

NeXus at ESRF

CALIPSOPlus JRA2/LEAPS WG3 IT/ UmbrellaID Meeting

Paul Scherrer Institute

May 2019

V. Armando SOLÉ
ESRF – Data Analysis Unit

This Talk

- ESRF NeXus interpretation for raw data
- ESRF NeXus interpretation for processed data
- Other ESRF NeXus Uses: Metadata storage
- Status

HDF5/NeXus – ESRF Interpretation for Raw Data

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 default 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

Exclusive **Acquisition** Domain

Almost exclusive **Acquisition** Domain

User/Scientist Domain

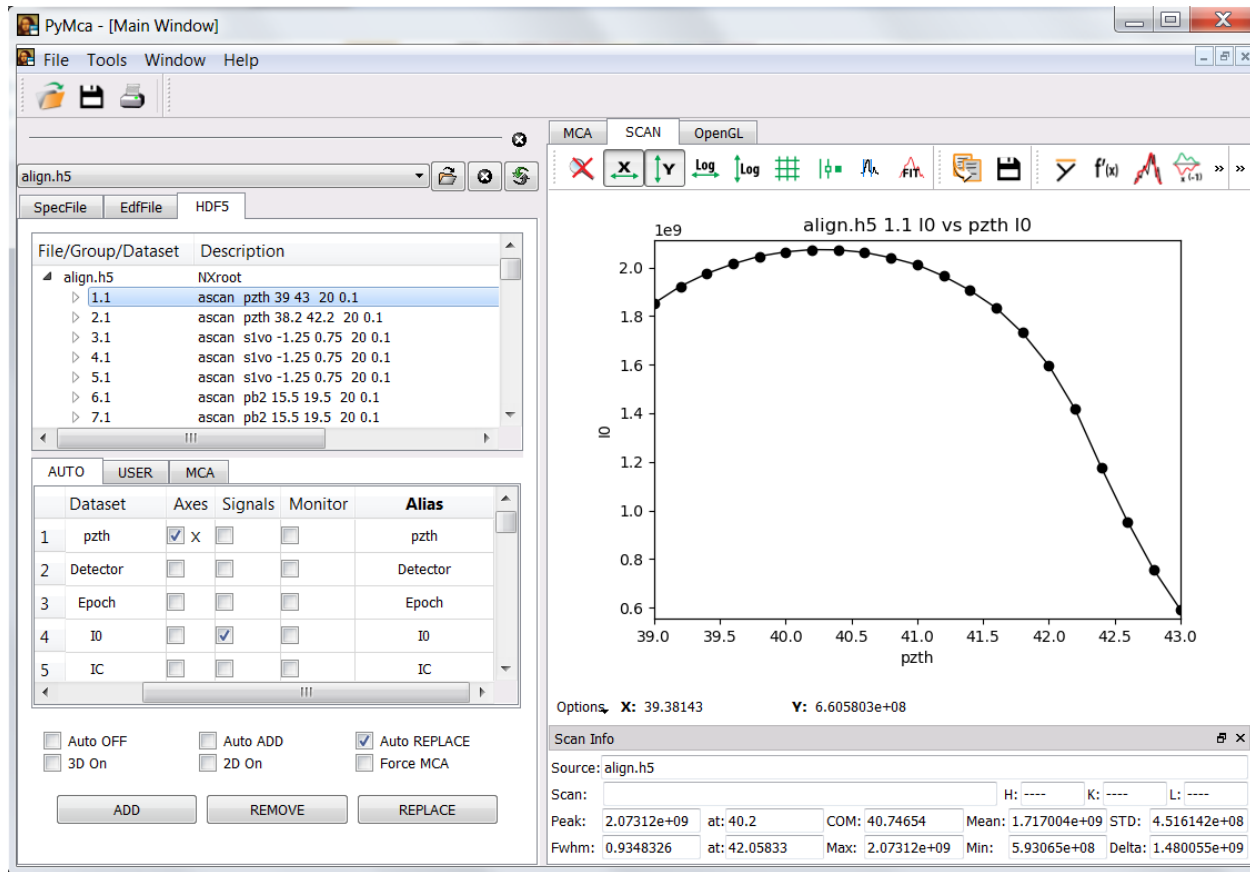
User/Scientist Domain

Administrative Domain (GDPR? DOI?)

Analysis Domain

Measurement Group Convention

- Name-based convention followed by ESRF and Sardana (MAX IV, ALBA...)



- Targets interactive use
- Applications can profit

Application Definitions

- Their goal is to enable interpretation and analysis of the data
- Unfortunately great ideas can be badly implemented

NXxas

entry

definition="NXxas"

start_time

title

instrument@NXinstrument

monochromator@NXmonochromator

energy

incoming_beam@NXdetector

data

absorbed_beam@NXdetector

data

....

Application Definitions

- In 2010 it was communicated to the NIAC that multiple techniques were quite common (SAXS/WAXS, FLUO/DIFF,...)
- It was decided to create a new field NXsubentry containing the relevant information for each technique

ESRF only considers application definitions in subentries

- But the NIAC kept imposing all the rest of the structure in the subentry (see nexusformat.org documentation on NXsubentry)

The actual analysis applications do not need the structure !!!!!

Application Definitions as Understood by the ESRF

- Only the relevant part for the analysis required
- If they are actual items or links to items is irrelevant
- If there are no programs using the definitions the later are useless
- They should come from developers or communities (not just NIAC)

NXxas

entry

whatever_name@NXsubentry

definition="NXxas"

energy

i0

it

DISCLAIMER

I'm not advocating the use of NXxas (in any of the shown forms)

HDF5/NeXus: Requirements for Processed Data

- NeXus conventions are fairly clear in what concerns raw data
- How to store processed data in HDF5 files?
 - Needs
 - Program used
 - Configuration parameters
 - Results
 - Minimize file creation
 - More than one data treatment step into the file
 - Describe data treatment sequence

NeXus: ESRF Implementation for Processed Data (v1)

- Goals can be achieved with “extended” NXprocess groups

entry

start_time

end_time

title

process_name@NXprocess

program_name

version

date

sequence_index

configuration@NXcollection if HDF5 supported by program

configuration@format=“ini” or “json” or ... if string

results@NXcollection or NXdata if plot

Just a name based convention added to NXprocess

NeXus: ESRF Implementation for Processed Data (v2)

- A 100% pure NeXus way to specify the configuration: NXnote entry

process_name@NXprocess

program_name

version

date

sequence_index

configuration@NXnote

file_name

type

data

results@NXcollection or NXdata if just a plot

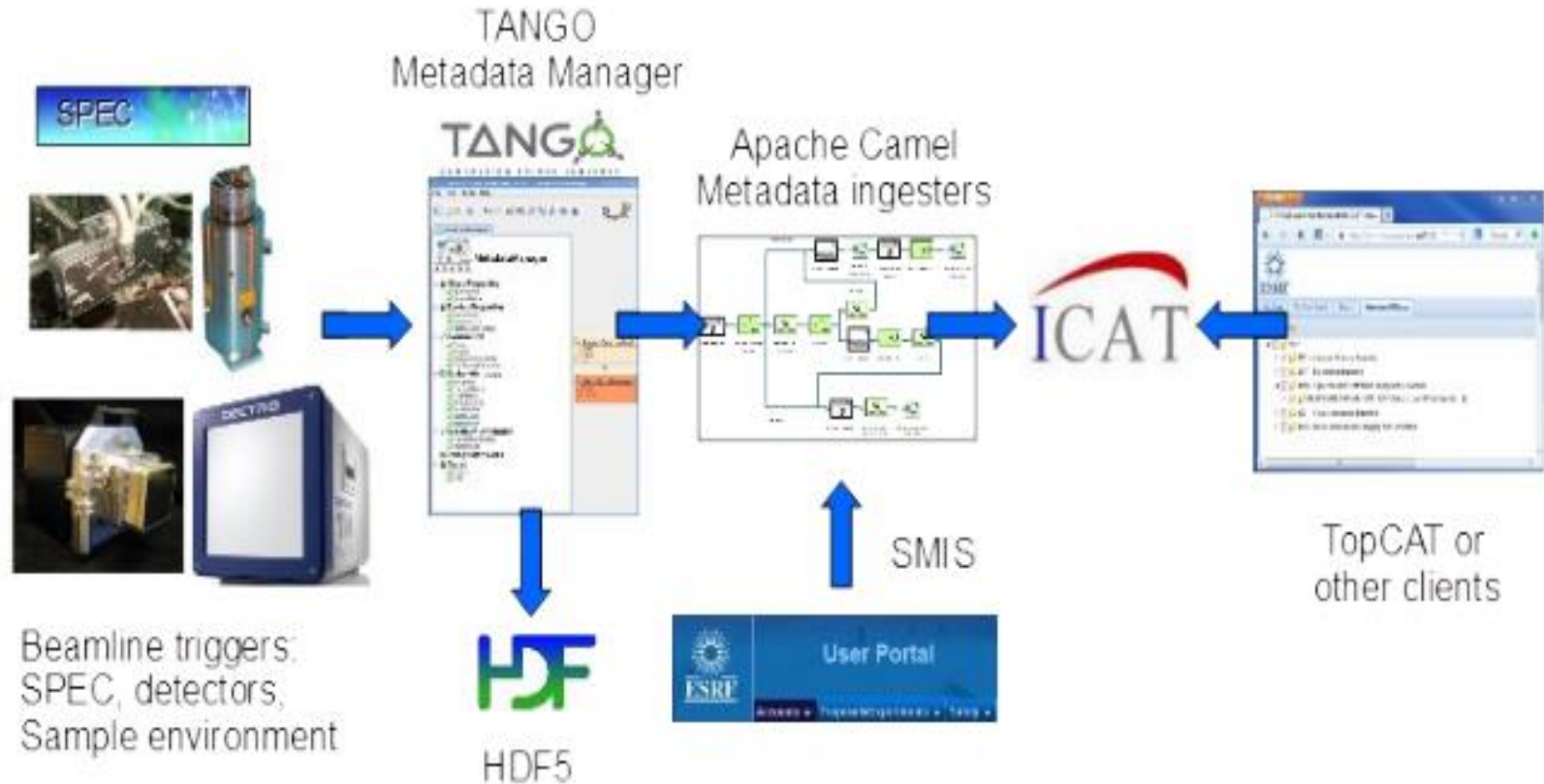
**The key point is that the configuration can be used back.
We have to be able to feed the original program with it.**



NeXus

ICAT

Metadata Collection Architecture



ICAT – NeXus Mapping

- Clear mapping from existing NeXus conventions to ICAT
 - ICAT key = Class1Class2Class3_dataset@attribute
 - NeXus current and mode in class Source inside class Instrument:
 - InstrumentSource_current
 - InstrumentSource_mode
- Technique or beamline specific information as NXsubentry based keys

```
<group NX_class="NXsubentry" groupName="EM">
  <protein_acronym ESRF_description="Protein acronym" NAPIttype="NX_CHAR">${EM_protein_acronym}</protein_acronym>
  <voltage ESRF_description="Voltage" NAPIttype="NX_CHAR">${EM_voltage}</voltage>
  <magnification ESRF_description="Magnification" NAPIttype="NX_CHAR">${EM_magnification}</magnification>
  <images_count ESRF_description="Number of images in movie" NAPIttype="NX_CHAR">${EM_images_count}</images_count>
  <position_x ESRF_description="Position X" NAPIttype="NX_CHAR">${EM_position_x}</position_x>
  <position_y ESRF_description="Position Y" NAPIttype="NX_CHAR">${EM_position_y}</position_y>
  <dose_initial ESRF_description="Dose initial" NAPIttype="NX_CHAR">${EM_dose_initial}</dose_initial>
  <dose_per_frame ESRF_description="Dose per frame" NAPIttype="NX_CHAR">${EM_dose_per_frame}</dose_per_frame>
  <spherical_aberration ESRF_description="Spherical aberration" NAPIttype="NX_CHAR">${EM_spherical_aberration}</spherical_aberration>
  <amplitude_contrast ESRF_description="Amplitude contrast" NAPIttype="NX_CHAR">${EM_amplitude_contrast}</amplitude_contrast>
  <sampling_rate ESRF_description="samplingRate" NAPIttype="NX_CHAR">${EM_sampling_rate}</sampling_rate>
</group>
```

ICAT – NeXus Mapping

- Collected files as list inside an NXcollection group named measurement

The screenshot displays a NeXus data browser interface. On the left, a hierarchical tree shows the structure of the data, with the 'measurement' group highlighted. The tree includes nodes for 'MX', 'definition', 'end_time', 'folder_path', 'Instrument', 'attenuator', 'positioners', 'detector01', 'detector02', 'insertion_device', 'gap', 'taper', 'monochromator', 'crystal', 'd_spacing', 'reflection', 'type', 'usage', 'energy', 'wavelength', 'name', 'optics', 'primary_slit', 'secondary_slit', 'slits', 'source', 'measurement', 'files', and 'proposal'.

On the right, a 'Text' window displays a list of files collected within the 'measurement' group. The list is organized into columns: 'Text' (line number), 'files' (file path), and 'Dat' (data type). The files are listed in a table format, showing the path to each file and its corresponding data type.

Text	files	Dat
0	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/ev325-CG_summer-CG_summer_hires1.h5	
1	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0028_0000_0000.edf	
2	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0131.edf	
3	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0170.edf	
4	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0018.edf	
5	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0052_0000_0000.edf	
6	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0024.edf	
7	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0105.edf	
8	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0181.edf	
9	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0038_0000_0000.edf	
10	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0090.edf	
11	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0083.edf	
12	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0076.edf	
13	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0187.edf	
14	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0096.edf	
15	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0026_0000_0000.edf	
16	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0130.edf	
17	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0160.edf	
18	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0012_0000_0000.edf	
19	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0042.edf	
20	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0038.edf	
21	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0172.edf	
22	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0137.edf	
23	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0194.edf	
24	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0097.edf	
25	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0027_0000_0000.edf	
26	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0164.edf	
27	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0119.edf	
28	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0109.edf	
29	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0009.edf	
30	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0059.edf	
31	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0113.edf	
32	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0032_0000_0000.edf	
33	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0058.edf	
34	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0045_0000_0000.edf	
35	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0171.edf	
36	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0034_0000_0000.edf	
37	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0006_0000_0000.edf	
38	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0137.edf	
39	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0164.edf	
40	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0021.edf	
41	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xmap_x2c_00_0001_0000.edf	
42	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0006.edf	
43	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0000.edf	
44	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0022_0000_0000.edf	
45	/data/visitor/ev325/ld21/CG_summer/CG_summer_hires1/zap/CG_summer_hires1_xia00_0001_0000_0086.edf	

FileEditViewHistoryBookmarksToolsHelp

ESRF Portal

https://data.esrf.fr/investigation/135816585/datasets

Search

ABP

Log out V. Armando SOLE

Closed Data

Investigations

EV-280

Beneficial symbiosis in tomato plants: its role on Fe translocation and speciation

Dataset List90

Logbook

Search

	Date ▾▴	Name ▾▴	Definition ▾▴	Files ▾▴	Size ▾▴	Download ▾▴	
<input type="checkbox"/>	10:16 Nov 5, 2018	fe2streptor2_XAScalib	SXM	5	5 MB	Download	
<input type="checkbox"/>	00:16 Nov 5, 2018	fe2streptor2_main_root	SXM	1343	2 GB	Download	

Summary

Instrument

Metadata List

Files1343

DOI

Monochromator

Energy	7.21972
Wavelength	1.7173
d_spacing	3.13542
Reflection	111
Type	Si
Usage	Bragg

European Synchrotron Radiation Facility

- Acquisition
 - SPEC
 - Not worth native output. Use *silx convert* if desired
 - Bliss
 - NeXus native output operational but concurrent access issues
 - Studying to externalize via REDIS + memcached
 - Data Analysis and not Control responsible of data writing?

Status of NeXus @ ESRF – Data Analysis

- Data Analysis
- ✓ • Capability to read HDF5 files (preferred data analysis I/O format)
- ✓ • Unified API to access all data formats
- ✓ • Support of NeXus NXdata I/O in viewers and analysis codes
- ✓ • Provide provenance via NXprocess (pyFAI, PyMca, PyNX,...)
- Only one NeXus application definition supported (NXcxi)

- Data Policy and NeXus
 - Mirror ICAT and NeXus master file done
 - External links between master file and raw HDF5 files desirable

Ideally one should aim at processing a dataset from its master file

The Weight of Legacy

Adoption of HDF5/NeXus has been slower at the ESRF than at other synchrotrons due to the raw data being acquired in different formats. Detector output in HDF5 and the deployment of Bliss are speeding things up.

User experience with HDF5 files has to be better than with legacy formats HDF5 should not be the question but the answer.

Concerning data analysis, ESRF started to provide HDF5 support in 2009. Currently making convenient **use of the NeXus formalism as output and as integral part of the ESRF data policy.**

Thank you for your attention!

