NeXus

NeXus is developed as an international standard by scientists and programmers representing major scientific facilities in Europe, Asia, Australia, and North America in order to facilitate greater cooperation in the analysis and visualization of neutron, x-ray, and muon data.

What is NeXus?

The NeXus data format has four components:

A set of design principles
to help people understand what is in the data files.

A set of data storage objects
(Base Class Definitions and Application Definitions) to allow the development of portable analysis software.

A set of subroutines
(Utilities and examples) to make it easy to read and write NeXus data files.

A Scientific Community
to provide the scientific data, advice, and continued involvement with the NeXus standard. NeXus provides a forum for the scientific community to exchange ideas in data storage.

A Set of Design Principles

NeXus data files contain four types of entity: groups, fields, attributes, and links.

Groups

Groups are like folders that can contain a number of fields and/or other groups.

Fields

Fields can be scalar values or multidimensional arrays of a variety of sizes (1-byte, 2-byte, 4-byte, 8-byte) and types (characters, integers, floats). Fields are represented as HDF5 *datasets*.
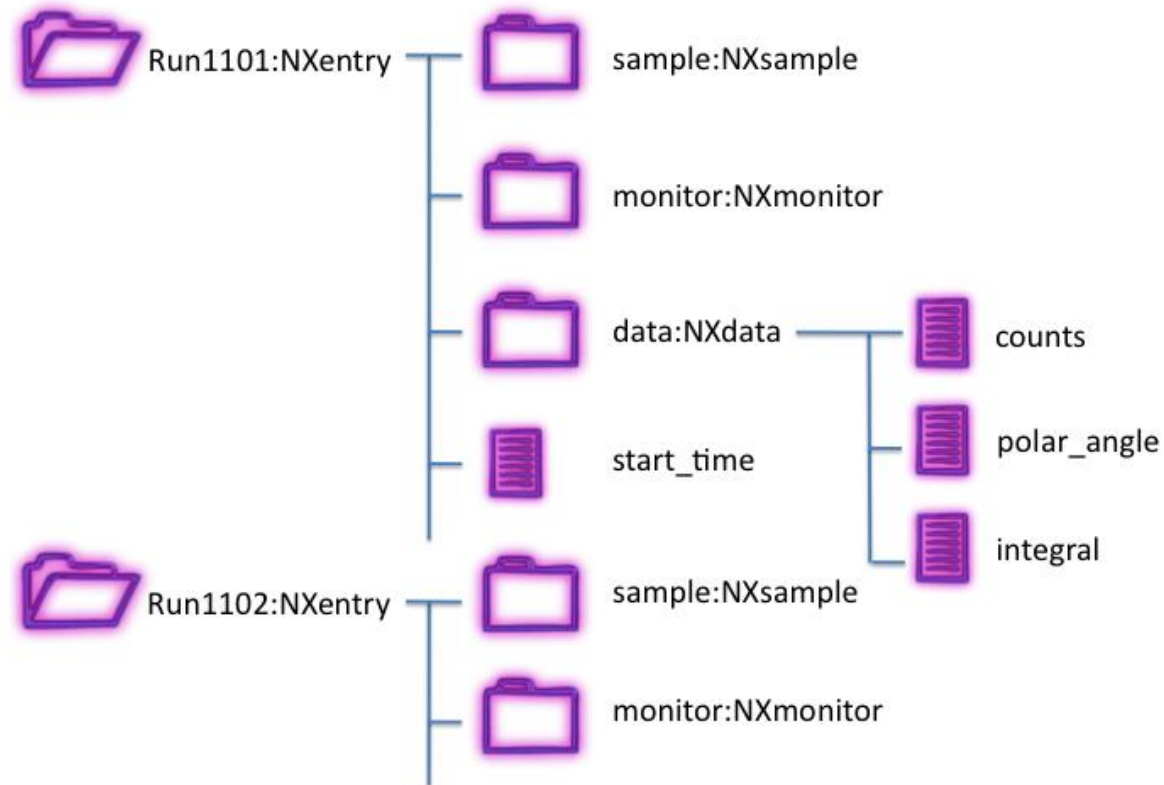
Attributes

Extra information required to describe a particular group or field, such as the data units, can be stored as a data attribute. Attributes can also be given at the file level of an HDF5 file.
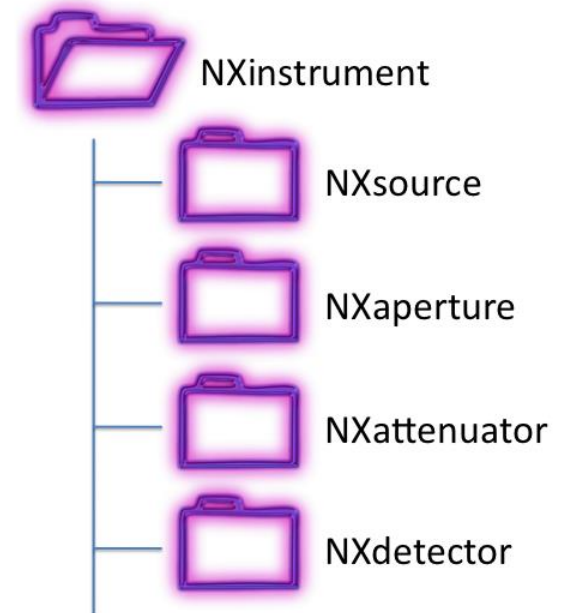
Links

Links are used to reference the plottable data from NXdata when the data is provided in other groups such as NXmonitor or NXdetector.
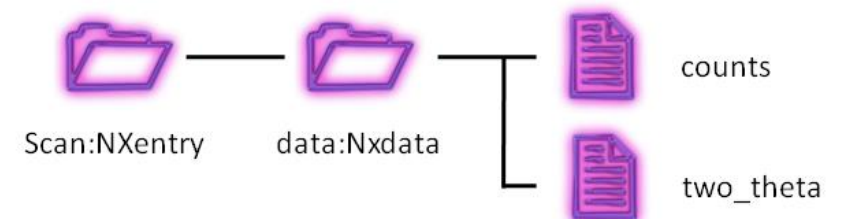
Example of a NeXus Data File

Important Classes

Example structure of a simple data file

Structure of a very simple NeXus Data file

```
1   verysimple.nx5 : NeXus data file
2     @default = entry
3     entry:NXentry
4       @NX_class = NXentry
5       @default = data
6       data:NXdata
7         @NX_class = NXdata
8         @signal = counts
9         @axes = two_theta
10        @two_theta_indices = [0]
11        counts:int32[15] = [1193, 4474, 53220, '...', 1000]
12          @units = counts
13          @long_name = photodiode counts
14        two_theta:float64[15] = [18.9094, 18.9096, '...', 18.9122]
15          @units = degrees
16          @long_name = two_theta (degrees)
```

Using Python to write a very simple NeXus HDF5 Data file

```python
1   #!/usr/bin/env python
2   '''uses h5py to build the verysimple.nx5 data file'''
3
4   import h5py
5
6   angle = [18.9094, 18.9096, 18.9098, 18.91,  18.9102,
7            18.9104, 18.9106, 18.9108, 18.911, 18.9112,
8            18.9114, 18.9116, 18.9118, 18.912, 18.9122]
9   diode = [1193, 4474, 53220, 274310, 515430, 827880,
10           1227100, 1434640, 1330280, 1037070, 598720,
11           316460, 56677, 1000, 1000]
12
13  f = h5py.File('verysimple.nx5', 'w')
14  f.attrs['default'] = 'entry'
15
16  nxentry = f.create_group('entry')
17  nxentry.attrs["NX_class"] = 'NXentry'
18  nxentry.attrs['default'] = 'data'
19
20  nxdata = nxentry.create_group('data')
21  nxdata.attrs["NX_class"] = 'NXdata'
22  nxdata.attrs['signal'] = 'counts'
23  nxdata.attrs['axes'] = 'two_theta'
24  nxdata.attrs['two_theta_indices'] = [0,]
25
26  tth = nxdata.create_dataset('two_theta', data=angle)
27  tth.attrs['units'] = 'degrees'
28  tth.attrs['long_name'] = 'two_theta (degrees)'
29
30  counts = nxdata.create_dataset('counts', data=diode)
31  counts.attrs['units'] = 'counts'
32  counts.attrs['long_name'] = 'photodiode counts'
33
34  f.close()
```

NeXus in FAIRmat

Examples of established metadata standards for materials science are **CIF**, **CSMD**, and **NeXus**. CIF (Crystallographic Information Framework) [31] is a file structure for the archiving and distribution of crystallographic information. CSMD (Core Scientific Metadata Model) [32], is a model designed to support data collected within a large-scale facility's scientific workflow. NeXus [33] is an international standard for the storage and exchange of neutron, x-ray, and muon experiment data. The structure of NeXus files is extremely flexible, allowing the storage of both simple data sets, such as a single data array, and highly complex data and their associated metadata, such as measurements on a multi-component instrument or numerical simulations.

While the above-mentioned standards are used in both experimental and computational applications, a few other metadata schemas have been introduced in the computational community. The Chemical Markup Language (CML) [34] defines about a hundred of metadata for the characterization of structures and properties of molecules. The European Theoretical Spectroscopy Facility (**ETSF**) Specifications [35] is a rich schema with hundreds of metadata, in particular for excited

NeXus in FAIRmat

WP B1.3 Parsers, normalizers, and converters

The NeXus [33] format will be used as an internal format for EM data in FAIRmat, and converters from other commonly used formats (e.g. DM3, DM4) will be developed and provided.

WP B2.1: Metadata ontologies, workflows

An ARPES-generic NeXus metadata format will predefine which information needs to be collected to acquire FAIR ARPES data using ELNs. As a service to the ARPES community, best-practice examples for hybrid software solutions will be developed, combining primary data acquisition, metadata logging, the collection of an experiment's narrative in an ELN, and also a platform for (often very sophisticated) data preprocessing pipelines, including a version control system.

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

FACHBEREICH PHYSIK
Prof. Dr. Martin Aeschlimann

Gebäude 46, Raum 459
Erwin-Schroedinger-Strasse
67663 Kaiserslautern
Deutschland
Telefon: +49 631 205 2322
Telefax: +49 631 205 3903
E-Mail: ma@physik.uni-kl.de
www.physik.uni-kl.de/aeschlimann

Technische Universität Kaiserslautern • Postfach 3049 • 67653 Kaiserslautern

Ihre Zeichen	Ihre Nachricht vom	Unsere Nachricht vom	Unsere Zeichen	Kaiserslautern
.	.	.	MA	26.09.2019

Letter of Commitment for the Research Data Initiative FAIRmat by the Transregio Collaborative Research Centre Spin+X

Dear Madam, Dear Sir,

Spin+X encompasses the whole range of spin research in a Transregio Collaborative Research Centre which as such directly relies on scientific exchange with the materials science community. Several Spin+X PIs participate within the coordinated structure of the NFDI. In this way, the concepts of data management within Spin+X are coordinated and shared with a broader scientific community. Thus, all relevant Spin+X data are handled according to the FAIR principles.

Several categories of data are generated in Spin+X, which include raw experimental data and evaluated experimental results derived from them, results of numerical simulations and, importantly, source codes and software procedures used for the treatment of the aforementioned. Therefore, structured concepts for each of these categories are developed jointly between Spin+X research groups to guarantee the transparent and uniform preservation of data.

Within Spin+X, the use of HDF5, a common self-describing data format, is being established. This well-developed open framework allows the compressed storage of huge amounts of arbitrary-dimensional data along with metadata which adds the complete context and physical meaning of the measurement. This is complemented by FAIRmat, where common conventions for the ordered structure of embedded metadata will be developed in the framework of NeXus format, which is based on HDF5. Since some of the numerous experimental techniques in Spin+X are covered in FAIRmat Area B, newly developed concepts can be implemented immediately and user experience can directly be contributed as feedback. Spin+X will develop analogous, and compatible conventions and interfaces as defined in Area D, which will help to expand the benefits of FAIRmat to a broader scientific community.

Spin+X actively encourages the principle of Open Source software by distributing commonly developed software packages under open source licenses and making them available to the public. In this way, the community will benefit from FAIRmat-compatible data acquisition, evaluation and maintenance tools.

Sincerely

Prof. Dr. Martin Aeschlimann
Speaker SFB/TRR 173 Spin+X

SPIN+X
SFB/TRR 173
Kaiserslautern • Mainz

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

trr173.spin-engineering.de

Bluesky/NeXus/HDF5/AreaDetector

- EPICS supports native HDF5 format
- you can write directly from Bluesky to HDF5 (from Jupyter notebook)
- As a science format NeXus is directly offered and fully supported
- https://git.aps.anl.gov/jemian/deployments/tree/093b01ee42e1b3f6ea89083ce9766b6207f9df9c/BlueSky