NIAC Quo Vadis

Mark Könnecke

NeXus International Advisory Committee

September 20, 2012

PAUL SCHERRER INSTITUT

The Future of NeXus

- In existence since 1996
- Uptake is very slooooooowwwwww
- Things have changed since 1996
- Do we leave NeXus as is?
- Or do we change it based on our experiences, feedback?

PAUL SCHERRER INSTITUT

- New things get done with NeXus
- Problem: few new things get done
- Problem: use of NeXus varies even across NeXus facilities
- BUT: whenever other people try to devise a data format the result looks much like NeXus: HDRI, tomography group, SAS

PAUL SCHERRER INSTITUT

PSI

Questionaire Feedback

- From things said to us and the questionaire
- We do not want NAPI: we want HDF-5
- Too complicated
  - NAPI
  - File hierarchy
- Users do not see benefits
- Bad documentation
- Old formats as a barrier
- Why change a running system?
- Slow I/O
- Lack of resources

PAUL SCHERRER INSTITUT

PSI

NeXus Usage

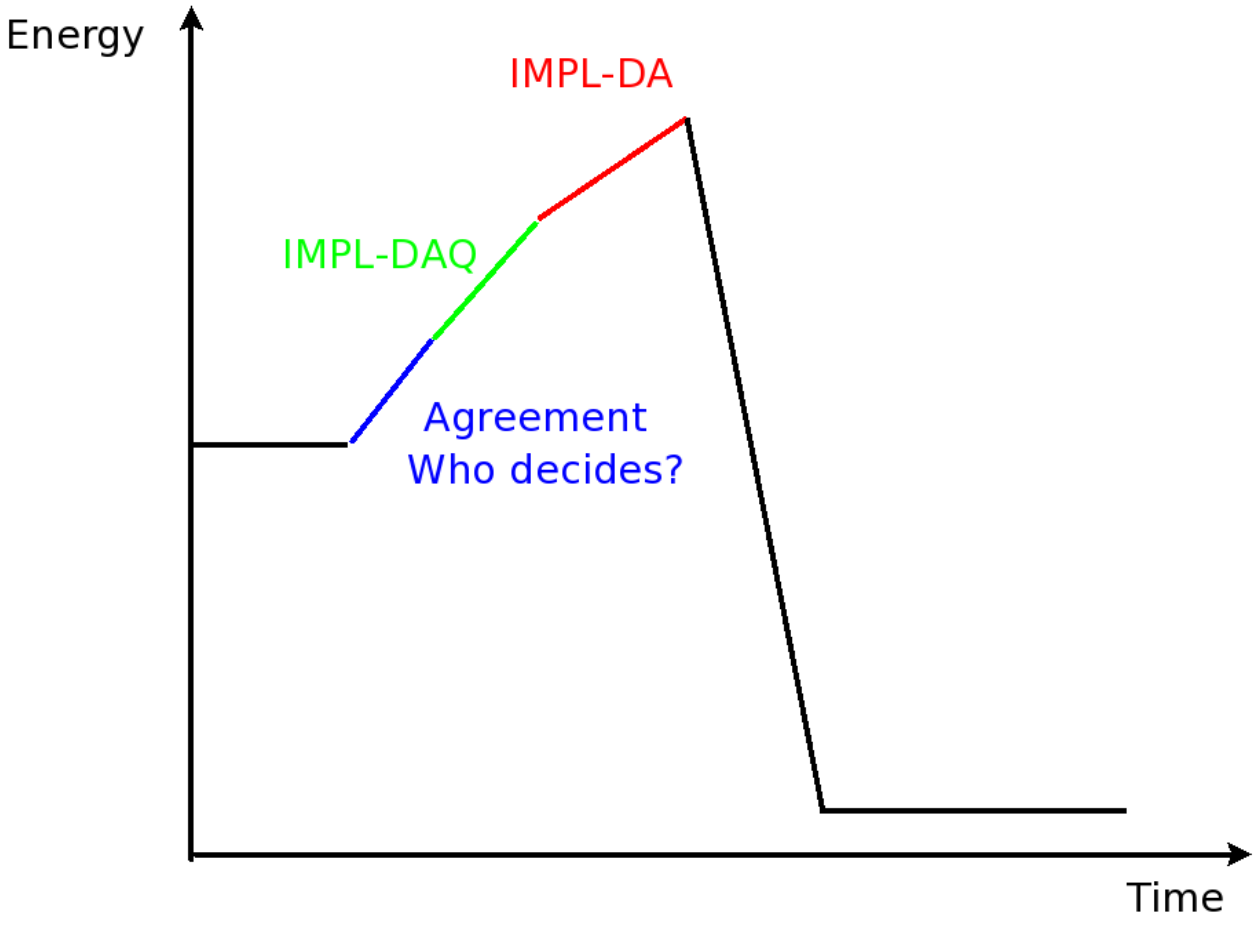
- Soleil: 20 out of 26 instruments do NeXus, 2 Mill Files
- PSI-SINQ: 11 from 16 instrument son NeXus, 1.4 Mill files
- PSI-SLS: 0, 2 planned,
- SLS: none yet, planning to
- KEK: 10, 6 planned
- ANSTO: 7 going to 10
- ESRF: 2 beamlines, limited to NXentry, NXcollection, NXdata, moving to 4
- HZB: 3N, 3 planned
- FRM 2: 0, moving to 1
- SNS: 14,3 in the pipeline
- DESY: 0, 11 in 2 Jahren
- Diamond: 7 NeXus only, 17 writing, moving to 18 as primary format
- ISIS: 8 using, 20 writing, planned: 20 using
- Muons: 4 instruments, reluctance to move on
- Lujan/LANL: 11 instruments, no change

NIAC Tasks

- Need to find something only NeXus can do
- ESRF would like a simpler NeXus: NeXus ultralight or NXexchange
- ESRF: NIAC as a custodian
- Collect examples
- Work better with SW groups at facilities in construction
- Work for reduced data interchange
- Provide NeXus tools as components to be integrated into other SW
- Try to integrate stronger with communities

PAUL SCHERRER INSTITUT

NeXus Energy Curve



PAUL SCHERRER INSTITUT

Reaching Agreement

- No scientific body to decide upon data formats
- Little interest from scientist in computing and data formats
  - MUST produce papers, papers, papers....
- Data format problems solved by computing slaves or PHD students
  - People with little political clout in facility hacking order

PAUL SCHERRER INSTITUT

PSI

Data Ownership Issues

- Producing facility generates and owns data
  - May need to do non standard things in order to optimize writing
  - My define a facility standard format to save man power
- Scientific communities own data analysis
  - They decide what they need

PAUL SCHERRER INSTITUT

PSI

Implementation Issues

- Facilities minimise computing resources
- Competition for scarce scientific computing resources
- More sexy then data formats:
  - Get that screwed instrument to work
  - Implement that new data analysis algorithm
  - Advice dumber and dumber users

PAUL SCHERRER INSTITUT

Conflicting Aims in NeXus

- Full NeXus organises a complete beamline description with possibly hundreds of parameters
  - Hierarchy a good thing
- An application definition ownly contains less then 30 parameters
  - Hierarchy looks like overkill
- Features like workflow in file, multiple experiments rarely used

PAUL SCHERRER INSTITUT

PSI

Dictionary Based Programming Techniques

- Soleil and ANSTO: Common Data Modell (CDM)

1. Use: H5Oopen(fid,"/entry/detector/data",NULL,NULL)
2. Externalize path strings into separate files

- Consequences:
  - By editing dictionary any file containing the data can be read
  - Fixed paths and names no longer important

PAUL SCHERRER INSTITUT

- Read in whole file hierarchy
- On demand loading for large arrays
- Then pick your data nuts and raisins from the tree

PAUL SCHERRER INSTITUT

Are Data Format Requirements Changing?

- 1996, diskspace expensive: snapshot of instrument was what was possible
- 2012, diskspace shit cheap:
  - Event mode for neutrons
  - On the fly scans at synchrotrons
  - FELS experiments log only
  - Full logging of experiment becomes possible (and desirable)
- Demand for more time stamped data
- Data files would be extracted from logs as interface to legacy DA software

PAUL SCHERRER INSTITUT

PSI

Value in NeXus

- The dictionary of names and terms
  - Needs better documentation and definition though
  - It does not matter where things are in the file but meaning is still important
- Application definitions as starting points for method specific standards
- Standard validation tools
- NeXus Rules
  - Scan storage rules
  - Axis assignment rules
  - Coordinate system rules

PAUL SCHERRER INSTITUT

NeXus Options

- Hold on tight!
- Simplified NeXus
- NIAC as a custodian
- Give up
- Your option here

PAUL SCHERRER INSTITUT

PSI

- Make hierarchy a recommendation
- Ultralight:
  - Use HDF-5
  - Use dictionary based programming
- Light:
  - Use NXentry and NXsubentry
  - Possibly flatten NXsubentry hierarchy
  - Collect data items for exchange
- Medium:
  - Flatten NXinstrument
  - Limit us to one way to define orientation and position of components

PAUL SCHERRER INSTITUT

NIAC as Custodian

- NIAC as forum for discussing data formats
- NIAC stores, documents and maintains community developed formats
- NIAC maintains NeXus dictionaries

PAUL SCHERRER INSTITUT

Questions to be answered

- Where do we want to go with NeXus?
- Do we reinvent NeXus?
- Do we provide more for time stamped data?
- Which role do you envisage for the NIAC?
- Do we invest in OO-NeXus?
- What do we do with NAPI?
- Opinions???

PAUL SCHERRER INSTITUT

PSI