# The State of NeXus

Tobias Richter

Diamond Light Source and NeXus International Advisory Committee

NOBUGS 2014

## What this talk is not about

There is no report from:

**NeXus Committee** no full meeting since the last NOBUGS
**Code Camp** no Code Camp happened either

(Both will happen this December at APS.)

# NeXus – what's in the box?

- data format aimed at all techniques in x-ray, neutron and muon science
- defining a data structure inside a container format
  (we strongly prefer HDF5)
- 300 page manual
- a library (API) and some tools for reading and writing
- two committees
- website, wiki, github, mailing lists
- logo (debatable quality)

Why cannot not be so easy?

## Changing Requirements

Measurements become routine.

Users want to focus on their science domain.

Data rates explode.
Writing the first PB at DLS took 7 years. Writing the second took 5 months.

Science continually develops.
Techniques change. Things need to be recorded differently.

# Changing Requirements? No Problem!

There is a lot of flexibility in NeXus!

Write what you want!
By default hardly anything is required in NeXus.
And you can add to that whatever you like.
*Is that true? How does that work?*

## base classes



*dictionary*

device level

pick what you like

## application definitions



*form*

file (entry) level

all fields required

What is actually used in practice?

Disclaimer: The survey is meant to provide a rough overview and does not follow strict scientific protocol.
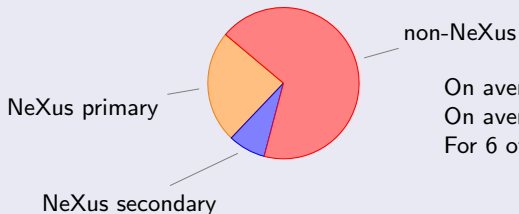
Results are presented anonymised.

# Survey Results

Facilities with a total of 360 beamlines care enough about NeXus to answer the survey.

## Use of NeXus



non-NeXus

NeXus primary

NeXus secondary

On average 24 beamlines per institution.
On average 8 of these write NeXus.
For 6 of these it is the primary data format.

## Facilities Plans

- Write NeXus at more end stations.
- Move away from NeXus library to a more direct HDF5 route
  (currently almost 2/3 use it for writing, 1/3 for reading)
- Work on better data analysis software and support for reading NeXus files.
- Implement new CIF geometry system.
- Develop ASCII-based storage backends.
- Create more application definitions.

# Survey Results

## Application Definitions

6 unique ones in reported use. 2 of them at more than one facility.

There are 30 documented application definitions in the manual.

## Number of Datasets

6 facilities with 14 NeXus stations *report* writing only a single dataset to the file.
That is a single number, spectrum or image, with no axis or other metadata.
At 3 of those facilities have "NeXus" as their primary output.

The median number of datasets in the remaining answers is 60.

# What does that mean?

NeXus is over the initial acceptance barrier.
Many facilities use it, mostly for real.

Some seem to fall for the dark side of mere formal compliance.

We need to aim for more convergence.

Undeclared content makes automatic processing hard.

# Making NeXus files more predictable and portable

## Definition

Options:

- Get facilities to increase the use of application definitions by peer pressure or magic.
- Create an inheritance model for definitions
- Create an alternative more light weight modular composition system

## Validation

Existing validation relies on formal types and naming.

With more flexible file contents a functional validation is required, checking objects and their dependencies.

# Getting Involved

`http://www.nexusformat.org/`

## Mailing List

Subscribe or browse the archive: `http://lists.nexusformat.org/mailman/listinfo/nexus`

## Github

`https://github.com/nexusformat`

- Fork:
    - Library and Tools
    - Definitions
    - Documentation
- Raise Issues
- Make Pull Requests

## Teleconferences (NeXus-Tech)

Runs roughly once a fortnight on
Google Hangout.

## NIAC Meetings (Committee)

2014-12-15 – 2014-12-19 @ APS
`http://wiki.nexusformat.org/NIAC2014`
Contact me or the mailing list
if your facility has no representative.

We have a lot to show: user base, API, definitions of file structure, manual, etc.

NeXus is not a magic bullet.
But done properly NeXus is better than yet another home grown file format.

We need to do some work to make the average NeXus file more mobile and the standard easier to follow.

Get involved in that process!

Take home message:

Make sure you write the best NeXus files you can at the moment!