NeXus show and tell summary and NeXus keyword & units discussion

Luis Maia

Grenoble, 19.09.2019

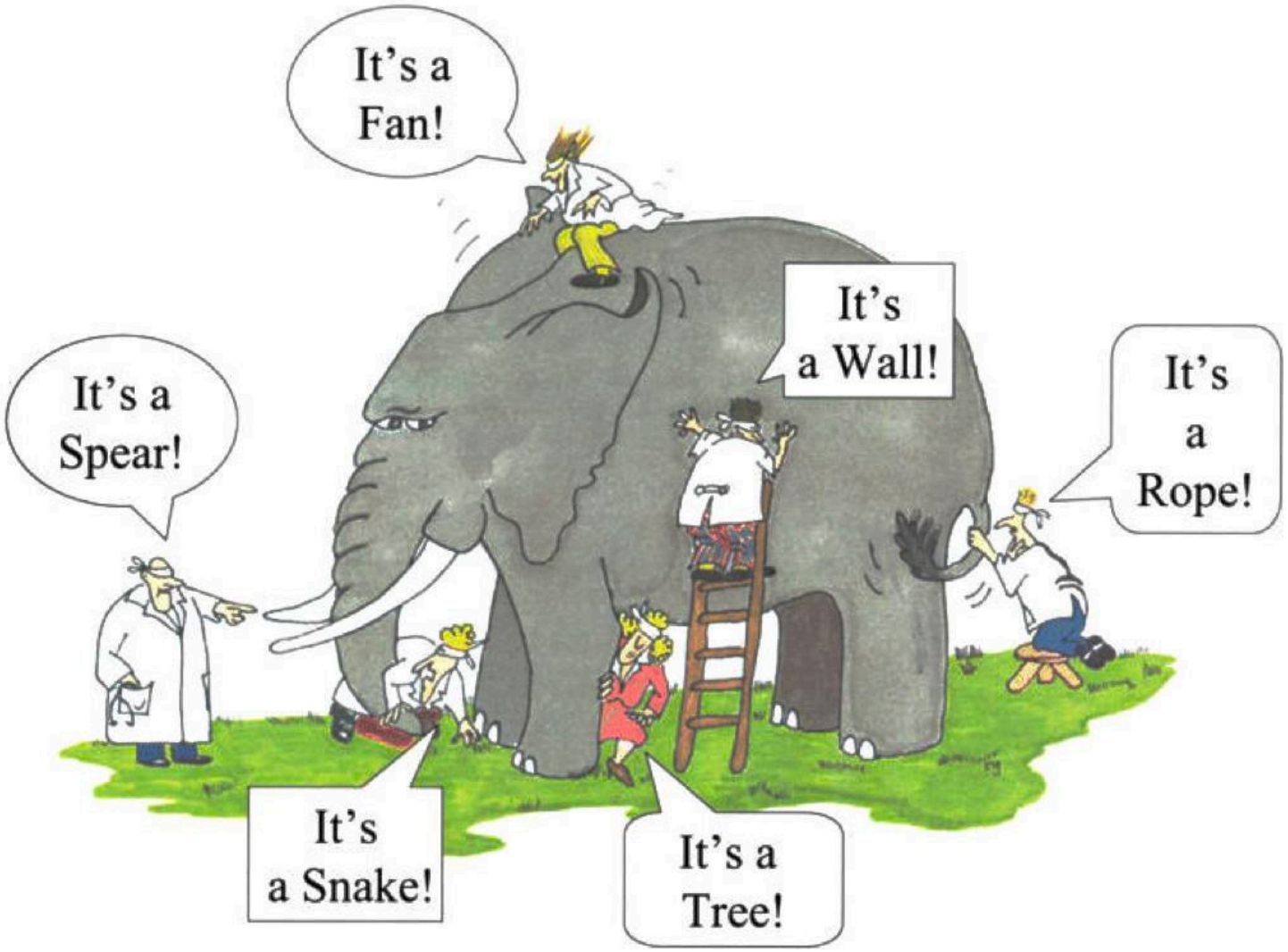
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Agenda

- Introduction
- Facilities Show and Tell - Summary
- Facilities NeXus Show and Tell
  - ELI (todo)
  - XFEL
  - CERIC
  - ESS
  - ILL
  - ESRF
- Conclusions

Introduction

- Much more than NeXus… it was about Metadata!
- Exchange of information
- Explaining each others perspective
- Establish common language
- Goals alignement



Facilities Show and Tell - Summary

	CERIC	ESS	ELI	ESRF	ILL	XFEL
User Portal	VUO	—	—	SMIS	ILL Own	UPEX
Metadata Catalog	VUO (ICAT is an option)	SciCat	TBD	ICAT	ILL Own	MyMdC
Datafiles	NeXus, HDF5, ASCII and many others	NeXus	—	EDF, SPEC, MCA, CBF, CCD, MCCD, HDF5, NeXus	NeXus and ILL Ascii	HDF5
DA Tools	WP4	WP4	WP4	WP4	WP4	WP4
NeXus	Yes*	Yes	No	Yes	Yes*	No
Logbook	DonkeyLog	SciChat	ELI Own	ESRF Own	ILL Own	Elog*

Facilities NeXus Show and Tell

- Facilities NeXus Show and Tell
  - ELI (todo)
  - XFEL
  - CERIC
  - ESS
  - ILL
  - ESRF

	CERIC	ESS	ELI	ESRF	ILL	XFEL
NeXus implementation (status on Sept. 2019)	Yes*	Yes	No	Yes	Yes*	No

ELI - General Remarks

- Instruments
  - Reaction Microscope (ReMi)
  - Magnetic bottle electron spectrometer (MBES)
  - Velocity-map imaging spectrometer (VMIS)
  - Ion Microscope (IM)

- Would like to use NeXus

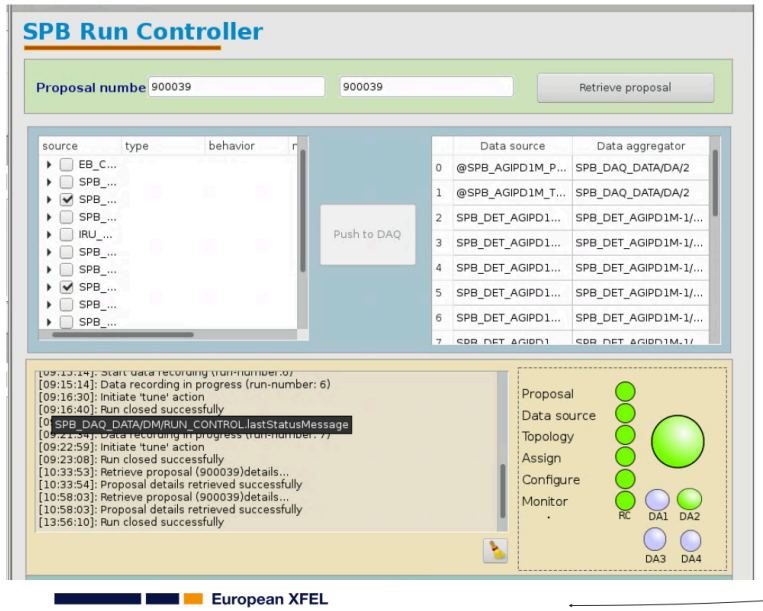
- Early phase of development

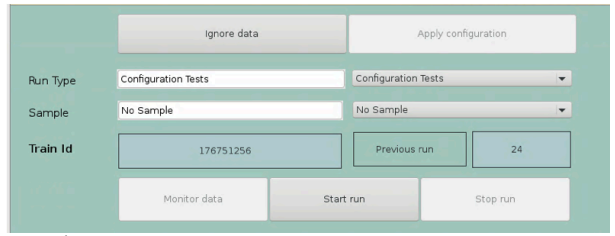
Note: Show and tell to be done on the 27th of September

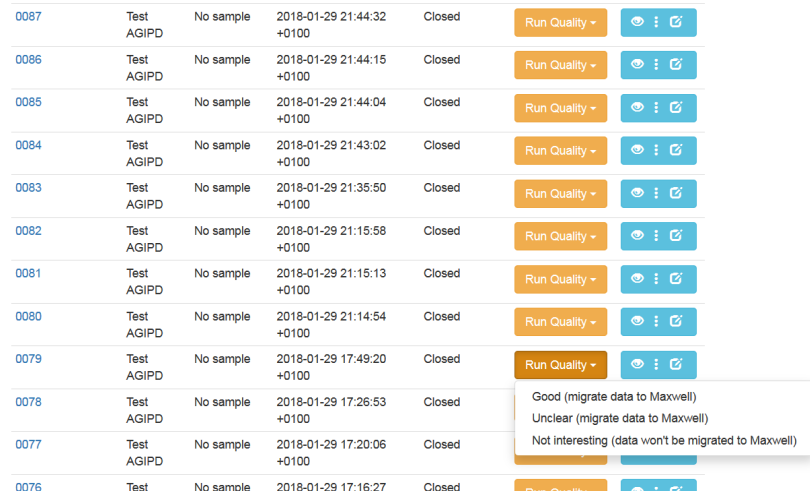
Courtesy ELI

XFEL - General Remarks

How is the existing metadata linked with the generated data?







Proposal no. 002359

Back | Edit | Clone | Add another | Runs | Delete | Beamtime status

General | Public Information | Runs | Team | History

Proposal Runs

Run Number (alias)	Run type	Sample Name	Start date	Run status	Data Assessment	Calibration	Run Comment	Edit
0092	Test DAQ	CsPbBr3 nanoparticles	2019-08-25 19:59:24 +0200	Closed	Good			
0091	Test DAQ	CsPbBr3 nanoparticles	2019-08-25 19:57:03 +0200	Closed	Good			
0090	Test DAQ	CsPbBr3 nanoparticles	2019-08-25 19:36:21 +0200	Closed	Good			
0089	Test DAQ	CsPbBr3 nanoparticles	2019-08-25 19:33:18 +0200	Closed	Good			
0088	Test DAQ	CsPbBr3 nanoparticles	2019-08-25 19:28:50 +0200	Closed	Good			
0087	Test DAQ	CsPbBr3 nanoparticles	2019-08-25 19:22:35 +0200	Closed	Good			
0086	Test DAQ	CsPbBr3 nanoparticles	2019-08-25 16:59:58 +0200	Closed	Good			

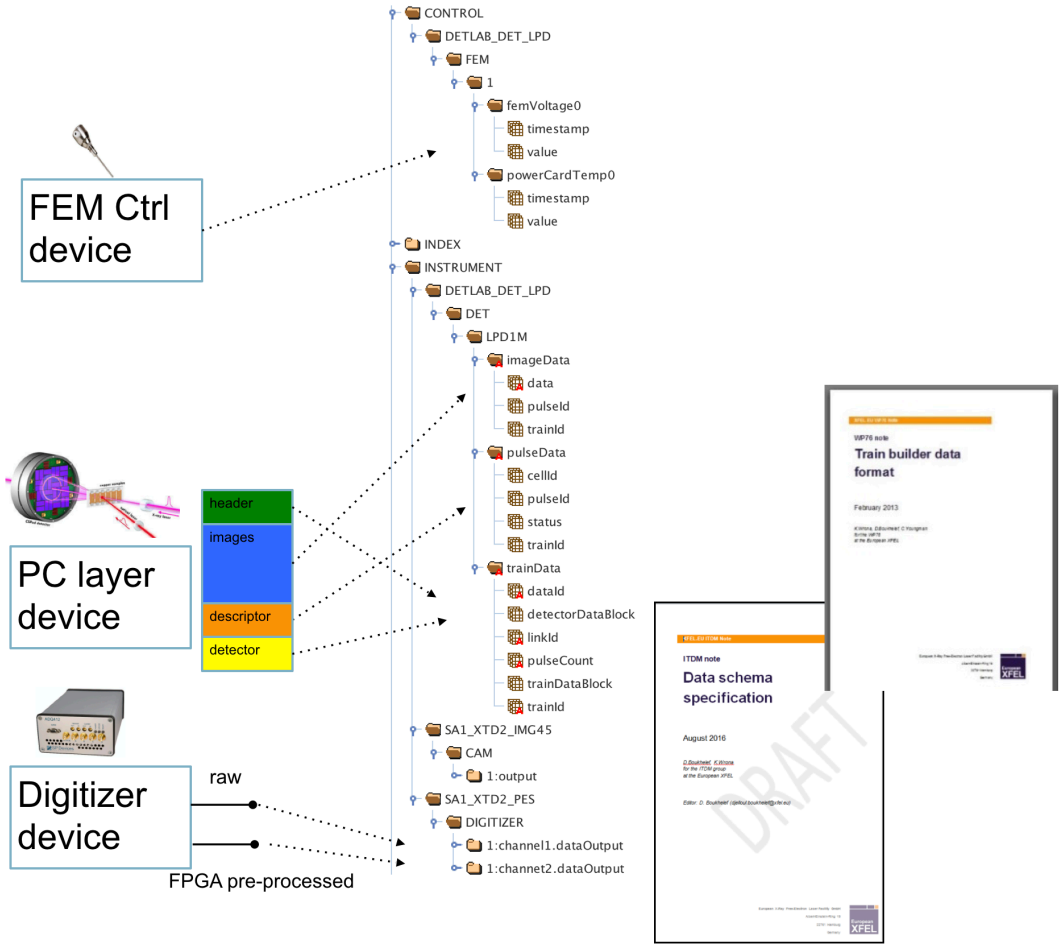
Courtesy XFEL

XFEL - HDF5

How is data organized?

- Control data
  - These are parameters of Karabo devices
    - ► PC layer aggregator uses the standard Karabo mechanism to register for any update of the parameter value

- Instrument data
  - Data acquired and processed by electronic devices (e.g. uTCA, Train Builder)
    - ► Available as a binary object described by XFEL Train Data Format (XTDF)
    - ► sent out to PC layer using application level Train Transfer Protocol (TTP) over UDP network protocol
  - Data acquired and sent out to PC layer by software device controlling the hardware via dedicated DAQ network
    - ► Available as Hash structure
    - ► Sent out to PC layer using Karabo Pipeline



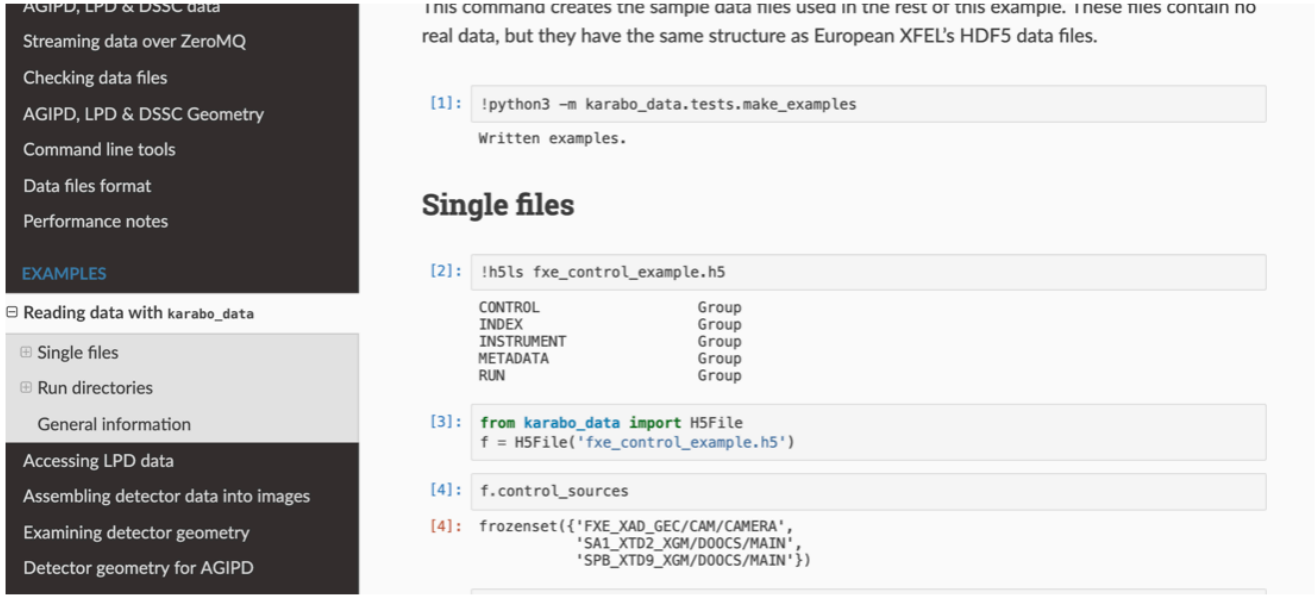
European XFEL

Courtesy XFEL

XFEL - Data Interaction

Tools to interact with the Data (e.g. Karabo Data)

https://karabo-data.readthedocs.io/en/latest/Demo.html



Courtesy XFEL

CERIC - General Remarks

CERIC data at glance

- CERIC is a consortium that allow users to choose 53 between instruments and laboratories distributed in 8 facilities of different nations.
- Start with statutory seat (Elettra) that counts 21 labs (beamlines)
- All CERIC users can use VUO in order to apply and manage their proposals
- Current state:
  - Just a beamline of Elettra facility (MCX) has some dataset saved in the online storage (under embargo, no NeXus)
  - We are discussing with the Elettra IT team which could be the best solution to transfer dataset from external facilities to Elettra IT infrastructure (VPN, WEBdav)

Data acquisition

- Mainly the acquisition software used to collect the data of the beamline is developed by scientist in charge
- Elettra beamline IT team provides support to the beamline scientists

Courtesy CERIC

CERIC - MSB Example

MSB (Material Science Beamline) used for experiments in materials science, surface physics, catalysts and organic molecules on various surfaces.

About the data:

- **Format**: XML file
- **Size:** A dataset has a size of **some (10^1-10^2) MBytes**
- **Metadata:** Are stored with the raw data, in the same xml file

MSB - Photoemission/absorption - Dataset examples

XML

```
...
<ulong name="num_scans">3</ulong>
<double name="scan_delta">0.05</double>
<double name="excitation_energy">410</double>
<double name="kinetic_energy">105</double>
<double name="kinetic_energy_base">0</double>
<double name="pass_energy">10</double>
<double name="bias_voltage">80</double>
<double name="detector_voltage">1825</double>
```

MetaData

```
...
<ulong type_id="IDL:specs.de/SurfaceAnalysis/Counts:1.0" type_name="Counts">
7
2
8
7
7
...
```

RawData

Electrons count

```
</ulong>
```

Courtesy CERIC

CERIC - SYRMEP (Tomography) Example

SYRMEP (Synchrotron Radiation for Medical Physics) is designed for research in medical diagnostic radiology and large number of different microtomography experiments.

About the data:

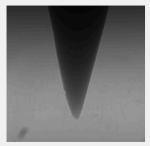
- **Format**:
  - TDF (HDF5)
  - HIS and TIFF
- **Size:** A dataset has a size of **some (10^0-10^1) GBytes**
- **Metadata:** Are stored with the raw data, in the same file

HIS

Binary file

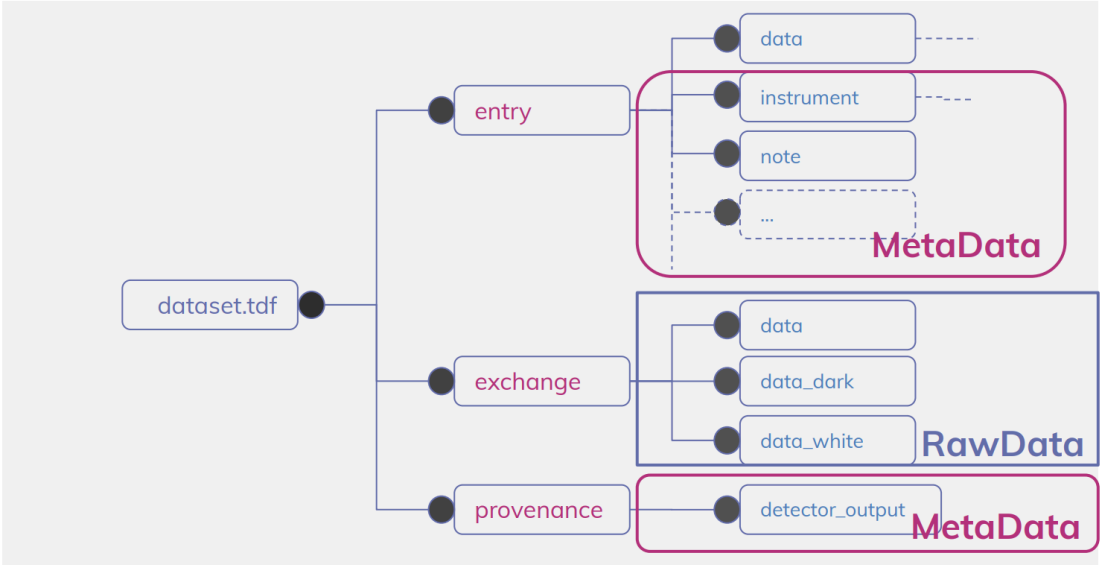
sequence of images

TIFF



Preprocessing

TDF (Hdf5) + NeXus



Courtesy CERIC

ESS - Base Classes

Contain parameters common for particulars type of equipment or sample, user, etc.

```
base_classes$ ls
NXaperture.nxdl.xml
NXattenuator.nxdl.xml
NXbeam.nxdl.xml
NXbeam_stop.nxdl.xml
NXbending_magnet.nxdl.xml
NXcapillary.nxdl.xml
NXcharacterization.nxdl.xml
NXcite.nxdl.xml
NXcollection.nxdl.xml
NXcollimator.nxdl.xml
NXcrystal.nxdl.xml
NXdata.nxdl.xml
NXdetector.nxdl.xml
NXdetector_group.nxdl.xml
NXdetector_module.nxdl.xml
NXdisk_chopper.nxdl.xml
NXentry.nxdl.xml
NXenvironment.nxdl.xml
NXevent_data.nxdl.xml
NXfermi_chopper.nxdl.xml
NXfilter.nxdl.xml
NXflipper.nxdl.xml
NXfresnel_zone_plate.nxdl.xml
NXgeometry.nxdl.xml
NXgrating.nxdl.xml
NXguide.nxdl.xml
NXinsertion_device.nxdl.xml
NXinstrument.nxdl.xml
NXlog.nxdl.xml
NXmirror.nxdl.xml
NXmoderator.nxdl.xml
NXmonitor.nxdl.xml
NXmonochromator.nxdl.xml
NXnote.nxdl.xml
NXobject.nxdl.xml
NXorientation.nxdl.xml
NXparameters.nxdl.xml
NXpinhole.nxdl.xml
NXpolarizer.nxdl.xml
NXpositioner.nxdl.xml
NXprocess.nxdl.xml
NXroot.nxdl.xml
NXsample.nxdl.xml
NXsensor.nxdl.xml
NXshape.nxdl.xml
NXslit.nxdl.xml
NXsource.nxdl.xml
NXsubentry.nxdl.xml
NXtransformations.nxdl.xml
NXtranslation.nxdl.xml
NXuser.nxdl.xml
NXvelocity_selector.nxdl.xml
NXxraylens.nxdl.xml
nxdlformat.xsl
```

With those you can build up a hierarchy describing a fairly complete description of an experiment.

Courtesy ESS

ESS - Application Definitions

Aim was to describe the file contents (mostly taken from base class definitions) required for the analysis of a particular “application”.

NXarchive.nxdl.xml	NXmonopd.nxdl.xml	NXsqom.nxdl.xml	NXtomoproc.nxdl.xml	NXxnb.nxdl.xml
NXarpes.nxdl.xml	NXmx.nxdl.xml	NXstxm.nxdl.xml	NXxas.nxdl.xml	NXxrot.nxdl.xml
NXcanSAS.nxdl.xml	NXrefscan.nxdl.xml	NXtas.nxdl.xml	NXxasproc.nxdl.xml	canSAS
NXdirecttof.nxdl.xml	NXreftof.nxdl.xml	NXtofnpd.nxdl.xml	NXxbase.nxdl.xml	nxdlformat.xsl
NXfluo.nxdl.xml	NXsas.nxdl.xml	NXtofraw.nxdl.xml	NXxeuler.nxdl.xml	
NXindirecttof.nxdl.xml	NXsastof.nxdl.xml	NXtofsingle.nxdl.xml	NXxkappa.nxdl.xml	
NXiqproc.nxdl.xml	NXscan.nxdl.xml	NXtomo.nxdl.xml	NXxlaue.nxdl.xml	
NXlauetof.nxdl.xml	NXspe.nxdl.xml	NXtomophase.nxdl.xml	NXxlaueplate.nxdl.xml	

Like the base classes they are defined via XML/XSD schema files with custom documentation elements that produce part of the NeXus manual.

Courtesy ESS

ESS - Locations and Orientations

depends_on
locates and orients components and is used to chain transformations corresponding to their physical setup

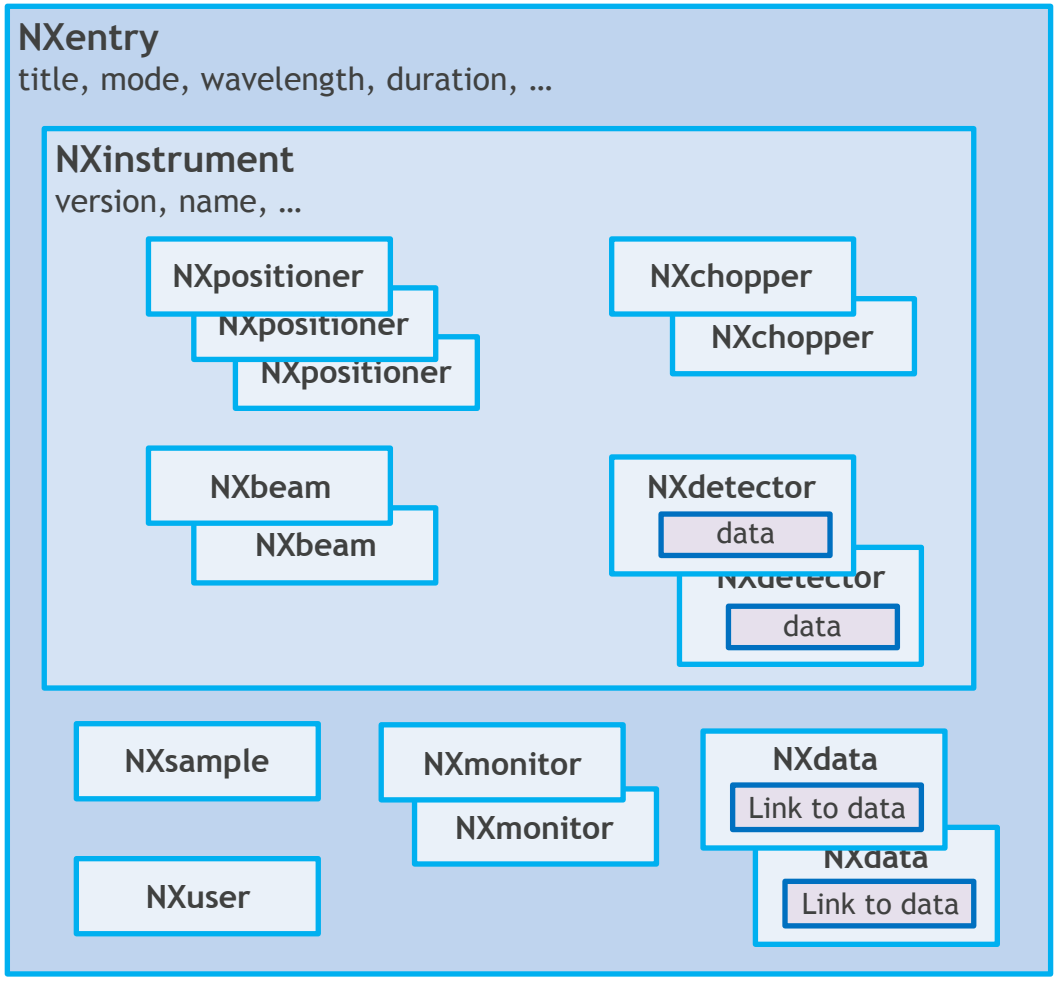
Transformations
describe the dynamic or static placement of components with @transformation_type, @vector, @depends_on, @units, @offset, etc

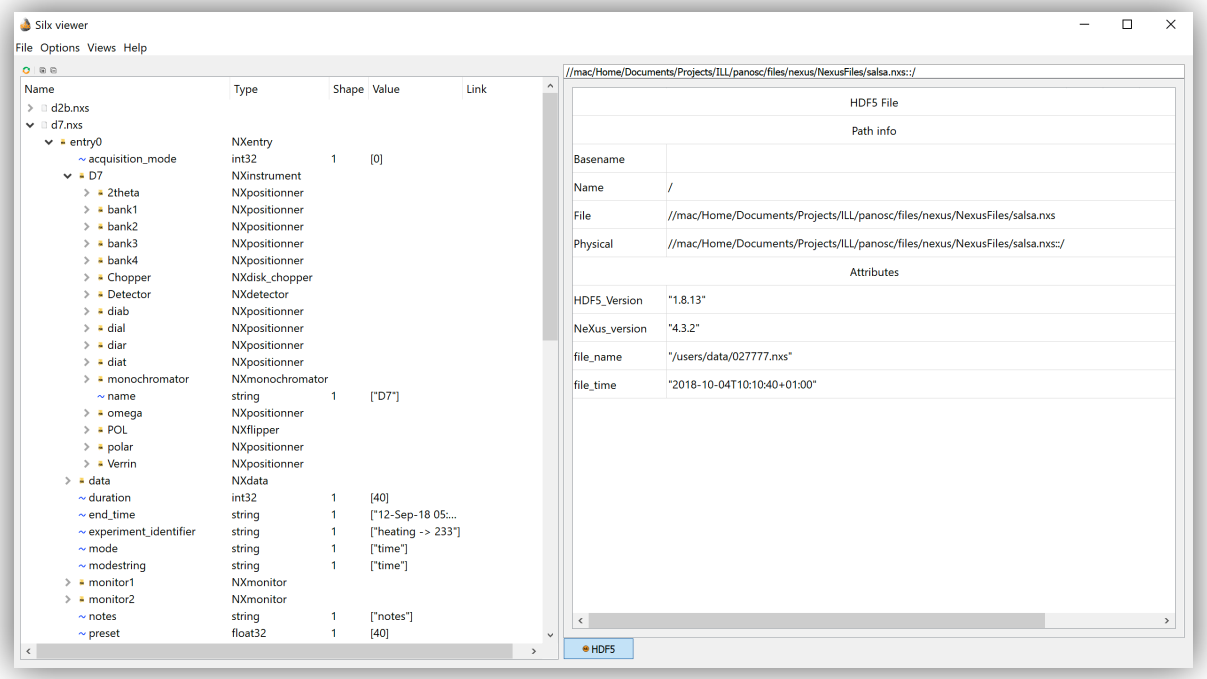
NXtransformations
used to group transformations, for example to have axes on one diffractometer together

```
entry:NXentry
  data:NXdata
    …
  instrument:NXinstrument
    detetctor:NXdetector
      …
  sample:NXsample
    depends_on=diffr/phi
    diffr:NXtransformations
      phi[…]
        @transformation_type=rotation
        @vector=0,1,0
        @depends_on=chi
      chi[…]
        @transformation_type=rotation
        @vector=0,0,1
        @depends_on=rotation_angle
      rotation_angle[…]
        @transformation_type=rotation
        @vector=0,1,0
```

Courtesy ESS

ILL - General NeXus structure of ILL data files





Courtesy ILL

ILL - Metadata in detail

Instrument matrix for root entry

Metadata	Type	Description	in4	in5	in6	in16b	d7	d9	d10	d11	d17	d19	d22	d33	d50	brisp	figaro	salsa
Entry																		
title/experiment_title	char	Title	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
subtitle/sample_description	char	Subtitle				■				■			■	■		■		
start_time	date	Start time	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
end_time	date	End time	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
experiment_identifier	char	Exp Id/Sub title/exp number	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
run_number	int32	File numor	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
wavelength	float32	Wavelength	■	■	■	■		■	■		■	■	■			■		■
Reactor power	float32					■				■			■	■				
duration	float32	Count duration	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
mode	char	Acquisition mode	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
modestring	char					■	■	■	■	■	■	■		■	■	■	■	■
acquisition_mode	int32	Acquisition mode		■	■	■	■			■	■		■		■	■	■	
preset	float32	Wanted count time	■			■		■	■		■	■	■	■		■	■	■
inhibit_time	float32	Inhibit time																
actual_time/time	float32	Count duration	■			■		■	■		■	■	■		■	■	■	■
instrument_name	char	Instrument name								■			■	■				

Instrument matrix for sample entry

Metadata	Type	Description	in4	in5	in6	in16b	d7	d9	d10	d11	d17	d19	d22	d33	d50	brisp	figaro	salsa
Sample																		
use_temp	int32	Use temperature						■	■			■						■
temperature	float32	Temperature	■	■	■	■	■	■	■	■	■	■	■	■		■	baths	■
regulation_temperature	float32	Regulation temperature	■	■	■	■	■	■	■	■	■	■	■	■		■		■
setpoint_temperature	float32	Setpoint temperature	■	■	■	■	■	■	■	■	■	■	■	■		■		■
use_mag	int32	Use magnetic field						■	■									■
field/field_actual	float32	Magnetic field		■		■	■	■	■	■	■		■	■				■
setpoint_field/field_requested	float32	Setpoint magnetic field		■		■	■	■	■	■	■		■	■				■
pressure	float32	Pressure		■	■	■												
setpoint_pressure	float32	Setpoint pressure		■	■	■												
sampleId	char	Sample Id	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■
chemicalFormula	char	Chemical formula	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■
volumeUnitCell	char		■	■	■	■	■	■	■	■	■	■	■	■		■	■	■
mass	char	Mass	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■
density	char	Density	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■
typeInstall	char		■	■	■	■	■	■	■	■	■	■	■	■		■	■	■
scatteringLengthDensity	char		■	■	■	■	■	■	■	■	■	■	■	■		■	■	■
unitCellClass	char		■	■	■	■	■	■	■	■	■	■	■	■		■	■	■
surfaceArea	char	Surface area	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■
consistance	char	Consistance	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■
size	char	Size	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■
sampledistance	float32	Sample distance	■							Lots of	Lots of		Lots of	Lots of				
cell_pressure	float32	Cell pressure			■					other ent	other entries		other ent	other entries				
piston_position	float32	Piston position			■					to sampl	to sample		to sampl	to sample				

Instrument matrix for user entry

Metadata	Type	Description	in4	in5	in6	in16b	d7	d9	d10	d11	d17	d19	d22	d33	d50	brisp	figaro	salsa
User																		
name	char	User name	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	
namelocalcontact	char	Local contact name	■	■	■	■	■	■	■	■	■	■	■	■	■		■	
proposal	char	Proposal number	■	■	■	■	■	■	■	■	■	■	■	■	■		■	
usernamelist	char					■							■					
localcontactlist	char					■							■					

Courtesy ILL

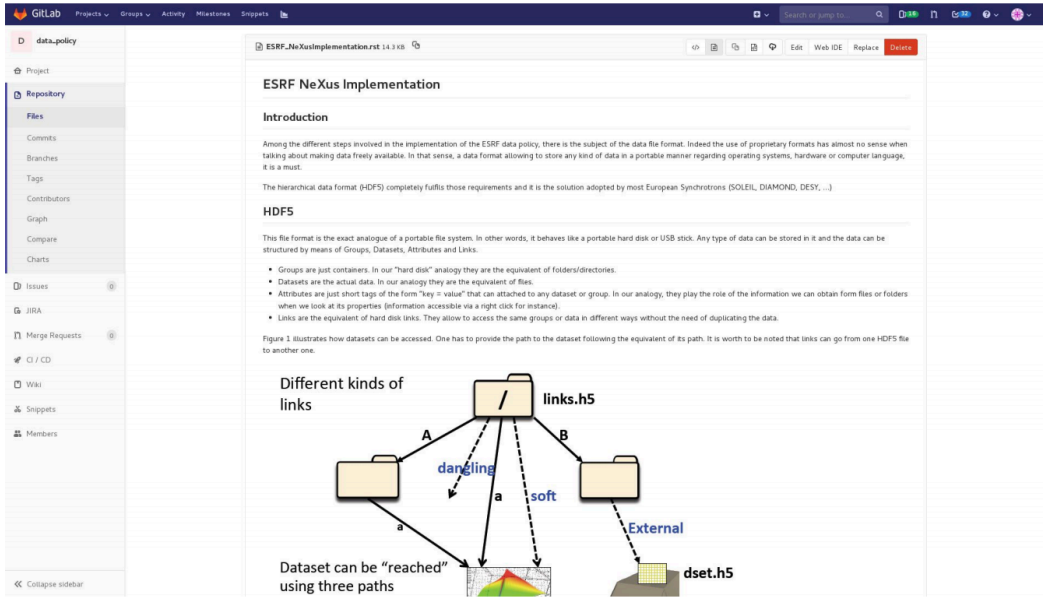
ESRF - Use of the nexus at the ESRF

		Data Policy Implementation			
Beamline	Status	Techniques	Metadata Collection*	Data archiving*	Raw Data in HDF5*	Open access to data*
ID01		KMAP	Implemented	Implemented		Implemented
BM01A						
BM01B						
ID02						
BM02						
ID03						
BM05		Tomography	Implemented	Implemented		Implemented
ID06-LVP						
ID06						
BM08						
ID09						
ID10						
ID11		Tomography	Implemented	Implemented		Implemented
ID12						
ID13						
BM14						
ID15A						
ID15B						
ID16A		Fluo, Tomo	Implemented	Implemented		Implemented
ID16B		Tomo	Implemented	Implemented		Implemented
ID17		MRT, Tomography	Implemented	Implemented		Implemented
ID18						
ID19		Tomography	Implemented	Implemented		Implemented
ID20		RIXS	In progress	In progress		In progress
ID21		Microscopy	Implemented	Implemented		Implemented
ID22						
ID23-1		MX	Implemented	Implemented		Implemented
ID23-2		MX	Implemented	Implemented		Implemented
BM23						
ID24						
BM25A						
BM25B						
ID26						
BM26A						
BM26B						
ID27						
ID28						
BM28						
ID29		MX	Implemented	Implemented		Implemented
BM29		BIOSAXS	Implemented	Implemented		Implemented
ID30A-1		MX	Implemented	Implemented		Implemented
ID30A-3		MX	Implemented	Implemented	Implemented	Implemented
ID30B		MX	Implemented	Implemented		Implemented
BM30A						
BM30B						
ID31						
ID32		RIXS	In progress	In progress		In progress
BM32						
CryoEM		Single Particle	Implemented	Implemented		Implemented

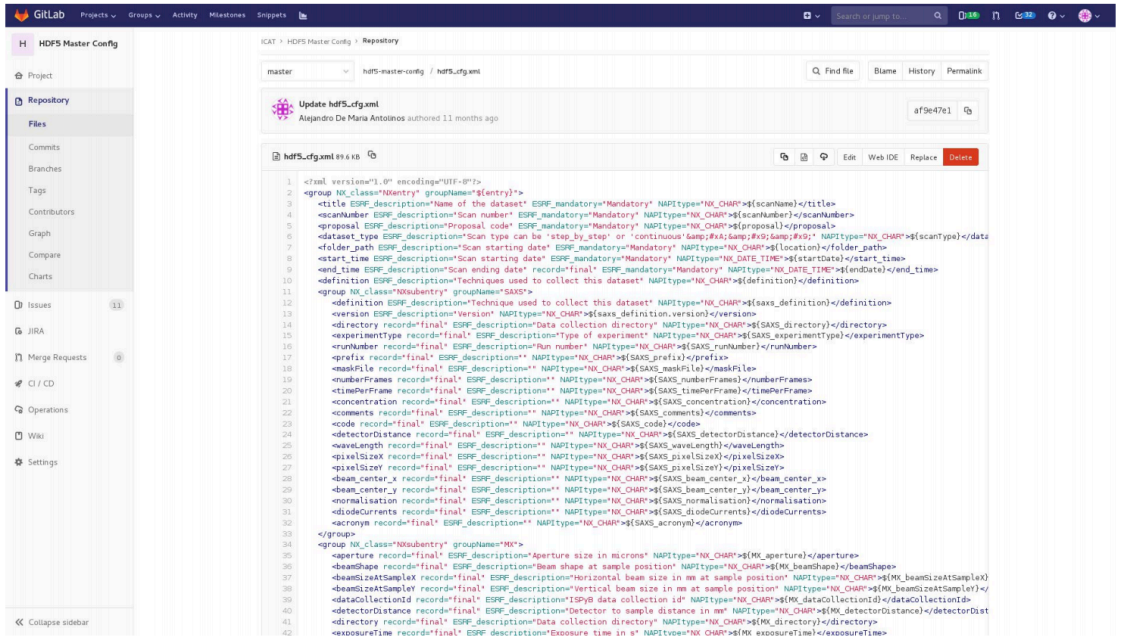
https://www.esrf.eu/datapolicy

Courtesy ESRF

ESRF - Nexus related documentation



https://gitlab.esrf.fr/sole/data_policy/blob/master/ESRF_NeXusImplementation.rst



https://gitlab.esrf.fr/icat/hdf5-master-config/blob/master/hdf5_cfg.xml

Courtesy ESRF

ESRF - HDF5, Nexus and ICAT integration

- HDF5 as a mirror of ICAT on the local beamline file system
- Following the NEXUS convention

```
−<group NX_class="NXentry" groupName="${entry}">
    <title ESRF_description="Name of the dataset" ESRF_mandatory="Mandatory" NAPItype="NX_CHAR">${scanName}</title>
    <scanNumber ESRF_description="Scan number" ESRF_mandatory="Mandatory" NAPItype="NX_CHAR">${scanNumber}</scanNumber>
    <proposal ESRF_description="Proposal code" ESRF_mandatory="Mandatory" NAPItype="NX_CHAR">${proposal}</proposal>
    <dataset_type ESRF_description="Scan type can be 'step_by_step' or 'continuous'&#xA;&#x9;&#x9;" NAPItype="NX_CHAR">${scanType}</dataset_type>
    <folder_path ESRF_description="Scan starting date" ESRF_mandatory="Mandatory" NAPItype="NX_CHAR">${location}</folder_path>
    <start_time ESRF_description="Scan starting date" ESRF_mandatory="Mandatory" NAPItype="NX_DATE_TIME">${startDate}</start_time>
    <end_time ESRF_description="Scan ending date" record="final" ESRF_mandatory="Mandatory" NAPItype="NX_DATE_TIME">${endDate}</end_time>
    <definition ESRF_description="Techniques used to collect this dataset" NAPItype="NX_CHAR">${definition}</definition>
  +<group NX_class="NXsubentry" groupName="SAXS"></group>
  +<group NX_class="NXsubentry" groupName="MX"></group>
  +<group NX_class="NXsubentry" groupName="PTYCHO"></group>
  +<group NX_class="NXsubentry" groupName="FLUO"></group>
  +<group NX_class="NXsubentry" groupName="TOMO"></group>
  +<group NX_class="NXsubentry" groupName="MRT"></group>
  +<group NX_class="NXsubentry" groupName="HOLO"></group>
  +<group NX_class="NXsubentry" groupName="WAXS"></group>
  +<group NX_class="NXsample" groupName="sample"></group>
  +<group NX_class="NXinstrument" groupName="instrument"></group>
  +<group NX_class="NXnote" groupName="notes"></group>
  </group>
```

NXroot

Top level. One per file.

NXentry

One group per measurement

NXinstrument

Describe the instrument.
Only one per NXentry

measurement (@NXcollection)

Flattened view of everything measured
Only one per NXentry

sample (@NXsample)

Define the physical state of the sample during the scan

NXdata

The data to be plotted.
One NXdata group per plot

user (@NXuser)

Details of a user, i.e., name, affiliation, email address,*etc*

NXsubentry

Data or links to data for particular analysis

Courtesy ESRF

Conclusions

- All facilities use or would like to use NeXus
  - Let's use it!
- NeXus evolution pace is too slow for Facilities pace
  - Implementation of data model as 'contributed definitions' in NEXUS Github repository. This is where the community can drop unofficial NEXUS definitions. NIAC can then review them before accepting and upgrading them to 'application definitions' or to 'base class definitions'.
- Defining the next steps
  - (?) Define which Data Models are the most common
  - (?) Tomography NeXus definition taking into account CERIC, ILL and ESRF experience
  - (?) Each site implements chosen Data Model

Questions?