT

Some Properties

- Transformations can be combined by matrix multiplications
- Individual matrices can be derived by looking at the situation when everything else is 0
- Absolute positions can be obtained by multiplying the resulting matrix with its transpose
- Defines new coordinate systems at components
- CIF contains a duplication: vector, offset scheme

PAUL SCHERRER INSTITUT

What Use Is This?

- Allows to calculate absolute positions of components in the laboratory coordinate systems
- Can directly convert from a detector coordinate system to vectors in Lab coordinate system
- Calculate things like impact of primary beam on detector, SAS
- Allows arbitray axis to be expressed
- Intuitively describe an instrument with angles and translations and still be able to recover absolute coordinates

PAUL SCHERRER INSTITUT

Information Required

type	rotation or translation
direction	vector around which rotated or translated
value	The angle of rotation or the length of translation
dependency	The order of operations to place a component

PAUL SCHERRER INSTITUT

Use For NeXus

- Use to document existing axis and polar coordinate system better
- Permits arbitrary axis to be defined
- Allows construction of transformation matrices and gain the advantages of using them
- Allows to express an instrument intuitively as a sequence of translations and rotations AND be able to reconstruct absolute positions
- The objective is to allow a full mapping from imageCIF to NeXus and back

PAUL SCHERRER INSTITUT

NeXus Axis Mapped

- rotation_angle, polar_angle, rotate 0 1 0
- azimuthal_angle, rotate 0 0 1
- distance, translate 0 0 1
- chi, rotate 0 0 1
- phi rotate, 0 1 0
- NeXus polar coordinate system: rotate azimuthal_angle, rotate polar_angle, translate by distance

PAUL SCHERRER INSTITUT

Axis Suggestions for NeXus

- NeXus stays with the McStas Laboratory coordinate system
- NeXus strongly encourages to use the named and documented NeXus axis
- Allow attributes type, direction in order to support arbitrary axis
- Add aequatorial_angle as a name to appear in base classes for rotation 1 0 0.
- Add y_translation (translate 0 1 0) and x_translation (translate 1 0 0) to base classes.

PAUL SCHERRER INSTITUT

PSI

Expressing Axis Dependency in NeXus

- Implied: use existing NeXus coordinate system
- dependson attribute pointing to depending axis
- transform field in base classes which becomes a comma separated list of the path to the transformations required to position this component
- Create a special container to hold axis dependencies, NXdependency, to collect the dependencies in one place for easy access. This is what CIF does

PAUL SCHERRER INSTITUT

```
sample,NXsample
        rotation_angle
        chi (dependson rotation_angle)
        phi (dependson phi)
```

```
sample,NXsample
        rotation_angle
        chi
        phi
        transform = rotation_angle,chi,phi
```

Separate Group Option

```
sample,NXsample
        rotation_angle
        chi
        phi
dependency,NXdependency
        sample/chi =
                sample/rotation_angle
        sample/phi =
                sample/chi
        instrument/detector/x_translation =
                instrument/detector/distance
        instrument/detector/distance =
                instrument/detector/polar_angle
```

PAUL SCHERRER INSTITUT

PSI

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

PAUL SCHERRER INSTITUT

Ideas for Project Plan 2012

- Refinement of application definitions with communities
  - Collect community feedback
  - Release of NXDL, base classes and application definitions 1.0
- nxvalidate 1.0
- Overhaul of documentation: Manual, NAPI, WWW-site
- OO base classes?
- Higher level NeXus-APIs?

PAUL SCHERRER INSTITUT