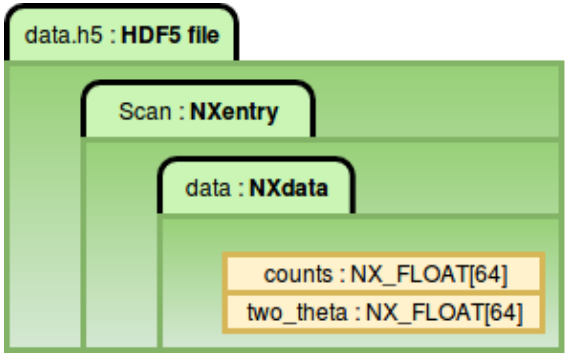
NeXus scientific data format: Introduction and applications



Pete R. Jemian, group leader
Beamline Controls & Data Acquisition Group
Advanced Photon Source, USA

www.nexusformat.org
github.org/nexusformat

Könnecke *et al.*, *J Appl Cryst* 48.1 (2015) 301–305, 10.1107/S1600576714027575

2017-08-25

This work was supported in part by the U.S. Department of Energy, Office of Science, under Contract No. DE-AC-02-06CH11357.

A few things to know about NeXus

Uses HDF5 as file format: https://www.hdfgroup.org

- NeXus is a common data format for neutron, x-ray, and muon science.
- HDF5 is a hierarchical format:
  - components: groups, datasets, attributes, links
- NeXus
  - Defines the layout of information in an HDF5 file
  - Uses the HDF5 group structure to organize data
  - Provides a dictionary of common terms
  - NeXus *application definitions* can further define terms and layout specific to a scientific community for the exchange of data

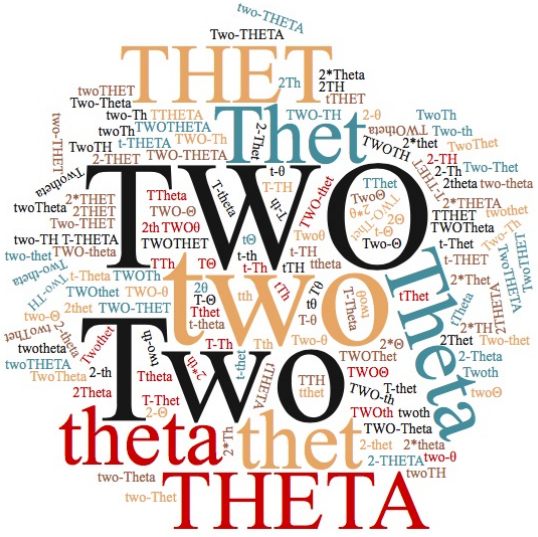
Basic NeXus structure

```
NeXus HDF5 data file
  NXentry
    NXdata
      # datasets
    NXinstrument  # optional
      # raw data structures
    NXsample # optional
      # sample metadata
    # other metadata
```

J. Appl. Cryst. (2015). **48**, 301-305
https://doi.org/10.1107/S1600576714027575

Overview of this talk

- Introduction to NeXus
  - Components
  - Design Principles
  - Data Storage Objects
  - Programs that read / write NeXus files
- Examples
- NeXus International Advisory Committee (NIAC)
- Communications



How many variations of this common term?

Introduction

NeXus

- NeXus is an effort by an international group of scientists
- Started in 1995
- common data exchange format for neutron, X-ray, and muon experiments
- built on top of the scientific data format HDF5 (hdfgroup.org)
- Purposes
  - raw data : any relevant data associated with a scientific instrument or beamline
  - processed data : structures for exchange of data between applications

The Components of NeXus

- Design Principles
  - group, field, attribute, link
- Data Storage Objects
  - base classes, application definitions, and contributed definitions
- Subroutines
  - NAPI, Python (and other) examples
- Scientific Community
  - you

NeXus Design Principles

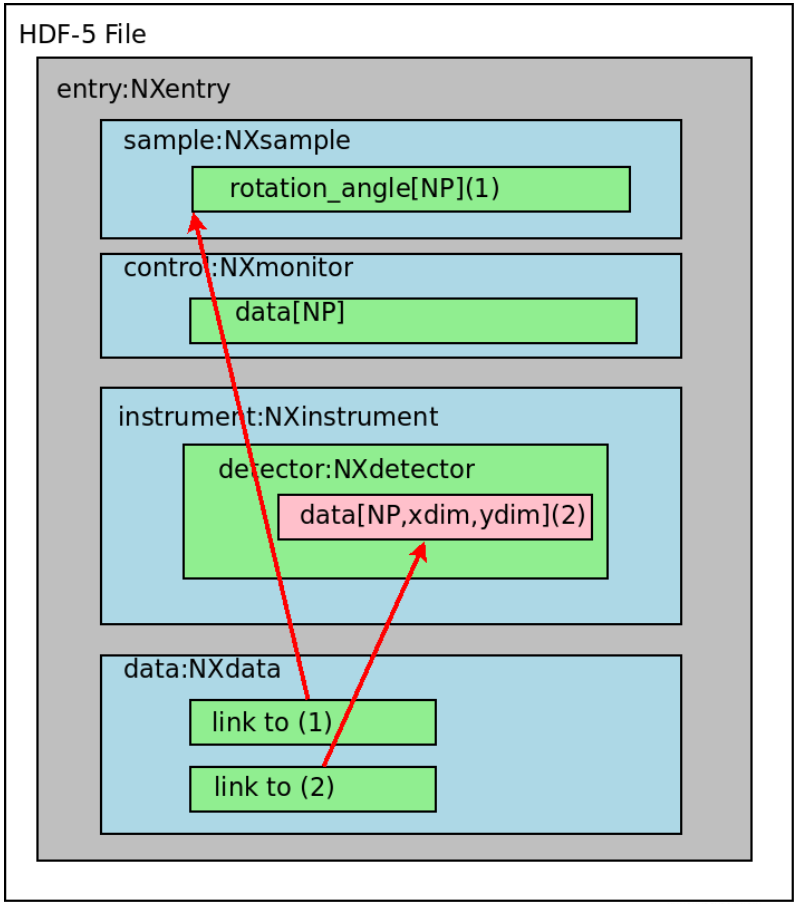
Underlying HDF5 Data Model in each class

- Group
  - subdirectory
- Field
  - dataset
- Attribute
  - describes a group or field
- Link
  - a pointer

```
entry:NXentry
  instrument:NXinstrument
    detector:NXdetector
      data:[]
        @long_name = "strip detector 1-D array"
      bins:[0, 1, 2, ... 1023]
        @long_name = "bin index numbers"
  sample:NXsample
    name = "zeolite"
  data:NXdata
    @signal = "data"
    @axes = ["bins", "bins"]
    @bins_indices = [0, 1]
    data --> /entry/instrument/detector/data
    bins --> /entry/instrument/detector/bins
```

NeXus Link
like a pointer

- Avoids writing same data many times
- Use as needed to build NXdata (or other)
- Implemented as HDF5 Hard Link
  - @target attribute added to identify “original” location
- HDF5 External file links allowed
  - *e.g.*, high-frame rate area detectors
  - Need to coordinate these files



NeXus Data Storage Objects

Describes how the groups, fields, attributes, and links are to be used

base classes (54)

- A set of components that are used to construct a data file
- Defines the set of terms that *might* be used in an instance of that class

application definitions (34)

- Specifies a data structure for a given domain, such as small-angle scattering
- Defines the minimum set of terms that must be used

contributed definitions (10)

- Propositions from the community for base classes or application definitions
- Other NXDL files for long-term archival by NeXus
- Considered as either in incubation or a special case not for general use

NeXus base classes

Different parts of a data file

- Set of terms that ***might*** be used in an instance of that class
- Can add additional terms
- Almost everything is optional
- Should cover most possible use cases
- Defines the spelling and meaning of such terms
- No single predefined place for *all* possible data
- Expect community contributions

NeXus

A few of the 54 base classes

NXcrystal

A crystal monochromator or analyzer.

NXdata

NXdata describes the plottable data and related dimension scales.

NXdetector

A detector, detector bank, or multidetector.

NXdetector_group

Logical grouping of detector elements.

NXdetector_module

Geometry and logical description of a detector module.

NXdisk_chopper

A device blocking the beam in a temporal periodic pattern.

NXentry

(**required**) NXentry describes the measurement.

NXenvironment

Parameters for controlling external conditions

application definitions

Interface with analysis or ...

- Minimum set of terms that ***must*** be used in an instance of that class
- Can add additional terms
- Definition declares what is optional
- Should cover most possible use cases
- Defines the spelling and meaning of such terms
- No single predefined place for *all* possible data
- Expect community contributions

A few of the 34 application def.

NeXus

NXiqproc

Application definition for any $I(Q)$ data.

NXlauetof

This is the application definition for a TOF laue diffractomet

NXmonopd

Monochromatic Neutron and X-Ray Powder diffractometer

NXmx

functional application definition for macromolecular crystallography

NXrefscan

This is an application definition for a monochromatic scanning reflectometer.

NXreftof

This is an application definition for raw data from a TOF reflectometer.

NXsas

raw, monochromatic 2-D SAS data with an area detector

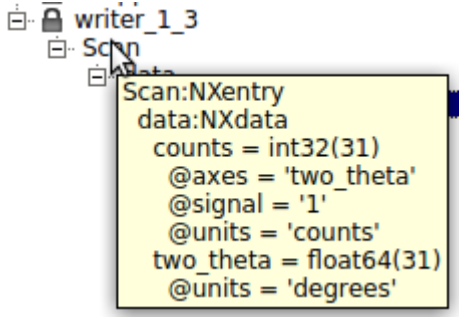
A very simple NXDL file

Each NeXus class is written in NeXus Definition Language (NXDL)

```
<?xml version="1.0" ?>
<definition
  xmlns="http://definition.nexusformat.org/nxdl/3.1"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://definition.nexusformat.org/nxdl/3.1 ../nxdl.xsd"
  category="base"
  name="verysimple"
  version="1.0"
  type="group" extends="NXobject">

  <doc>
    A very simple NeXus NXDL file
  </doc>
  <group type="NXentry">
    <group type="NXdata">
      <field name="counts" type="NX_INT"  units="NX_UNITLESS">
        <doc>counts recorded by detector</doc>
      </field>
      <field name="two_theta" type="NX_FLOAT" units="NX_ANGLE">
        <doc>rotation angle of detector arm</doc>
      </field>
    </group>
  </group>
</definition>
```

- NXDL file is XML
- Defines dictionary
  - Data type
  - Documentation
- Describes how data are related
- Marks data if item is required
- Each NXDL is versioned
- Governed by a Schema
- XML can be validated
- Continuous integration used to ensure integrity

Data Units



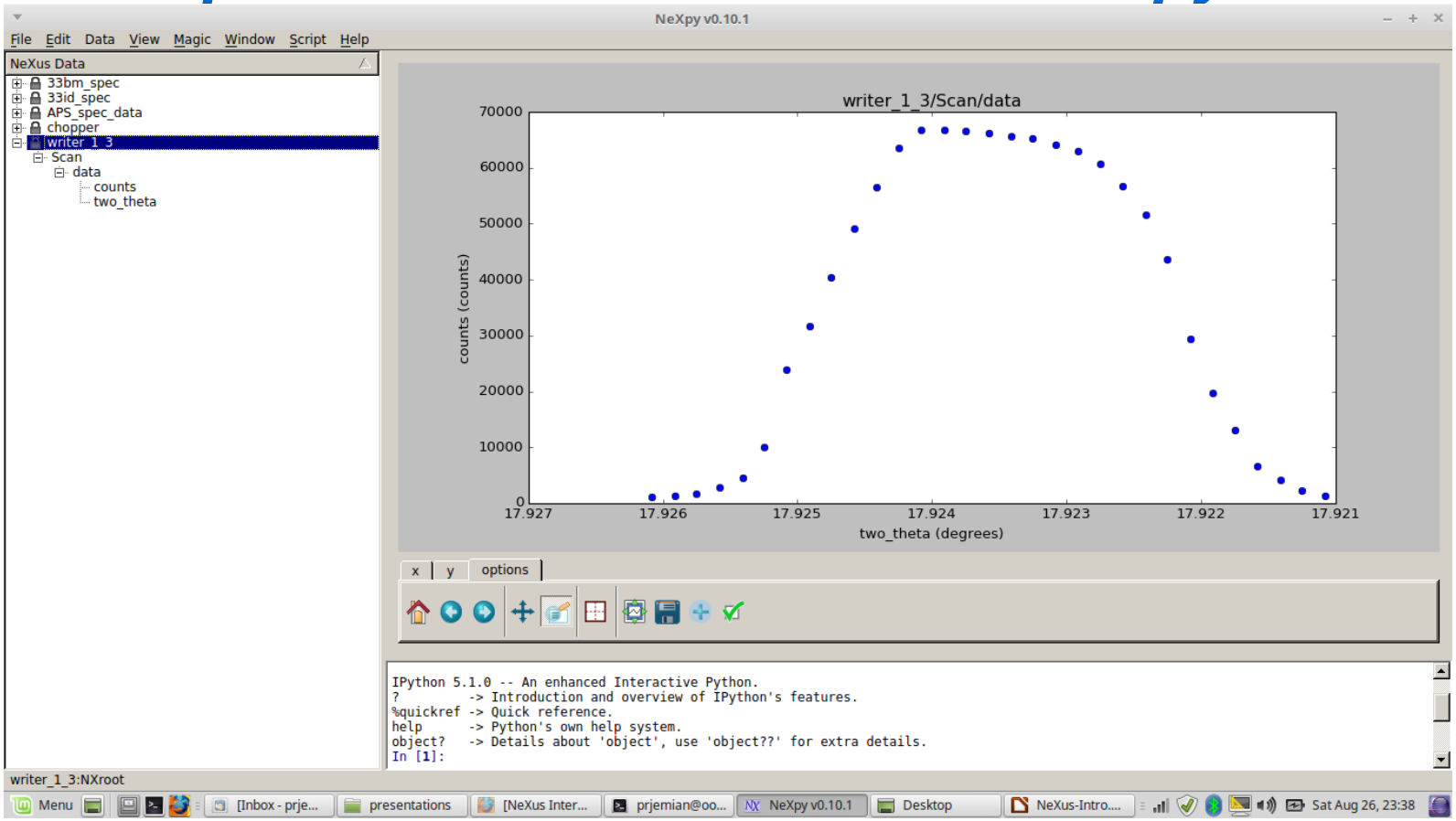
- NeXus uses the UDUNITS standard (http://www.unidata.ucar.edu/software/udunits)
  - Very flexible
  - Allows mathematical expressions such as g/cm^3
  - Supported by C library for programmatic conversion (also Python)
- Units appropriate to field: wavelength has NX_WAVELENGTH type units
- Units are not presently validated
  - But, there are plans to check them using the C library code

Programs that read & write NeXus files

see: http://download.nexusformat.org/doc/html/utilities.html

- Anything that can read or write HDF5 files
- Data Analysis applications:
  - At least 14 (including IDL, Matlab, PyMCA, and NeXpy)
- Tools from the HDF Group (hdfgroup.org): HDFview, h5dump, …
- Python package: h5py
- NeXus command-line utilities (from https://github.com/nexusformat/code)
- Validation of NeXus files: cnxvalidate & punx

Example: 1-D data shown with NeXpy



NeXus raw data file structure

Standard dictionary to describe your instrument

- NXentry: data belonging to one run or scan
- NXinstrument: data that describes the instrument
- NXmonitor: incident intensity
- NXsample: all about the sample
- NXuser: all about the sample
- NXdata: (typically) links to plottable data

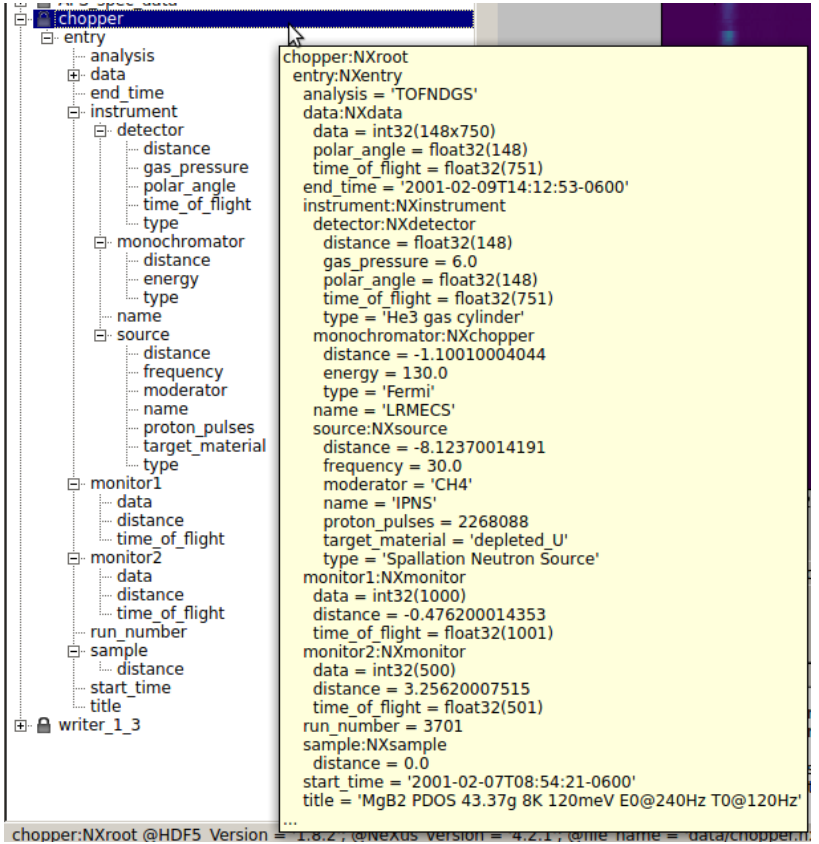
```
HDF5 file
  NXentry
    NXinstrument
      NXsource
      NXcollimator
      NXattenuator
      NXdetector
      ...
    NXsample
    NXmonitor
    NXuser
    NXdata
  NXentry
```

- Only NXentry and NXdata are required

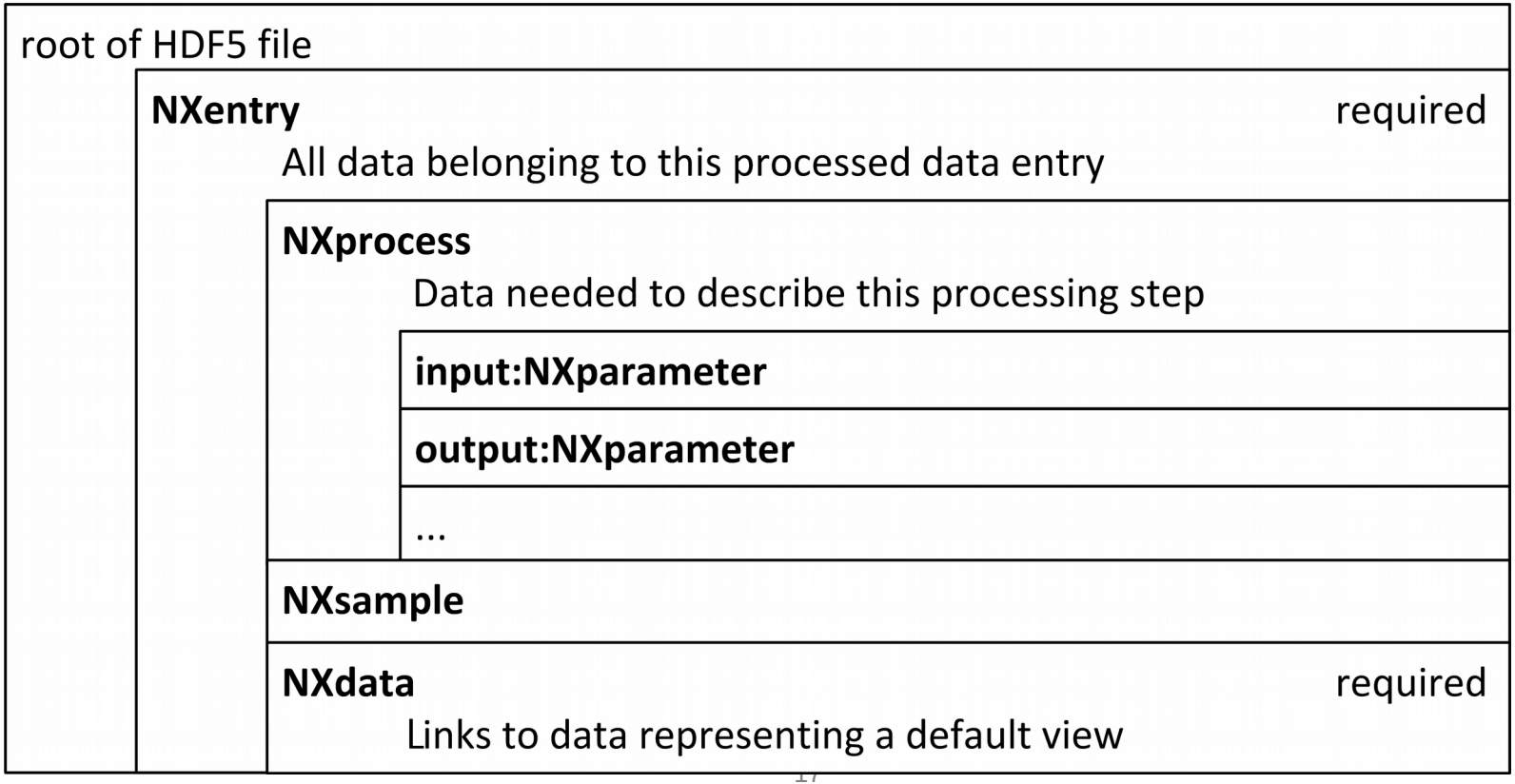
Example raw data tree

LRMECS, IPNS, ANL: Chopper Spectrometer Pulsed Neutron Source

- Data collected in 2001
- For clarity here, no attributes shown

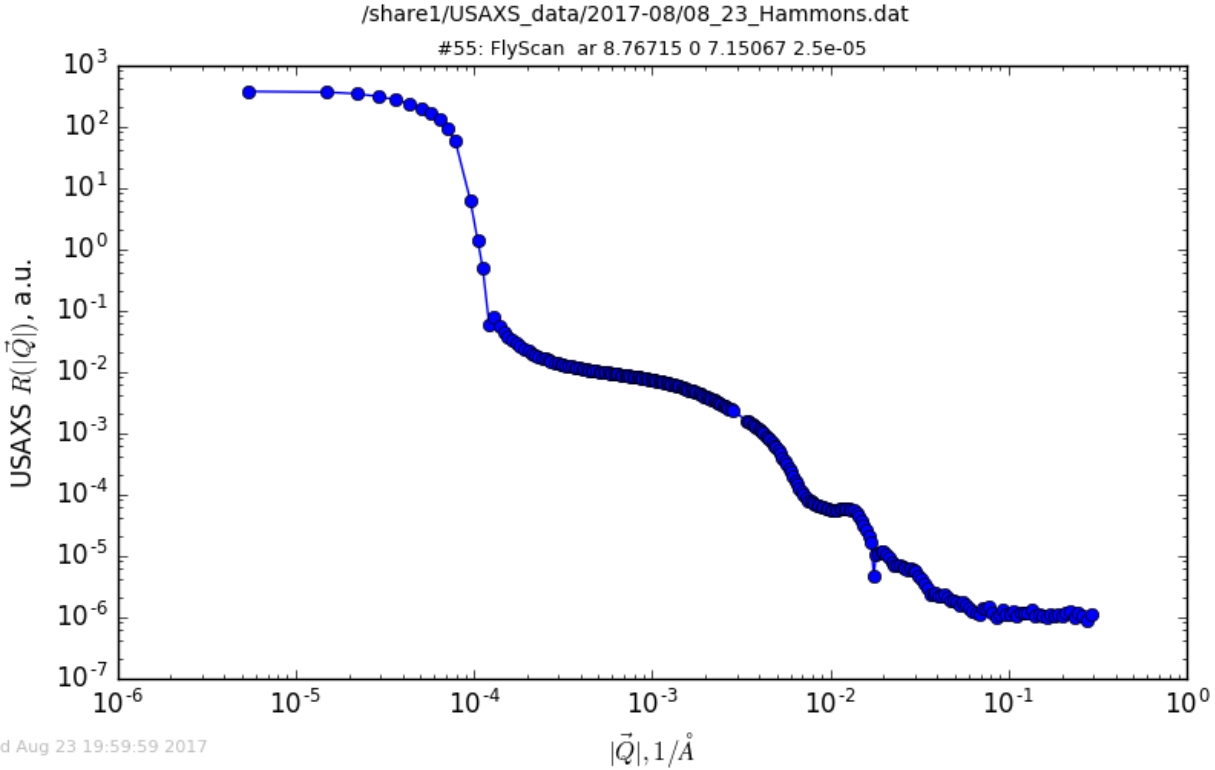


Structure of a NeXus processed data file



Example: EPICS Processed Data

USAXS, APS, Python and h5py and MatPlotLib



Example: multi-dimensional data

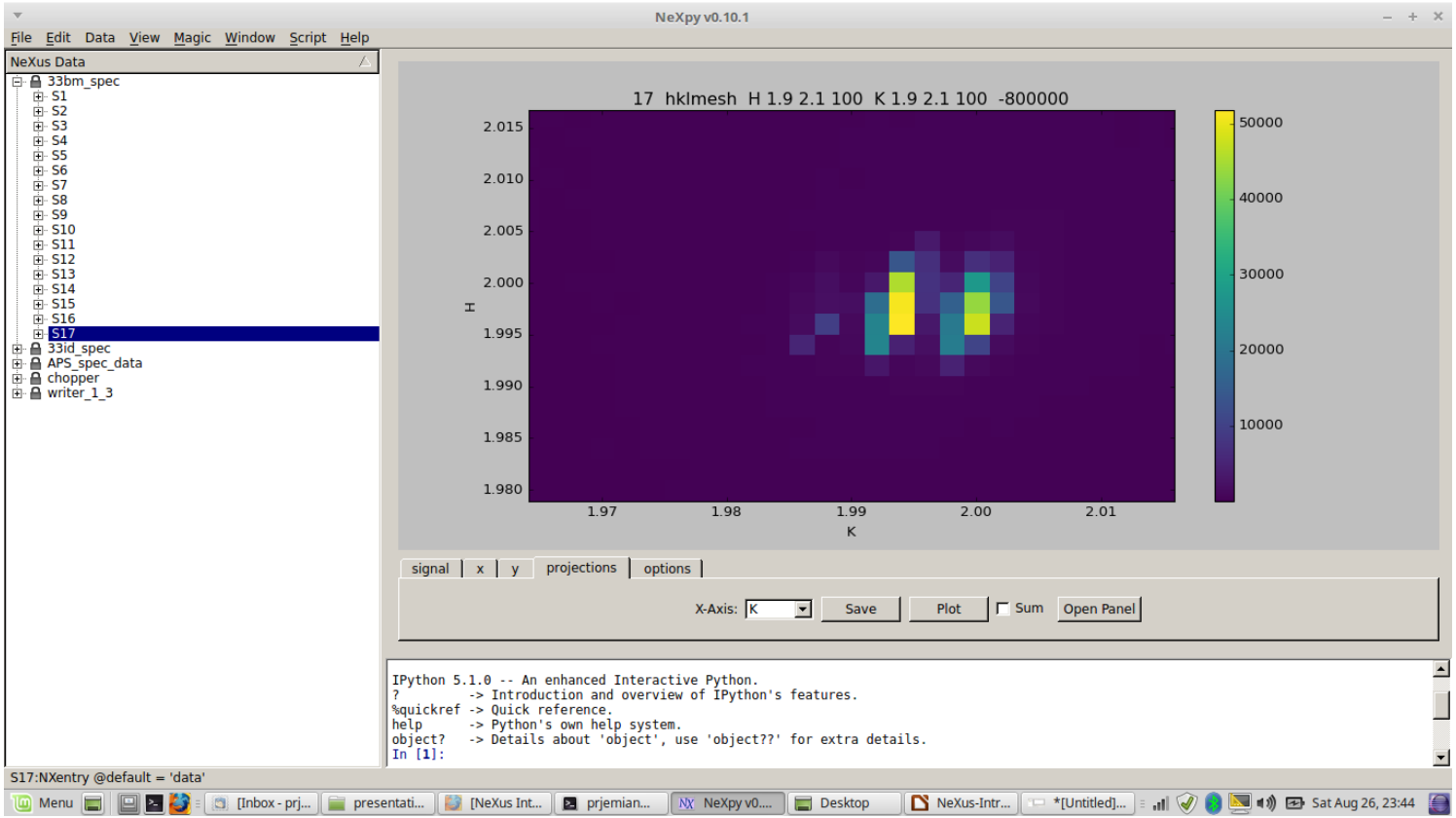
NeXus

- The AF1410 SANS data might be re-written as a multi-dimensional data structure:
  - Sector angle (nuclear *v.* nuclear+magnetic scattering)
  - Aging time
  - *Q*
- The previous SANS data would need to be binned to the same *Q* steps
- Eliminates extra structure:
  - 1 NXentry group remains
  - 1 NXdata group remains
- Could include `sector_angle` and/or `aging_time` in the analysis model

AF1410 steel SANS as multi-D

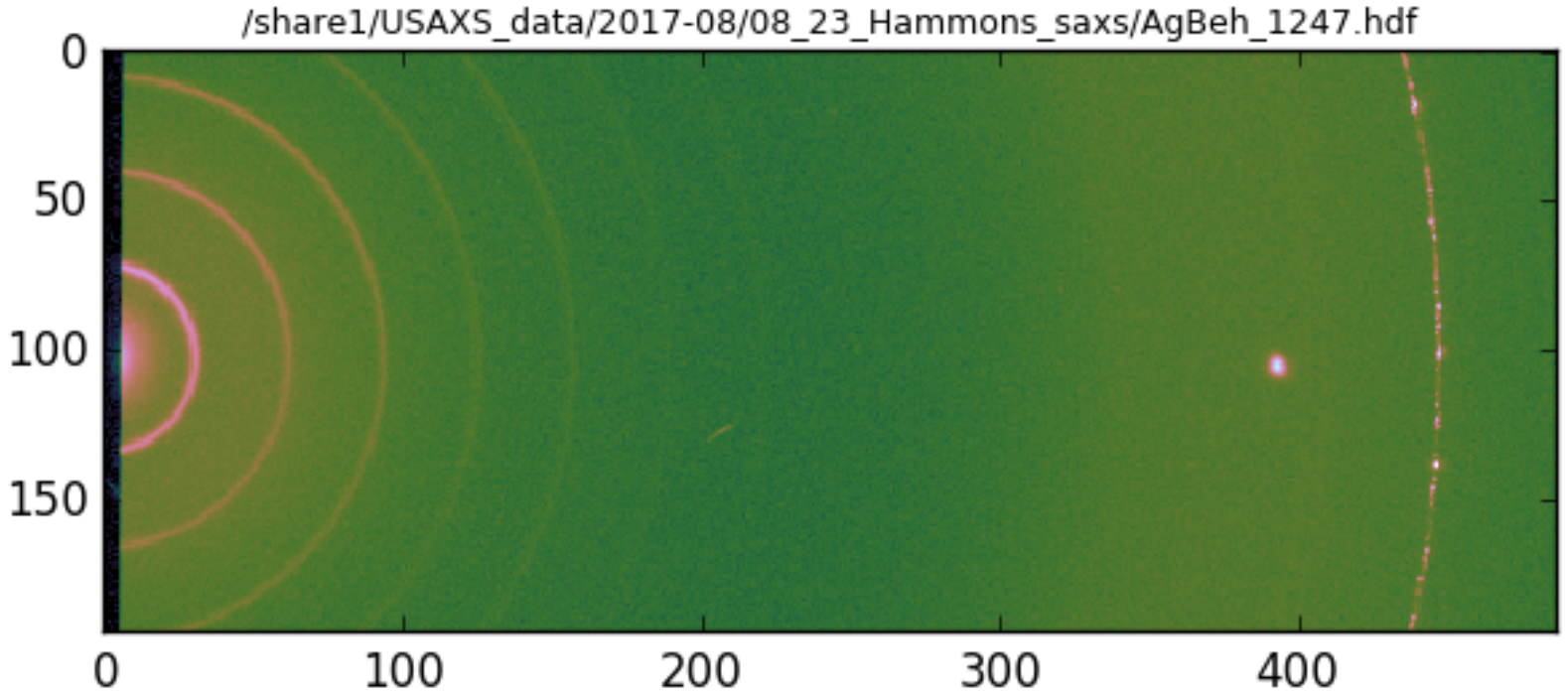
```
cs_af1410_multi.h5 : NeXus data file
  @creator = "hypothetical example"
  @default = AF1410
  AF1410:NXentry
    @default = AF1410
    AF1410:NXdata
      @signal = I
      @I_axes = Q
      @Q_indices = 2
      @aging_time_indices = 0
      @sector_angle_indices = 1
      I:NX_FLOAT64[10, 2, 77]
      Idev:NX_FLOAT64[10, 2, 77]
      Q:NX_FLOAT64[77]
      aging_time:NX_FLOAT64[10] = [0, .25,
         0.5, 1, 2, 5, 8, 10, 20 50, 100]
        @units = h
      sector_angle:NX_FLOAT64[2] = [0, 90]
        @units = degrees
    sasinstrument:NXinstrument
      sascollimation:NXcollimator
      sasdetector:NXdetector
      sassource:NXsource
    sassample:NXsample
```

Example: 2-D data from SPEC hklmesh



Example: EPICS Area Detector

Pilatus USAXS, APS, Dectris 1k, Python and h5py and MatPlotLib



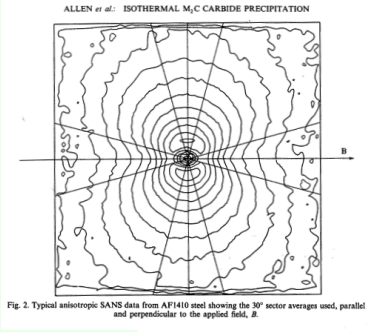
Example: many samples

AF1410 steel
10 samples (aging series)
two (NIST) SANS each sample

- Steel sample, aging series
- Area detector image at each aging time
- Binned in horizontal & vertical sectors
- 1-D I(Q) have different lengths
- (lots of other metadata in example file)
- file: cs_af1410.h5
- https://github.com/canSAS-org/NXcanSAS_examples/tree/master/1d_standard

Acta Metall **41** (1993) 1869-1884

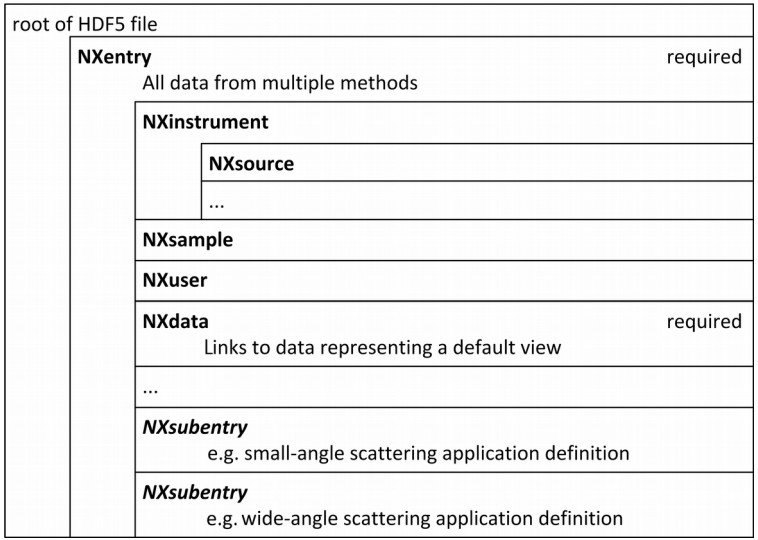

```
cs_af1410.h5 : NeXus data file
  @creator = xml2hdf5.py
  @default = AF1410_10
  AF1410_10:NXentry
    AF1410_a10:NXdata
      I:NX_FLOAT64[77]
      Idev:NX_FLOAT64[77]
      Q:NX_FLOAT64[77]
    AF1410_b10:NXdata
      I:NX_FLOAT64[76]
      Idev:NX_FLOAT64[76]
      Q:NX_FLOAT64[76]
    sasinstrument:NXinstrument
      sascollimation:NXcollimator
      sasdetector:NXdetector
      sassource:NXsource
    sassample:NXsample

  AF1410_1h:NXentry
    AF1410_a1h:NXdata
    AF1410_b1h:NXdata
  AF1410_20:NXentry
  AF1410_2h:NXentry
  AF1410_50:NXentry
  AF1410_5h:NXentry
  AF1410_8h:NXentry
  AF1410_cc:NXentry
  AF1410_hf:NXentry
  AF1410_qu:NXentry
```



Example: multi-modal data

NeXus

- NeXus structure is flexible
  - describe many sets of measurements
- Use data groups under a single entry group if from the same sample
- (Sample and wavelength data not shown for brevity)
- Or, use application definitions



2-D SAS/WAS images

```
sas-was-example.h5 : NeXus data file
  @canSAS_class = "SASroot"
  entry:NXentry
    @canSAS_class = "SASentry"
    sasdata:NXdata
      @canSAS_class = "SASdata"
      @I_axes = Q, Q
      @Q_indices = 0, 1
      @signal = I
      I:NX_FLOAT64[100,512]
      Qx:NX_FLOAT64[100,512]
      Qy:NX_FLOAT64[100,512]
      Qz:NX_FLOAT64[100,512]
    wasdata:NXdata
      @I_axes = Q, Q
      @Q_indices = 0, 1
      @signal = I
      I:NX_FLOAT64[256,256]
      Q_x:NX_FLOAT64[256,256]
      Q_y:NX_FLOAT64[256,256]
      Q_z:NX_FLOAT64[256,256]
```

Examples: How to describe uncertainty

NeXus

- These values are the *estimates* of uncertainty.
- By default, assumed as estimated standard deviation, but could specify as other
- Can be added to any field with an attribute:
  - Exact name of uncertainty dataset is flexible

simple uncertainty

```
I:NX_FLOAT64[91]
  @units = 1/cm
  @uncertainties = Iesd
Iesd:NX_FLOAT64[91]
  @units = 1/cm
Q:NX_FLOAT64[91]
  @units = 1/nm
  @resolutions = Qdev
Qdev:NX_FLOAT64[91]
  @units = 1/nm
```

uncertainty from multiple sources

```
I : float[m,n]
  @uncertainties=Idev
Idev : float[m,n]
 @components=I_uncertainties
I_uncertainties : (group)
 electronic : float[m,n]
  @basis="Johnson noise"
 counting_statistics : float[m,n]
  @basis="shot noise"
 secondary_standard : float[m,n]
  @basis="esd"
```

Example: metadata

Rich metadata can be added, but is optional

```
sassample:NXsample
     @NX_class = NXsample
     @canSAS_name = sassample
     @canSAS_class = SASsample
     ID:NX_CHAR[34] = AF1410-10 (AF1410 steel aged 10 h)
     details:NX_CHAR[128] =
       transverse saturation magnetic field (1.6 T) applied in
       horizontal direction to clear magnetic domain scattering
```

```
sasinstrument:NXinstrument
     sascollimation:NXcollimator
     sasdetector:NXdetector
     sassource:NXsource
       incident_wavelength:NX_FLOAT64 = 0.85
         @units = nm
       radiation:NX_CHAR[7] = neutron
       wavelength_spread:NX_FLOAT64 = 25.0
         @units = percent
```

Example: metadata

citation

- NeXus provides the *NXnote* to describe any additional freeform information not covered by the other base classes

- These tags and attributes were supplied *ad hoc*

```
sasnote:NXnote
     citation:NXcollection
       @NX_class = NXcollection
       @canSAS_name = citation
       @tag = citation
       journal:NX_CHAR[11] = Acta Metall
         @tag = journal
       pages:NX_CHAR[9] = 1869-1884
         @tag = pages
       title:NX_CHAR[92] = Small-Angle Neutron Scattering
         Studies of Carbide Precipitation in
         Ultrahigh-Strength Steels
         @tag = title
       volume:NX_CHAR[2] = 41
         @tag = volume
       year:NX_CHAR[4] = 1993
         @tag = year
       authors:NXcollection
         @NX_class = NXcollection
         @canSAS_name = authors
         @tag = authors
         author_0:NX_CHAR[10] = A.J. Allen
           @tag = author
         author_1:NX_CHAR[11] = D. Gavillet
           @tag = author
         author_2:NX_CHAR[13] = J.R. Weertman
           @tag = author
```

NeXus International Advisory Committee

http://www.nexusformat.org/NIAC.html

Terms of Reference

The purpose of the NeXus International Advisory Committee is to supervise the development and maintenance of the NeXus common data format for neutron, x-ray, and muon science.

How to contact the NIAC

Send an email to: nexus-committee@nexusformat.org

Membership

NIAC seeks a balanced representation of the international community. It will consist of at most one voting representative from each major neutron, synchrotron x-ray, and muon facility.

The NIAC invites PSL-II to become a member.

Communications

Mailing lists

See: http://download.nexusformat.org/doc/html/mailinglist.html

Bi-weekly teleconferences

Discuss timely issues: http://www.nexusformat.org/Teleconferences.html

Next one: Tuesday, 12 September, 16:30 CEST (+0200)

NeXus Code Camp

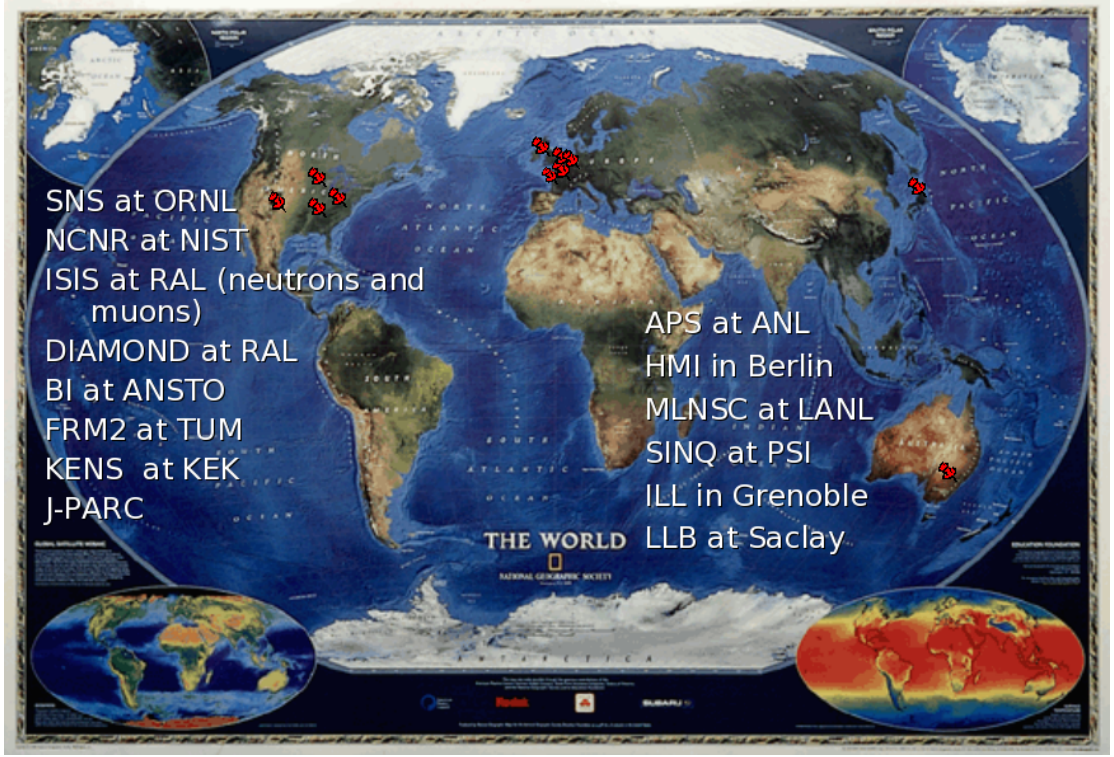
Approximately yearly, next one October 24-26, 2017, Diamond Light Source, UK

NIAC Meeting

Every two years, next one 2018, location not decided yet

Who uses NeXus?

Only a partial list from 2013



Add to this list:

- EPICS Area Detector HDF5 plugin
- Petra III at DESY, Hamburg, Germany

Acknowledgement

- Mark Basham, Diamond Light Source, UK
- Herbert Bernstein, CIF (non-facility member)
- Aaron Brewster, Lawrence Berkeley Laboratory, USA
- Stuart Campbell, Brookhaven National Laboratory/NSLS-2, USA
- Bjorn Clausen, Los Alamos National Laboratory, USA
- Stephen Cottrell, Rutherford Appleton Laboratory, UK (Muon Representative)
- Ricardo Ferraz-Leal, SNS and HFIR at ORNL, USA
- Jens-Uwe Hoffmann, Helmholtz Zentrum Berlin, Germany
- Pete Jemian, Advanced Photon Source, USA (Documentation Release Manager)
- Mark Könnecke, Paul Scherrer Institut, Switzerland (Executive Secretary)
- David Männicke, Australian Nuclear Science and Technology Organisation, Australia
- Raymond Osborn, Argonne National Laboratory, USA (non-facility member)
- Tobias Richter, European Spallation Source, Sweden (Chair)
- Armando Sole, European Synchrotron Radiation Facility, France
- Jiro Suzuki, KEK, Japan
- Benjamin Watts, Swiss Light Source, Switzerland
- Eugen Wintersberger, DESY, Germany (Technical Manager)

Thank you for your attention.