HUN-REN
Hungarian Research Network
Research.
Innovation.
Impact.

Standards-Compatible Data Storage in Laboratory-Based X-ray Instruments

András Wacha

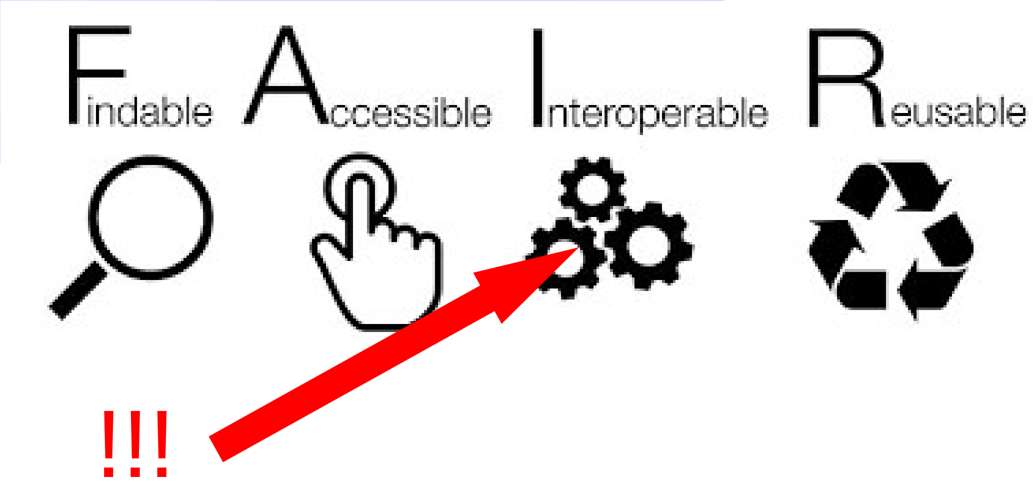
HUN-REN | ttk

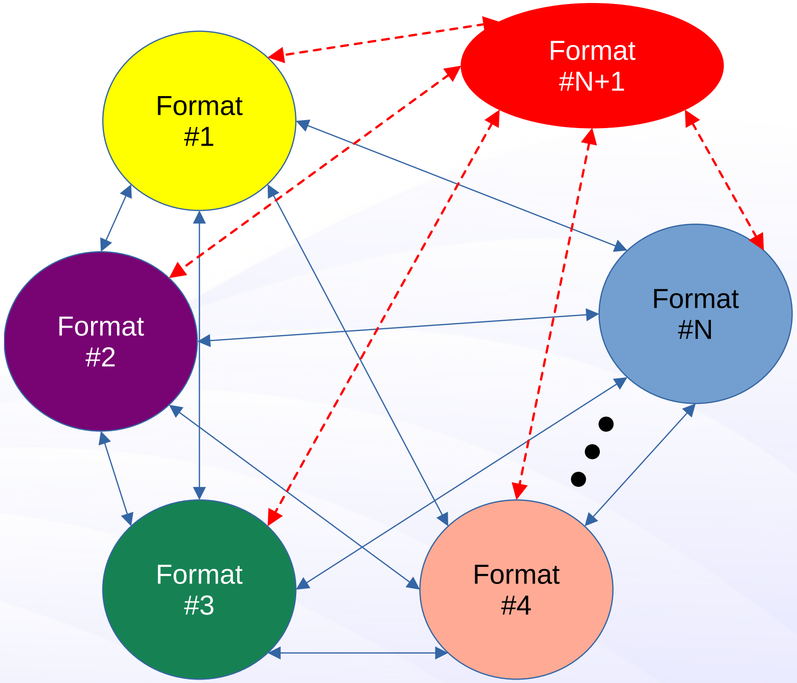
MTA
Centre of Excellence

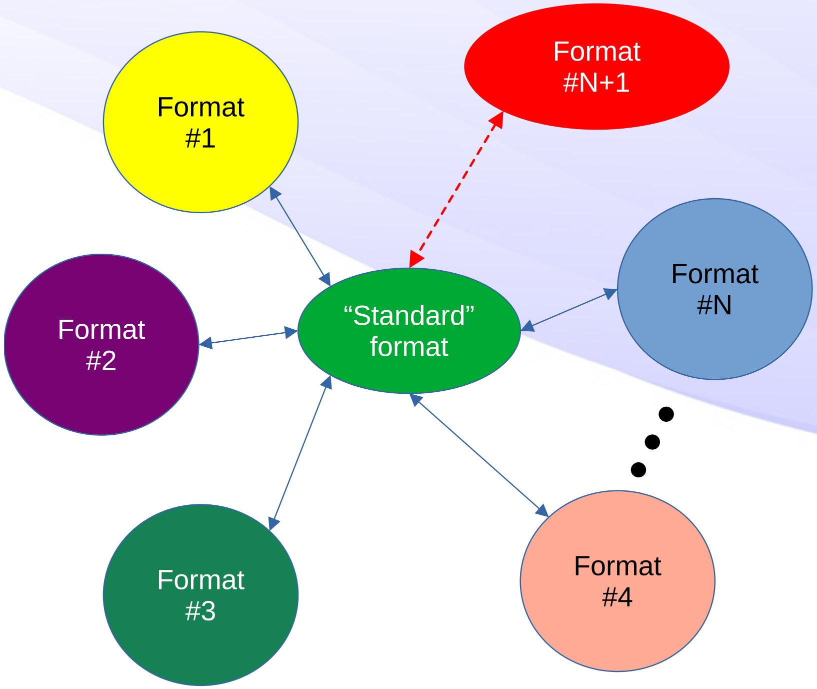
EXCELLENT RESEARCH INFRASTRUCTURE
HUNGARY

Motivation: Standardized Data Storage Formats

- "It's good that we have so many standards to choose from"
- The evolution of storage formats together with hardware and software
- The question of interoperability
- Without standardization:
  - N formats, N×(N-1) converters
  - New format: 2N new converters needed
- With a standard format agreed upon:
  - N formats, 2N converters.
  - New format: 2 new converters needed
- FAIR principles of open data

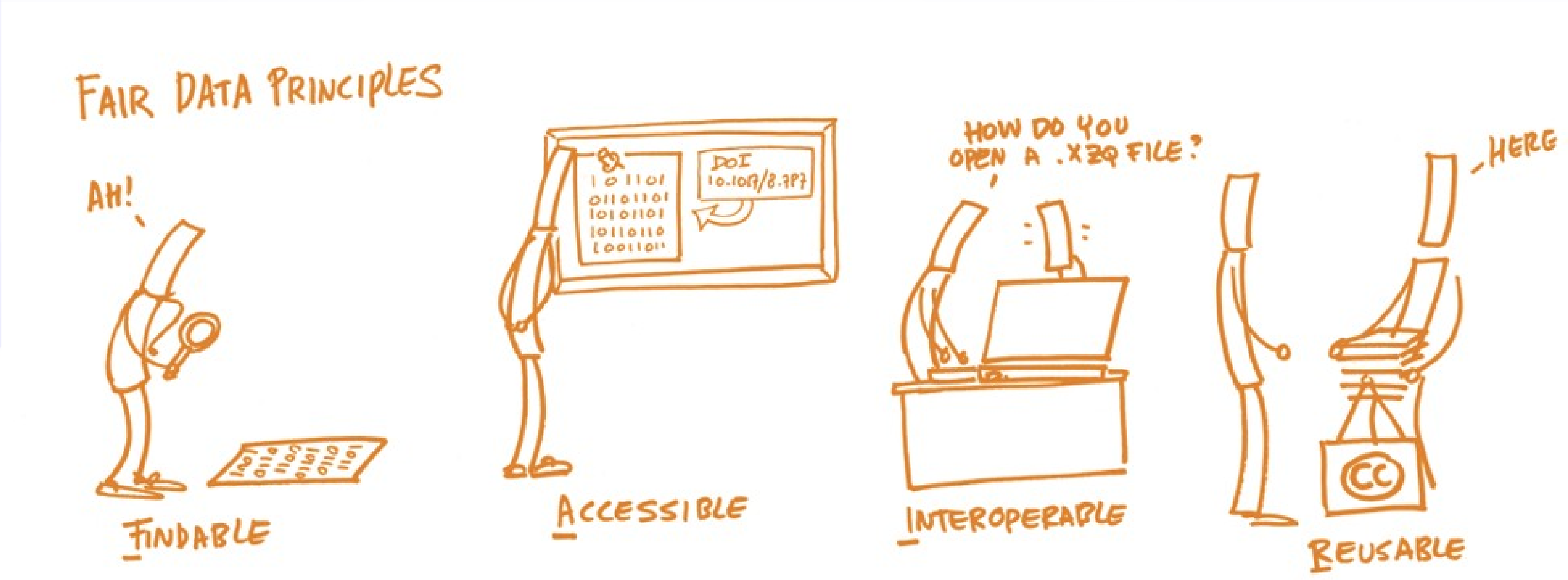




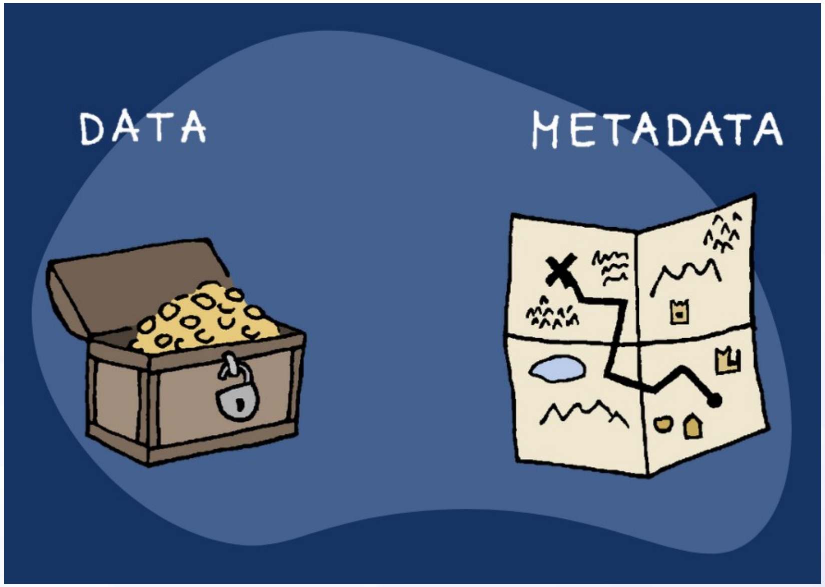


FAIR Principles in the Practice (https://www.openaire.eu/what-is-fair-data)

- **Findable:** “Discoverable with metadata, identifiable and locatable by means of a standard identification mechanism”
- **Accessible:** “Always available and obtainable; even if the data is restricted, the metadata is open”
- **Interoperable:** “Both syntactically parseable and semantically understandable, allowing data exchange and reuse between researchers, institutions, organizations or countries”
- **Reusable:** “Sufficiently described and shared with the least restrictive licences, allowing the widest reuse possible and the least cumbersome integration with other data sources”



Data and metadata, i.e., what to store



- Primary data: what we're most interested in
  - Typically non-scalar, possibly N-dim
  - Dependent variable as a function of the independent variable(s)
  - Fluorescence/absorption spectrum
  - Scattering pattern, scattering curve
  - Image
  - …
  - Raw vs. processed
- Metadata: describing the primary data
  - Information about:
    - The sample (name, composition, longer description…)
    - The instrument (name, state variables…)
    - Experimental conditions (temperature, *in situ* parameters)
    - Data reduction/evaluation procedures used
  - Ensuring correctness, reliability and reproducibility of the primary data

FAIR Data in Photon (and Neutron) Science

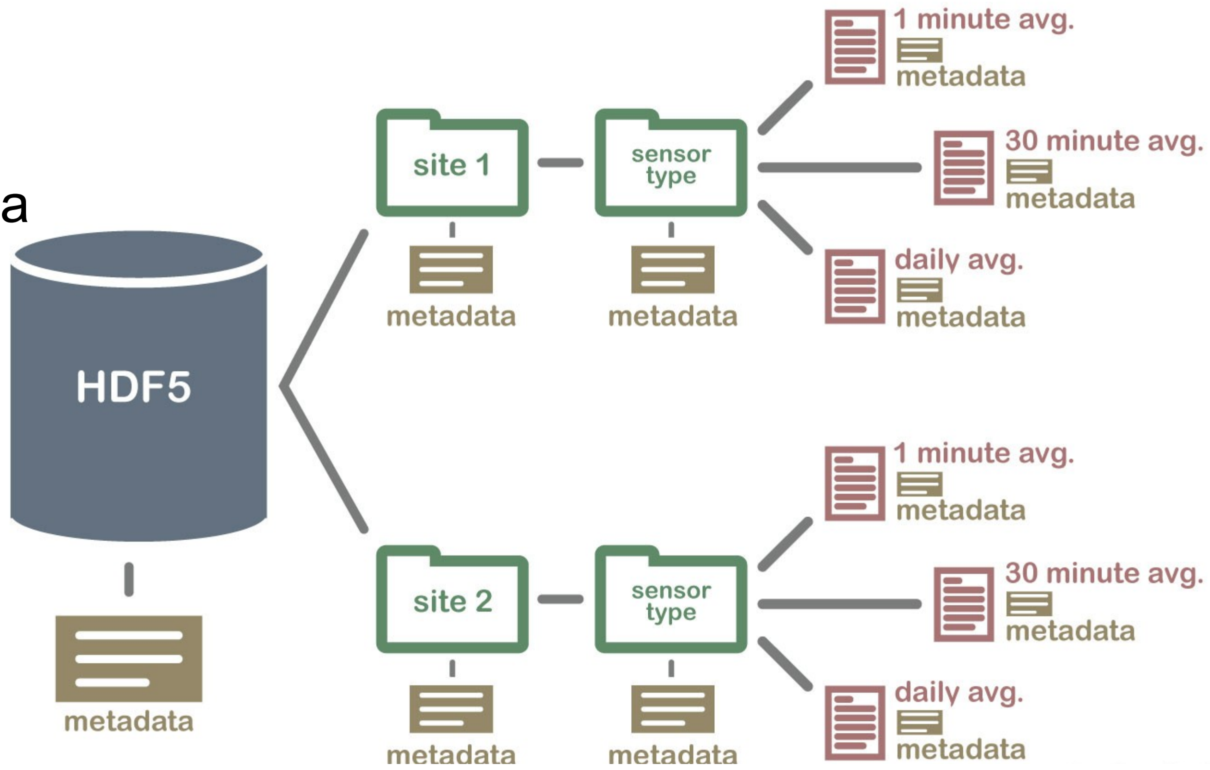
- Photon and neutron open science cluster: PaNOSC (https://www.panosc.eu)
  - Part of the European Open Science Cloud (EOSC)
  - PaNOSC project: EC-financed, 2018-2022
  - Representing European photon and neutron research infrastructures
  - PaN-data Europe Deliverable D2.1: Common policy framework on scientific data
    - a generic data management policy
    - can be tailored by facilities to their own needs
  - Recommends the NeXus/HDF5 format for storing data and metadata
- CanSAS (https://www.cansas.org)
  - “collective action for nomadic small angle scatterers”
  - Providing the small-angle scattering user community with shared tools and information
  - First meeting: 1998
  - N-dimensional data: NeXus-based (NXcanSAS) 2017-06-06 (announced)
  - 1D data: XML-based format
    - cansas1d (v1.0: 2009-05-12, v1.1: 2013-03-29)
    - Recommendation from early 2017 on: store 1D also as NxcanSAS

How to choose a data format?

- Home-brewn format
  - Easy to implement if you are in a hurry, but usually not *future-proof*
  - Not well thought-out (not all relevant aspects are stored)
  - Difficult to extend (becomes “patchy”, needs workarounds)
  - Difficult to maintain (protection against hardware and software obsolescence)
- Choose from an already existing standard
  - Takes efforts to implement
  - Be kind to your “nomadic users”
  - Good chances that many pieces of software already support it
- Other requirements
  - **Open:** specification, algorithms, libraries freely available
  - **Self-describing:** data are labeled, intuitively stored
  - **Compression:** store large datasets, preferably in a seamless way
  - **Fast read/write:** high throughput
  - **Fault tolerance / detection:** redundancy, checksums
  - **Straightforward API:** easy to access the data in many programming languages (scientists are not programmers)

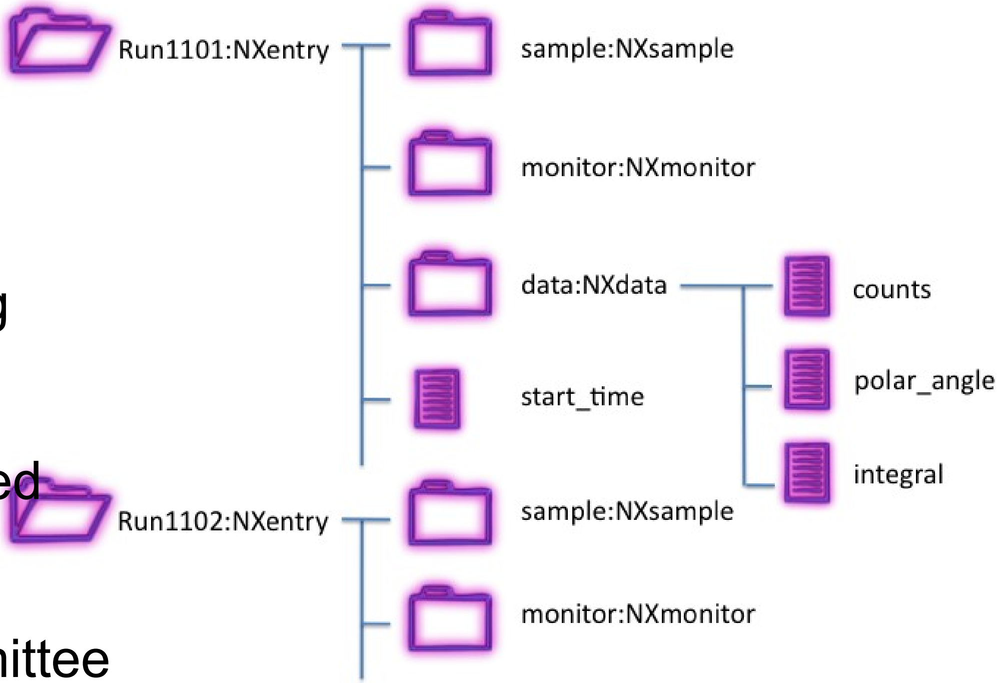
Hierarchical Data Format (https://www.hdfgroup.org/solutions/hdf5/)

- Binary container format
- Cross-platform
- Many programming languages supported (C, C++, Java, Python…)
- Tree structure
  - Groups (~ folders)
    - containers
  - N-dimensional datasets (~ files)
    - N ≥ 0
    - Size, shape, data type
  - Symbolic or hard links
    - Pointers to the same data
  - External links
- Metadata (attributes) can be attached to groups and datasets
- Transparent IO filters
  - Compression
  - Shuffling
  - Fletcher32 checksum



The NeXus Format

- *De facto* standard in photon, neutron and muon sciences
  - PaNOSC, canSAS…
- Common data exchange format
- Aims
  - Container for raw data, associated with a scientific instrument
  - Container for processed data
- Based on the HDF5 format
  - HDF5 is only a container
  - standardized structure
  - Standardized nomenclature
- Aims:
  - Domain-specific rules for organizing and arranging data
  - Quick default visualization
  - Standard definitions that can be used to validate files
- Governing body: NIAC
  - Nexus International Advisory Committee
  - Convenes every other year
- https://www.nexusformat.org
- J Appl Crystallogr (2015) 48(1) 301-305 (doi:10.1107/S1600576714027575)

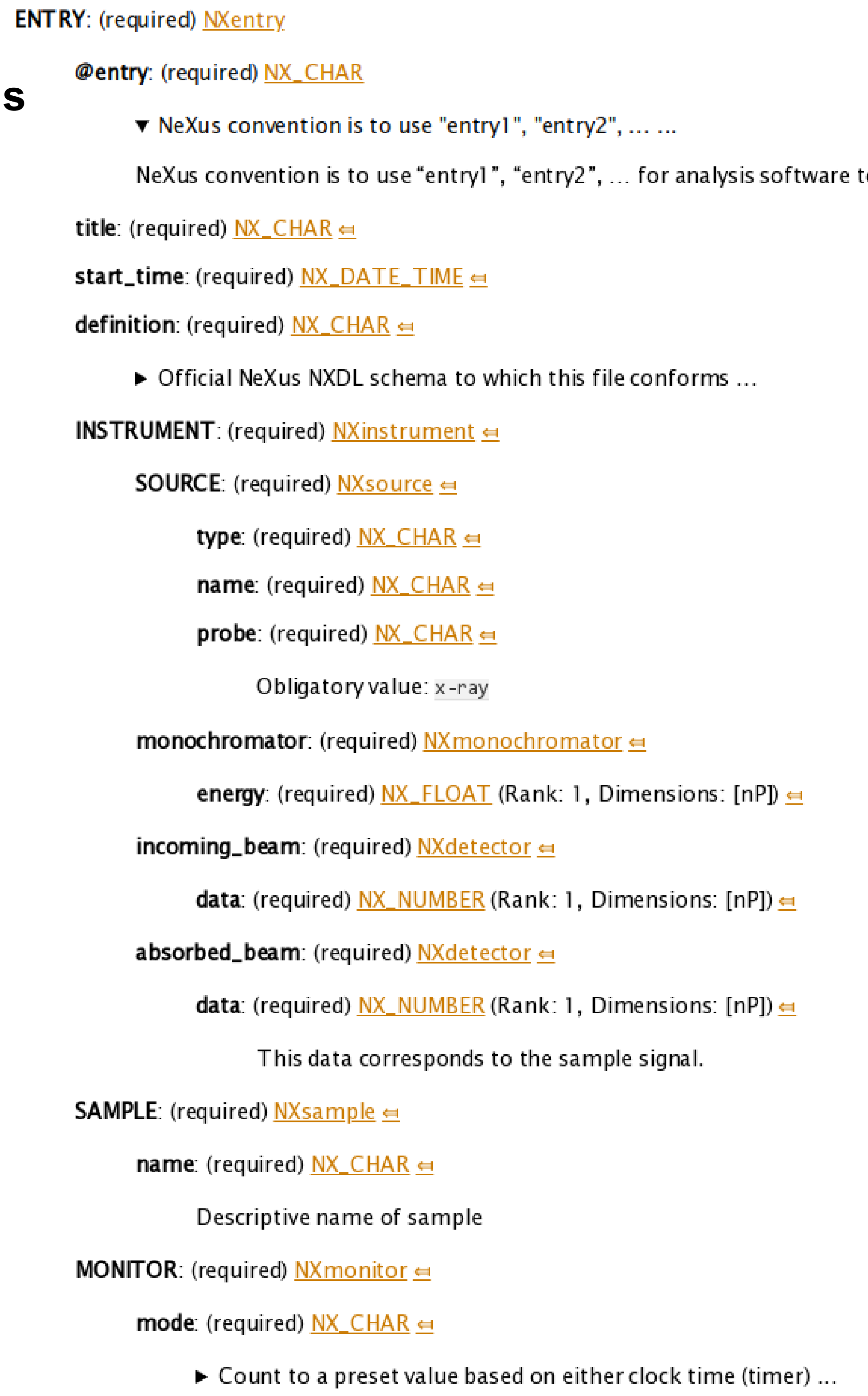


NeXus Base Concepts I: NeXus base classes

- The “type” of a HDF5 group, stored in the “NX_class” attribute of the group
- Prescription of the possible fields (datasets) and attributes
  - Field types are also specified
- Corresponds to real-world objects (sample, instrument, data reduction step…)
- https://manual.nexusformat.org/classes/base_classes/index.html
- Example 1: NXentry (describe the measurement)
  - @NX_class=”NXentry” (attribute)
  - title (NX_CHAR): title of the entry
  - start_time (NX_DATE_TIME): starting time of the measurement
  - end_time (NX_DATE_TIME): ending time of the measurement
  - program_name (NX_CHAR): the program used to generate this file
  - Sample (NXsample): a group describing the sample
  - Instrument (NXinstrument): a group describing the instrument
  - ...
- Example 2: NXsample (describe the sample)
  - @NX_class=”NXsample” (attribute)
  - name (NX_CHAR): name of the sample
  - chemical_formula (NX_CHAR): chemical formula
  - temperature (NX_FLOAT): temperature of the sample
  - ...

NeXus Base Concepts II: Application definitions

- Base classes only define the nomenclature, not requirements
- Application definitions: domain-specific rules on obligatory and optional data
- Declared in the “definition” field of the NXentry
- Example: NXxas (X-ray absorption spectroscopy measurements)
- https://manual.nexusformat.org/classes/applications/NXxas.html
- Lower case names: required name
- Upper case names: arbitrary name
- Path specification in the file:
  - By name:
    - entry145/vonhamos/xraygenerator/name
  - Using NXclass attributes:
    - Nxentry/Nxinstrument/Nxsource/name
- Multiple methods (e.g. raw SAXS and processed SAXS): NXsubentry



ENTRY: (required) NXentry

- **@entry**: (required) NX_CHAR
  - ▼ NeXus convention is to use "entry1", "entry2", … …
  - NeXus convention is to use “entry1”, “entry2”, … for analysis software t
- **title**: (required) NX_CHAR
- **start_time**: (required) NX_DATE_TIME
- **definition**: (required) NX_CHAR
  - ▶ Official NeXus NXDL schema to which this file conforms …
- **INSTRUMENT**: (required) NXinstrument
  - **SOURCE**: (required) NXsource
    - **type**: (required) NX_CHAR
    - **name**: (required) NX_CHAR
    - **probe**: (required) NX_CHAR
      - Obligatory value: `x-ray`
  - **monochromator**: (required) NXmonochromator
    - **energy**: (required) NX_FLOAT (Rank: 1, Dimensions: [nP])
  - **incoming_beam**: (required) NXdetector
    - **data**: (required) NX_NUMBER (Rank: 1, Dimensions: [nP])
  - **absorbed_beam**: (required) NXdetector
    - **data**: (required) NX_NUMBER (Rank: 1, Dimensions: [nP])
      - This data corresponds to the sample signal.
- **SAMPLE**: (required) NXsample
  - **name**: (required) NX_CHAR
    - Descriptive name of sample
- **MONITOR**: (required) NXmonitor
  - **mode**: (required) NX_CHAR
    - ▶ Count to a preset value based on either clock time (timer) …

NeXus Base Concepts III: Contributed definitions

- Tentative, suggested extensions to the NeXus specification
- Proposed by the community
- Not yet standardized
- Both base classes and application definitions
- Example fields:
  - Optical spectroscopy
  - Multi-dimensional photoemission spectroscopy
  - Atom probe microscopy
  - Electron microscopy
  - …
- Curated, commented on and finally incorporated into the NeXus standard by the NIAC

NeXus Base Concepts IV: Default Visualization

- Each measurement should have a default visualization...
- … which should be declared in the data file
- The NXdata class
  - “Encapsulates all the information required for a set of data to be plotted”
  - *Signals*: dependent variables (1- or more dimensions)
    - Default signal: “signal” attribute
  - *Axes*: independent variables (typically 1D, but can be more)
  - Names freely chosen (but *cf* application definitions)
  - https://manual.nexusformat.org/classes/base_classes/NXdata.html
  - Example:

```
data: NXdata
  @signal = “data”
  @axes = [“x”, “y”]
  data: float[10, 20]
  x: float[10]
  y: float[20]
```

```
data: NXdata
  @signal = “data”
  @axes = [“x”, “y”]
  @x_indices = 0
  @y_indices = 1
  data: float[10, 20]
  x: float[10]
  y: float[20]
```

More complex NXdata

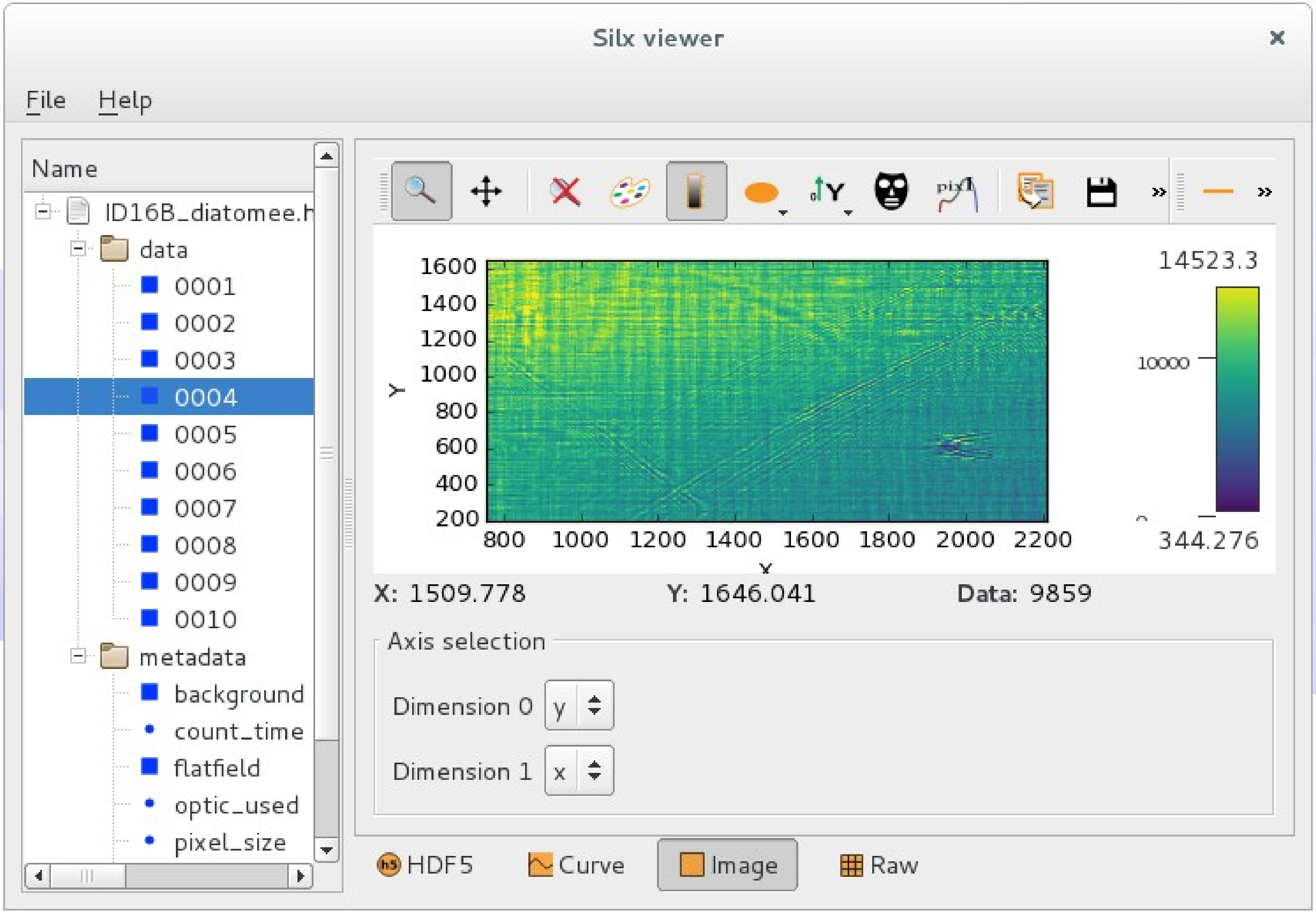
- Multi-dimensional data, e.g., scans

```
data:NXdata
  @signal = “data”
  @auxiliary_signals = [“data2”, “data3”]
  @axes = [“x”, “y”, “energy”, “wavelength”]
  @x_indices = 0
  @y_indices = 1
  @energy_indices = 2
  @wavelength_indices = 2
  data: float[10, 20, 30]
  x: float[10]
  y: float[20]
  energy: float[30]
  wavelength: float[30]
  data2: float[10, 20, 30]
  data3: float[10, 20, 30]
```

```
data:NXdata
  @signal = “absorption”
  @axes = [“xpos”, “ypos”, “energy”]
  @xpos_indices = [0, 1, 2]
  @ypos_indices = [0,1,2]
  @energy_indices = 2
  absorption: float[10, 20, 30]
  xpos: float[10, 20, 30]
  ypos: float[10, 20, 30]
  energy: float[30]
  absorption_errors: float[10, 20, 30]
  xpos_errors: float[10, 20, 30]
  ypos_errors: float[10, 20, 30]
```

NeXus utilities – Silx view

- Displays HDF5 tree and the default (customizable) plot
- https://www.silx.org



X: 1509.778 Y: 1646.041 Data: 9859

Dimension 0 y

Dimension 1 x

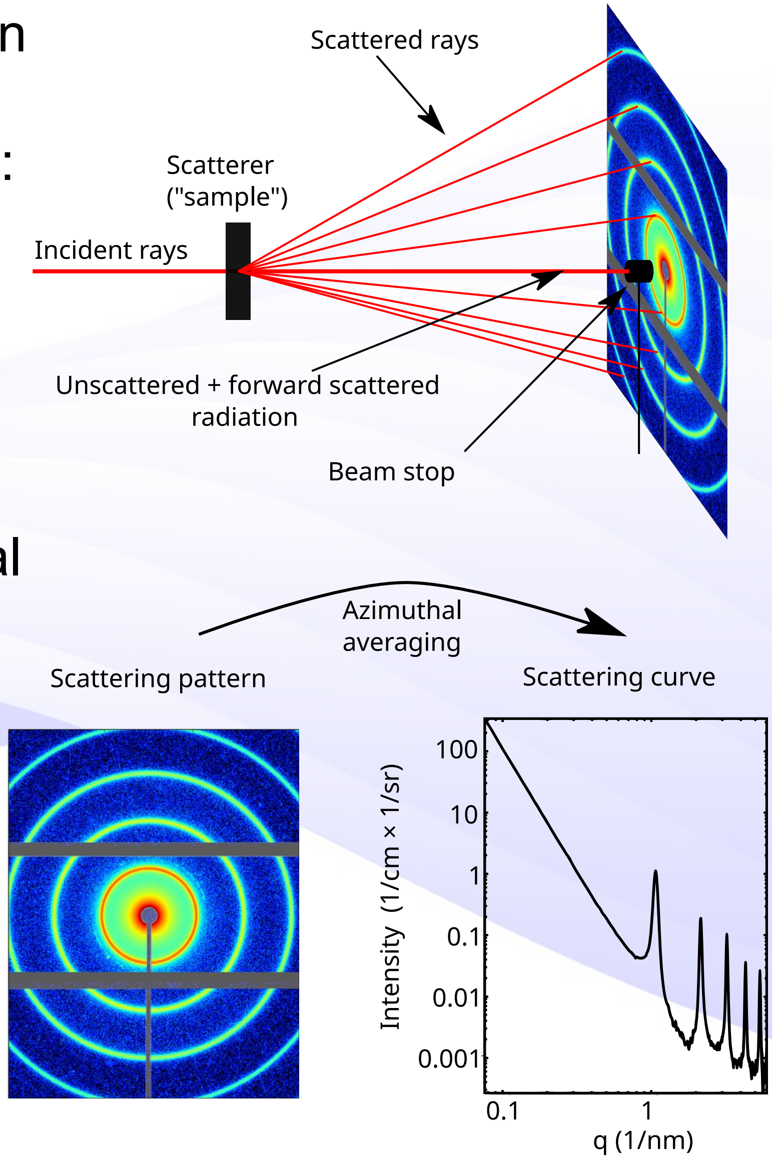
Other NeXus utilities: programming & whatnot

- Reading and writing in Python
  - h5py (https://www.h5py.org): HDF5 wrappers for Python
  - PyTables (https://www.pytables.org): alternate HDF5 wrappers, sposnored by NUMFOCUS (Numpy, Scipy & Co.)
  - NeXpy (https://nexpy.github.io/nexpy/): high-level Python interface (+ GUI)
- Validation + other utilities
  - Punx (https://github.com/prjemian/punx)
  - NeXus command-line utilities: nxbrowse, nxconvert, nxdir…
- Data analysis programs supporting NeXus files:
  - DAVE (https://www.ncnr.nist.gov/dave/): for inelastic neutron scattering
  - DAWN (https://www.dawnsci.org): generic visualization, domain-specific processing
  - Mantid (http://mantidproject.org): high-performance computing on neutron and muon data
  - PyMCA (https://pymca.sourceforge.net): X-ray fluorescence data analysis
  - … (see https://manual.nexusformat.org/utilities.html)

How We Store NeXus Files in the CREDO System?

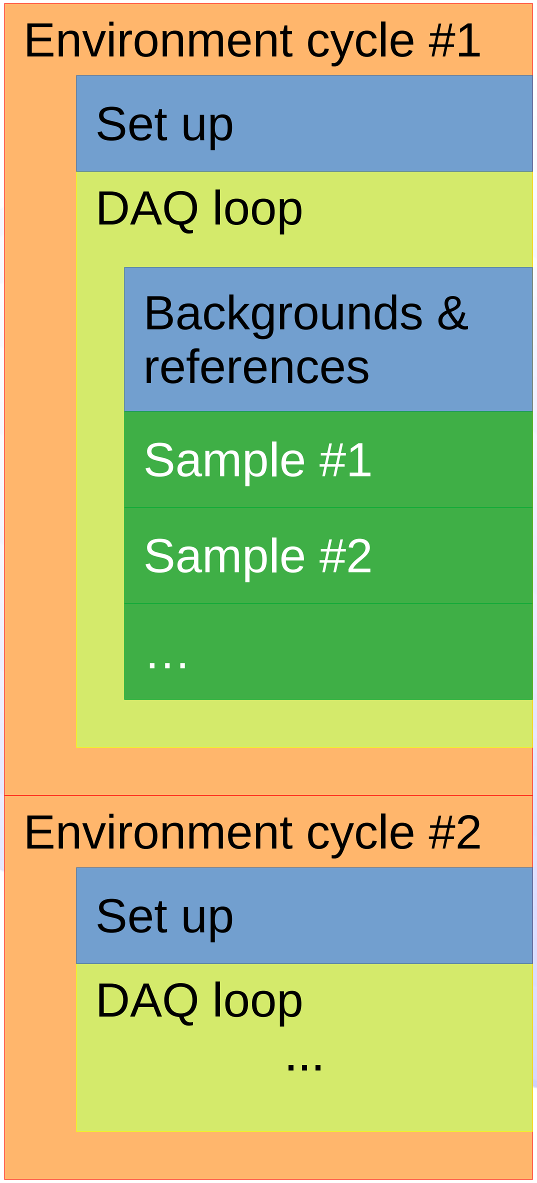
SAXS data

- Modern SAXS measurements: 2D position sensitive detectors
- Isotropic (unoriented, powder, …) sample: dependence only on $q=|\boldsymbol{q}|$.
  - Azimuthal average → radial scattering curve (“intensity vs. *q*”)
- Anisotropic scattering patterns:
  - Azimuthal averages in sectors
  - Radial average in an annulus: azimuthal scattering curve (“intensity vs. *φ*”)
- Typical detector format: ~ 4-9 Mpixel
  - File size ~ 30-50 Mbyte
- Data reduction
  - External and internal background
  - Geometrical corrections
  - Normalization by exposure time and beam intensity
  - Correct for X-ray absorption by the sample



Storing NeXus files

- Application definitions:
  - Raw data: NXsas
  - Processed data: home-brewn (moving to NXcansas)
- https://nexdatas.github.io
  - Developed at DESY
  - A set of Tango servers:
    - NeXus file writer
    - Configuration server for NeXus file writer
    - Component selector
  - GUI
    - Component designer
  - Sardana/Taurus extensions
    - Macro GUI
    - Sardana Recorder which uses the Tango servers
- In-house developed Sardana macros
  - nxsbegin – ct – nxsend: most metadata written while the exposure is being made



Environment cycle #1

Set up

DAQ loop

Backgrounds & references

Sample #1

Sample #2

…

Environment cycle #2

Set up

DAQ loop

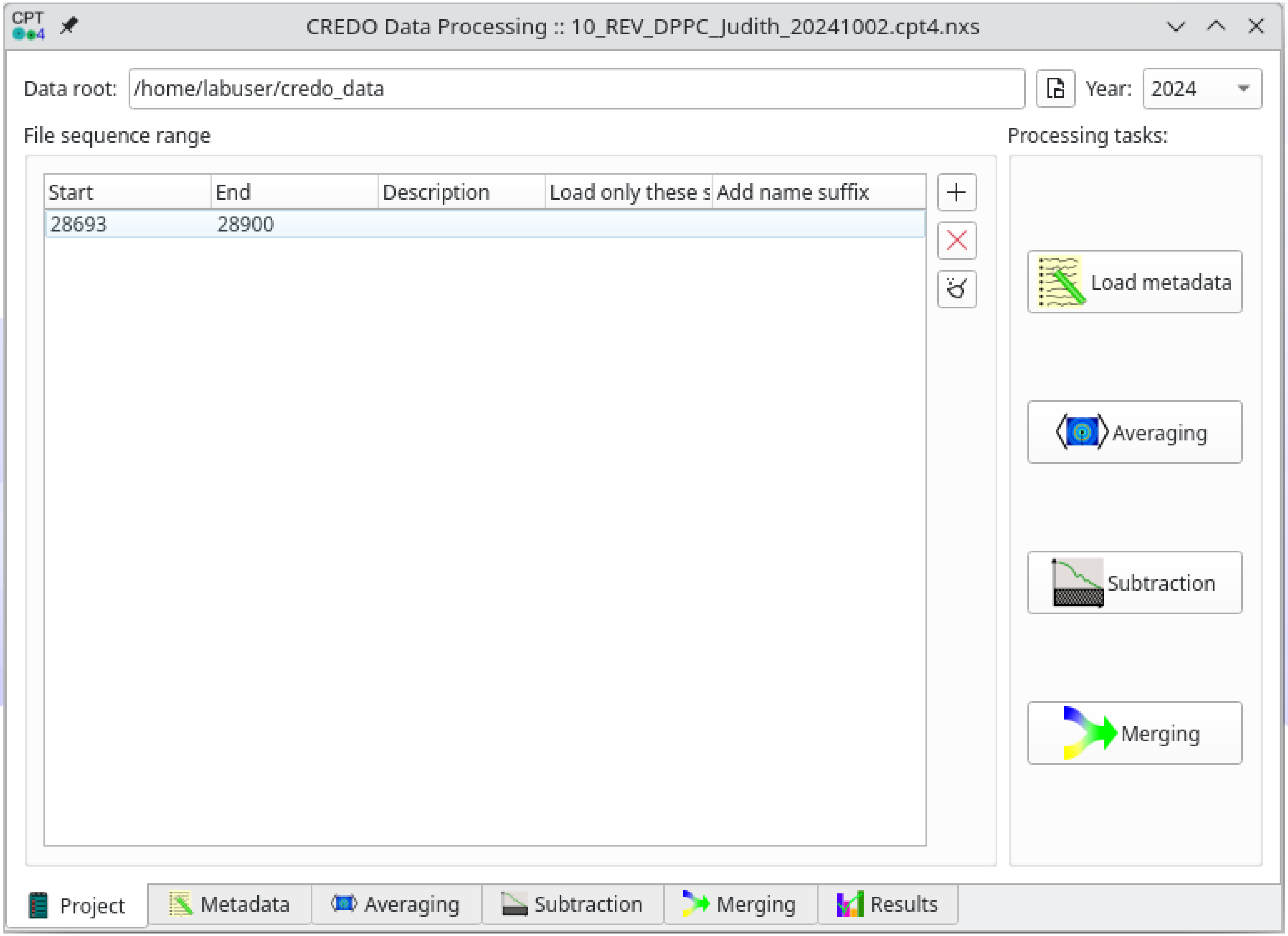
...

Data storage strategy

- File sequence: <prefix>_<file_sequence_index>.<extn>
  - Examples: crd_52133.cbf (raw detector image), crd_52133.nxs (NeXus file, separate NXsubentry for raw and processed)
  - Prefix: independently counted images, based on the use (we use the same 2D detector for everything!):
    - scn: scan measurements
    - tra: transmission measurements
    - crd: “production”, true measurements
    - tst: test shots
    - gsx: GISAXS
  - A SQL database is also written (raw; updated with processed)
    - Basis for **F**indability later on
- End result: a single dataset for each sample
  - ? time-evolution ?
  - Transitioning to NXcansas in the near future

Data post-processing GUI

- Load a range of a file sequence



CREDO Data Processing :: 10_REV_DPPC_Judith_20241002.cpt4.nxs

Data root: /home/labuser/credo_data

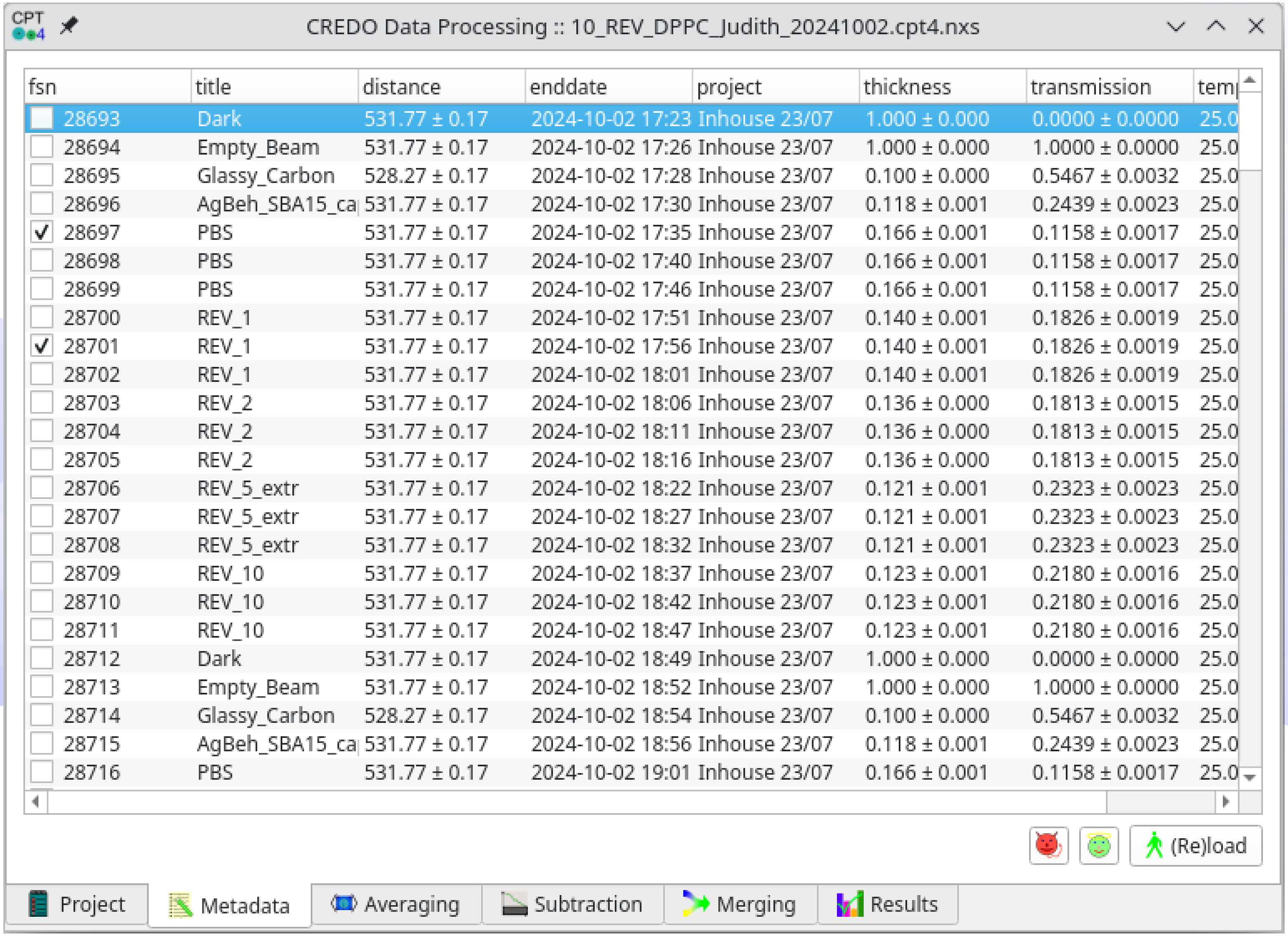
Year: 2024

File sequence range

Start	End	Description	Load only these s	Add name suffix
28693	28900			

Data post-processing GUI

- Analyze metadata

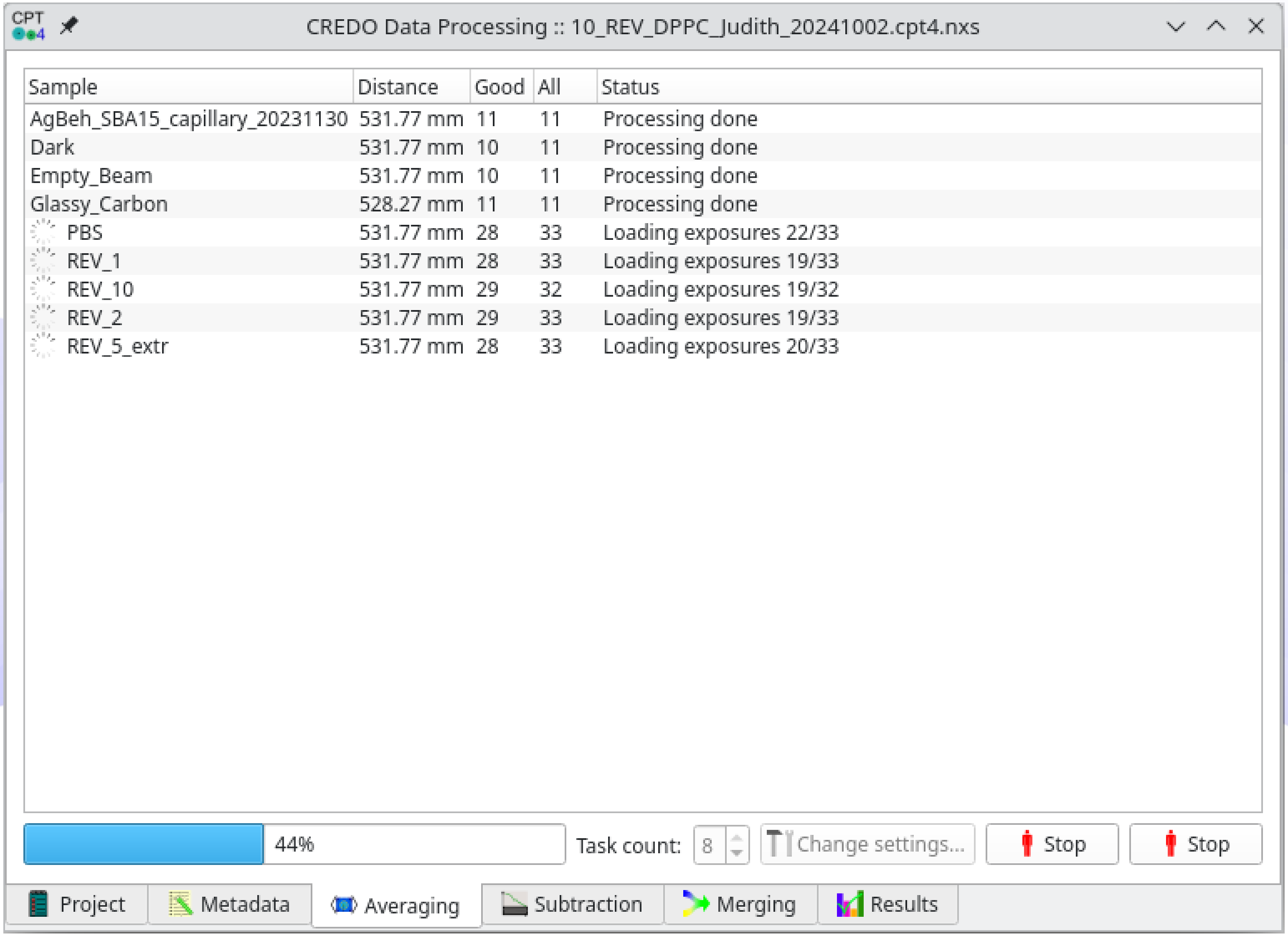


CREDO Data Processing :: 10_REV_DPPC_Judith_20241002.cpt4.nxs

fsn	title	distance	enddate	project	thickness	transmission	tem
☐ 28693	Dark	531.77 ± 0.17	2024-10-02 17:23	Inhouse 23/07	1.000 ± 0.000	0.0000 ± 0.0000	25.0
☐ 28694	Empty_Beam	531.77 ± 0.17	2024-10-02 17:26	Inhouse 23/07	1.000 ± 0.000	1.0000 ± 0.0000	25.0
☐ 28695	Glassy_Carbon	528.27 ± 0.17	2024-10-02 17:28	Inhouse 23/07	0.100 ± 0.000	0.5467 ± 0.0032	25.0
☐ 28696	AgBeh_SBA15_ca	531.77 ± 0.17	2024-10-02 17:30	Inhouse 23/07	0.118 ± 0.001	0.2439 ± 0.0023	25.0
☑ 28697	PBS	531.77 ± 0.17	2024-10-02 17:35	Inhouse 23/07	0.166 ± 0.001	0.1158 ± 0.0017	25.0
☐ 28698	PBS	531.77 ± 0.17	2024-10-02 17:40	Inhouse 23/07	0.166 ± 0.001	0.1158 ± 0.0017	25.0
☐ 28699	PBS	531.77 ± 0.17	2024-10-02 17:46	Inhouse 23/07	0.166 ± 0.001	0.1158 ± 0.0017	25.0
☐ 28700	REV_1	531.77 ± 0.17	2024-10-02 17:51	Inhouse 23/07	0.140 ± 0.001	0.1826 ± 0.0019	25.0
☑ 28701	REV_1	531.77 ± 0.17	2024-10-02 17:56	Inhouse 23/07	0.140 ± 0.001	0.1826 ± 0.0019	25.0
☐ 28702	REV_1	531.77 ± 0.17	2024-10-02 18:01	Inhouse 23/07	0.140 ± 0.001	0.1826 ± 0.0019	25.0
☐ 28703	REV_2	531.77 ± 0.17	2024-10-02 18:06	Inhouse 23/07	0.136 ± 0.000	0.1813 ± 0.0015	25.0
☐ 28704	REV_2	531.77 ± 0.17	2024-10-02 18:11	Inhouse 23/07	0.136 ± 0.000	0.1813 ± 0.0015	25.0
☐ 28705	REV_2	531.77 ± 0.17	2024-10-02 18:16	Inhouse 23/07	0.136 ± 0.000	0.1813 ± 0.0015	25.0
☐ 28706	REV_5_extr	531.77 ± 0.17	2024-10-02 18:22	Inhouse 23/07	0.121 ± 0.001	0.2323 ± 0.0023	25.0
☐ 28707	REV_5_extr	531.77 ± 0.17	2024-10-02 18:27	Inhouse 23/07	0.121 ± 0.001	0.2323 ± 0.0023	25.0
☐ 28708	REV_5_extr	531.77 ± 0.17	2024-10-02 18:32	Inhouse 23/07	0.121 ± 0.001	0.2323 ± 0.0023	25.0
☐ 28709	REV_10	531.77 ± 0.17	2024-10-02 18:37	Inhouse 23/07	0.123 ± 0.001	0.2180 ± 0.0016	25.0
☐ 28710	REV_10	531.77 ± 0.17	2024-10-02 18:42	Inhouse 23/07	0.123 ± 0.001	0.2180 ± 0.0016	25.0
☐ 28711	REV_10	531.77 ± 0.17	2024-10-02 18:47	Inhouse 23/07	0.123 ± 0.001	0.2180 ± 0.0016	25.0
☐ 28712	Dark	531.77 ± 0.17	2024-10-02 18:49	Inhouse 23/07	1.000 ± 0.000	0.0000 ± 0.0000	25.0
☐ 28713	Empty_Beam	531.77 ± 0.17	2024-10-02 18:52	Inhouse 23/07	1.000 ± 0.000	1.0000 ± 0.0000	25.0
☐ 28714	Glassy_Carbon	528.27 ± 0.17	2024-10-02 18:54	Inhouse 23/07	0.100 ± 0.000	0.5467 ± 0.0032	25.0
☐ 28715	AgBeh_SBA15_ca	531.77 ± 0.17	2024-10-02 18:56	Inhouse 23/07	0.118 ± 0.001	0.2439 ± 0.0023	25.0
☐ 28716	PBS	531.77 ± 0.17	2024-10-02 19:01	Inhouse 23/07	0.166 ± 0.001	0.1158 ± 0.0017	25.0

Data post-processing GUI

- Average exposures corresponding to the same sample and same sample-to-detector distance



CREDO Data Processing :: 10_REV_DPPC_Judith_20241002.cpt4.nxs

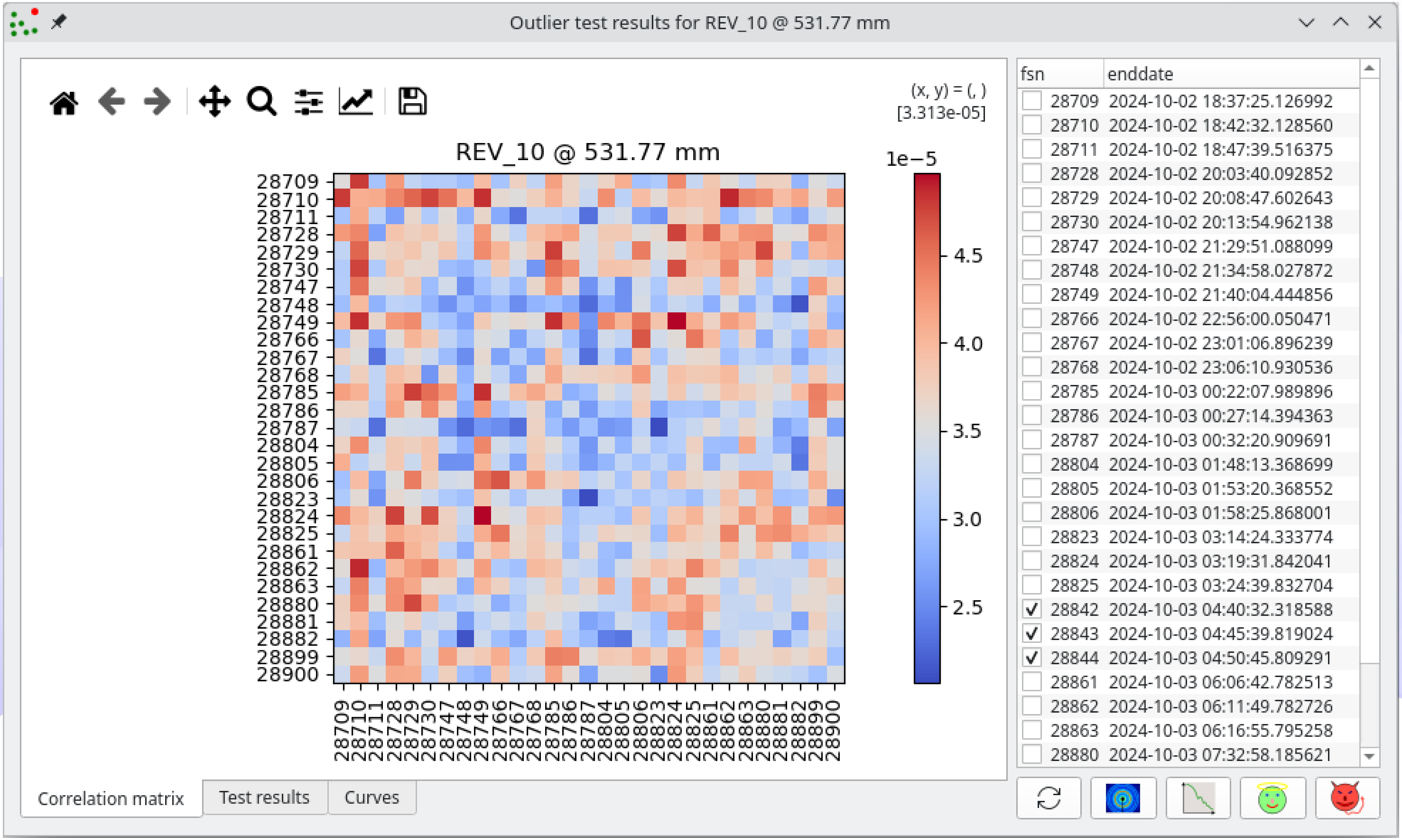
Sample	Distance	Good	All	Status
AgBeh_SBA15_capillary_20231130	531.77 mm	11	11	Processing done
Dark	531.77 mm	10	11	Processing done
Empty_Beam	531.77 mm	10	11	Processing done
Glassy_Carbon	528.27 mm	11	11	Processing done
PBS	531.77 mm	28	33	Loading exposures 22/33
REV_1	531.77 mm	28	33	Loading exposures 19/33
REV_10	531.77 mm	29	32	Loading exposures 19/32
REV_2	531.77 mm	29	33	Loading exposures 19/33
REV_5_extr	531.77 mm	28	33	Loading exposures 20/33

44%

Task count: 8

Data post-processing GUI

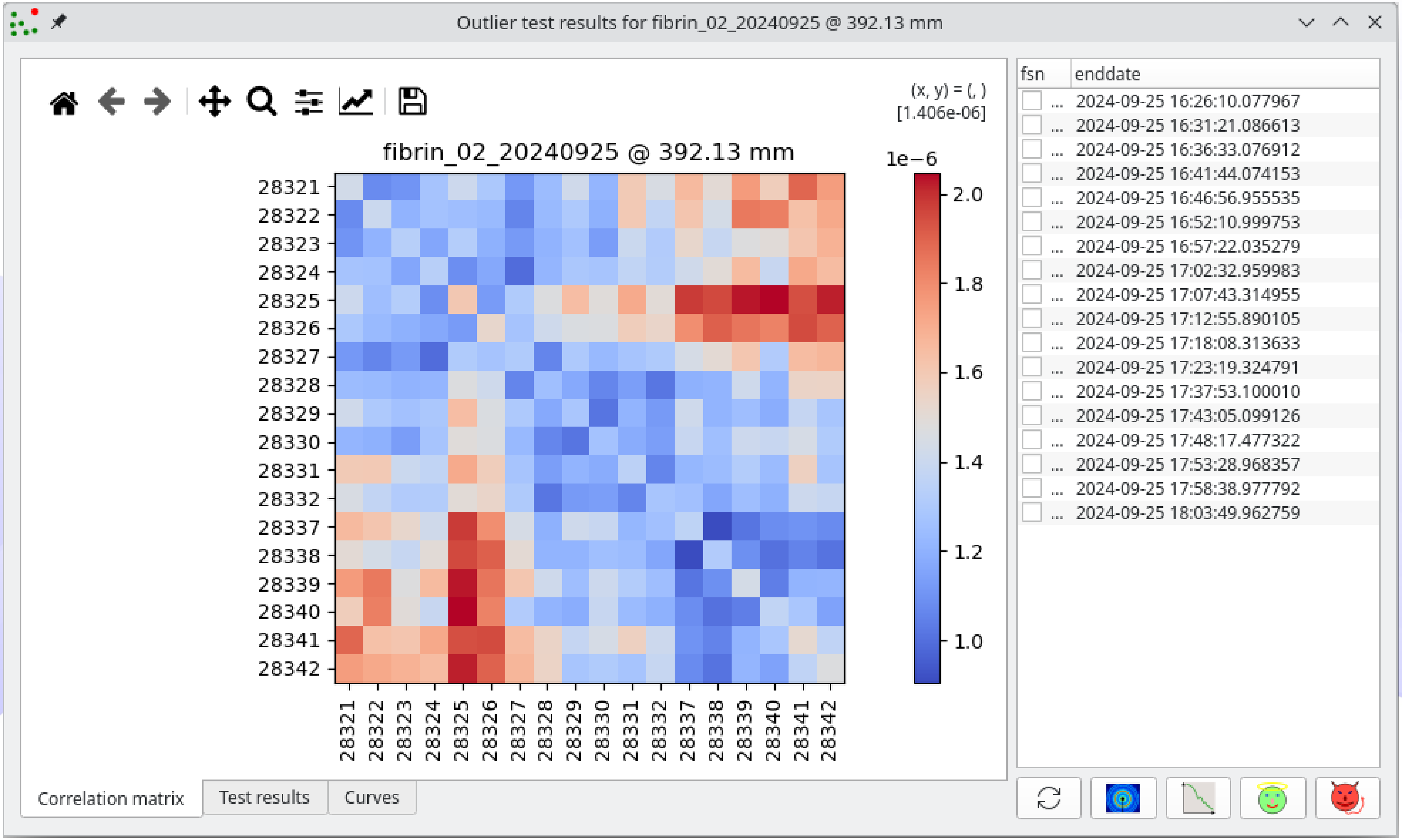
- Filter exposures with artefacts (Pilatus chip flashes), assess sample stability



	fsn	enddate
□	28709	2024-10-02 18:37:25.126992
□	28710	2024-10-02 18:42:32.128560
□	28711	2024-10-02 18:47:39.516375
□	28728	2024-10-02 20:03:40.092852
□	28729	2024-10-02 20:08:47.602643
□	28730	2024-10-02 20:13:54.962138
□	28747	2024-10-02 21:29:51.088099
□	28748	2024-10-02 21:34:58.027872
□	28749	2024-10-02 21:40:04.444856
□	28766	2024-10-02 22:56:00.050471
□	28767	2024-10-02 23:01:06.896239
□	28768	2024-10-02 23:06:10.930536
□	28785	2024-10-03 00:22:07.989896
□	28786	2024-10-03 00:27:14.394363
□	28787	2024-10-03 00:32:20.909691
□	28804	2024-10-03 01:48:13.368699
□	28805	2024-10-03 01:53:20.368552
□	28806	2024-10-03 01:58:25.868001
□	28823	2024-10-03 03:14:24.333774
□	28824	2024-10-03 03:19:31.842041
□	28825	2024-10-03 03:24:39.832704
☑	28842	2024-10-03 04:40:32.318588
☑	28843	2024-10-03 04:45:39.819024
☑	28844	2024-10-03 04:50:45.809291
□	28861	2024-10-03 06:06:42.782513
□	28862	2024-10-03 06:11:49.782726
□	28863	2024-10-03 06:16:55.795258
□	28880	2024-10-03 07:32:58.185621

Data post-processing GUI

- Filter exposures with artefacts (Pilatus chip flashes), assess sample stability

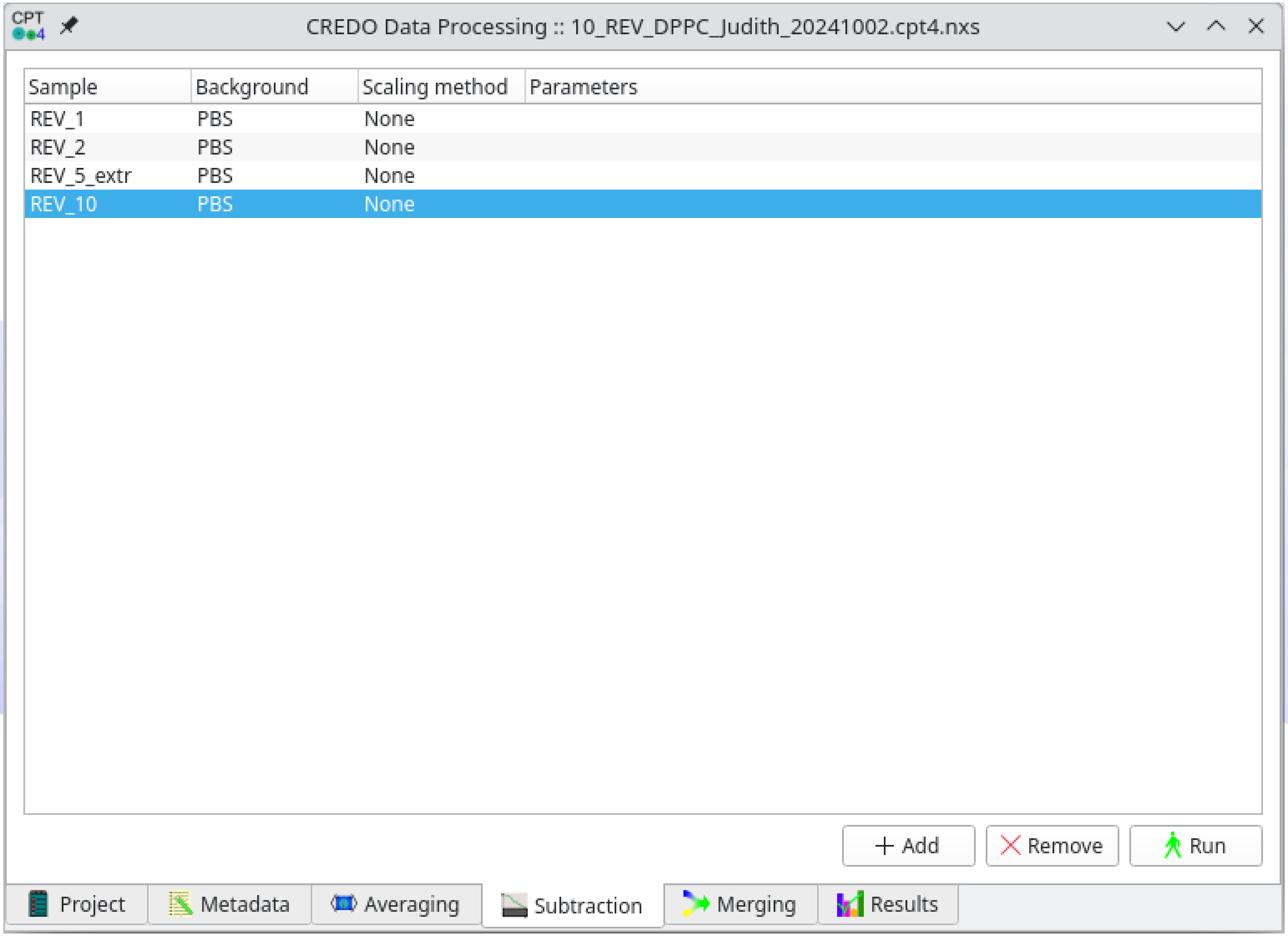


Outlier test results for fibrin_02_20240925 @ 392.13 mm

fsn	enddate
□ ...	2024-09-25 16:26:10.077967
□ ...	2024-09-25 16:31:21.086613
□ ...	2024-09-25 16:36:33.076912
□ ...	2024-09-25 16:41:44.074153
□ ...	2024-09-25 16:46:56.955535
□ ...	2024-09-25 16:52:10.999753
□ ...	2024-09-25 16:57:22.035279
□ ...	2024-09-25 17:02:32.959983
□ ...	2024-09-25 17:07:43.314955
□ ...	2024-09-25 17:12:55.890105
□ ...	2024-09-25 17:18:08.313633
□ ...	2024-09-25 17:23:19.324791
□ ...	2024-09-25 17:37:53.100010
□ ...	2024-09-25 17:43:05.099126
□ ...	2024-09-25 17:48:17.477322
□ ...	2024-09-25 17:53:28.968357
□ ...	2024-09-25 17:58:38.977792
□ ...	2024-09-25 18:03:49.962759

Data post-processing GUI

- Subtract background (optional)

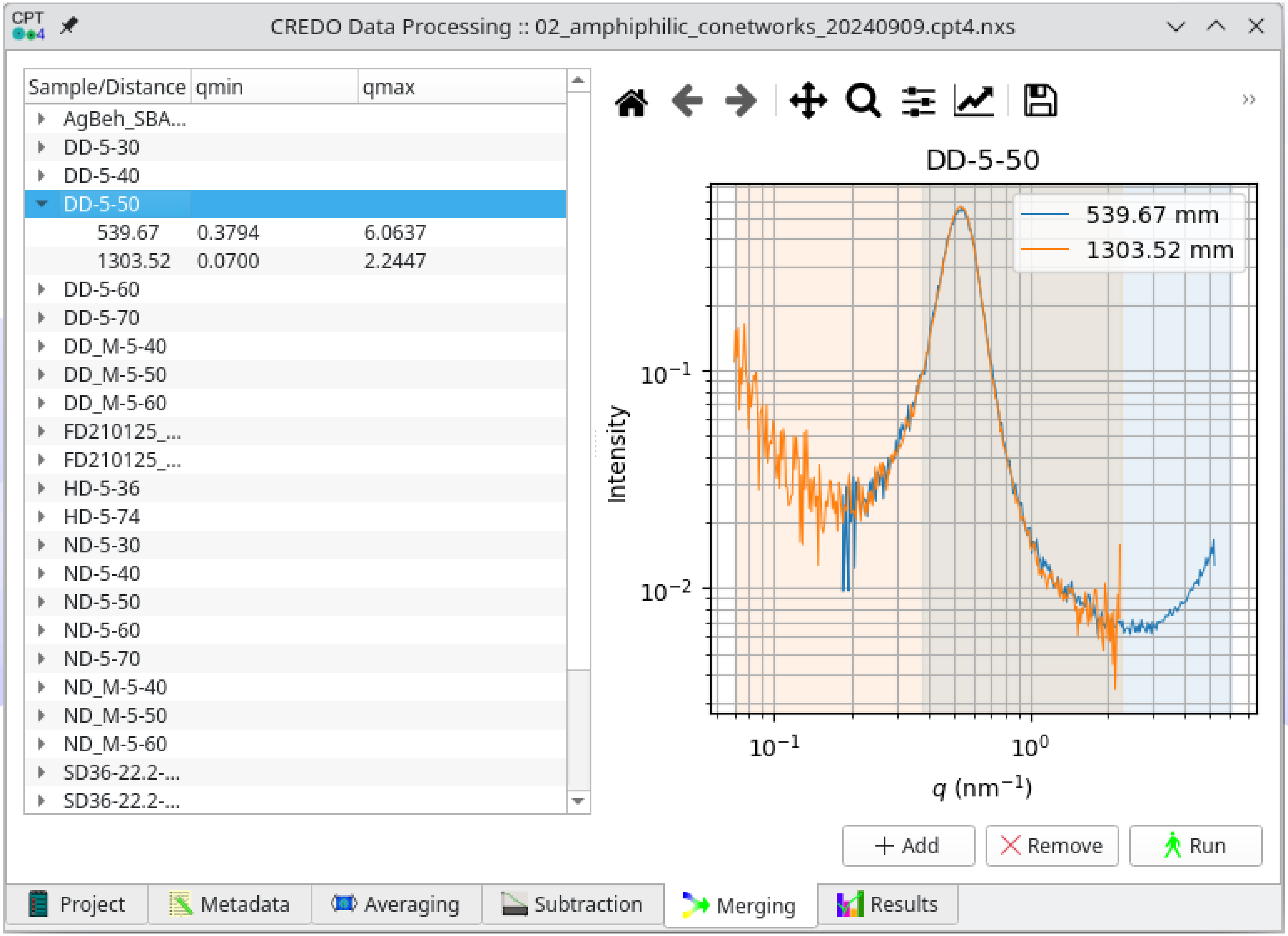


CREDO Data Processing :: 10_REV_DPPC_Judith_20241002.cpt4.nxs

Sample	Background	Scaling method	Parameters
REV_1	PBS	None	
REV_2	PBS	None	
REV_5_extr	PBS	None	
REV_10	PBS	None	

Data post-processing GUI

- Merge curves from multiple sample-to-detector distances

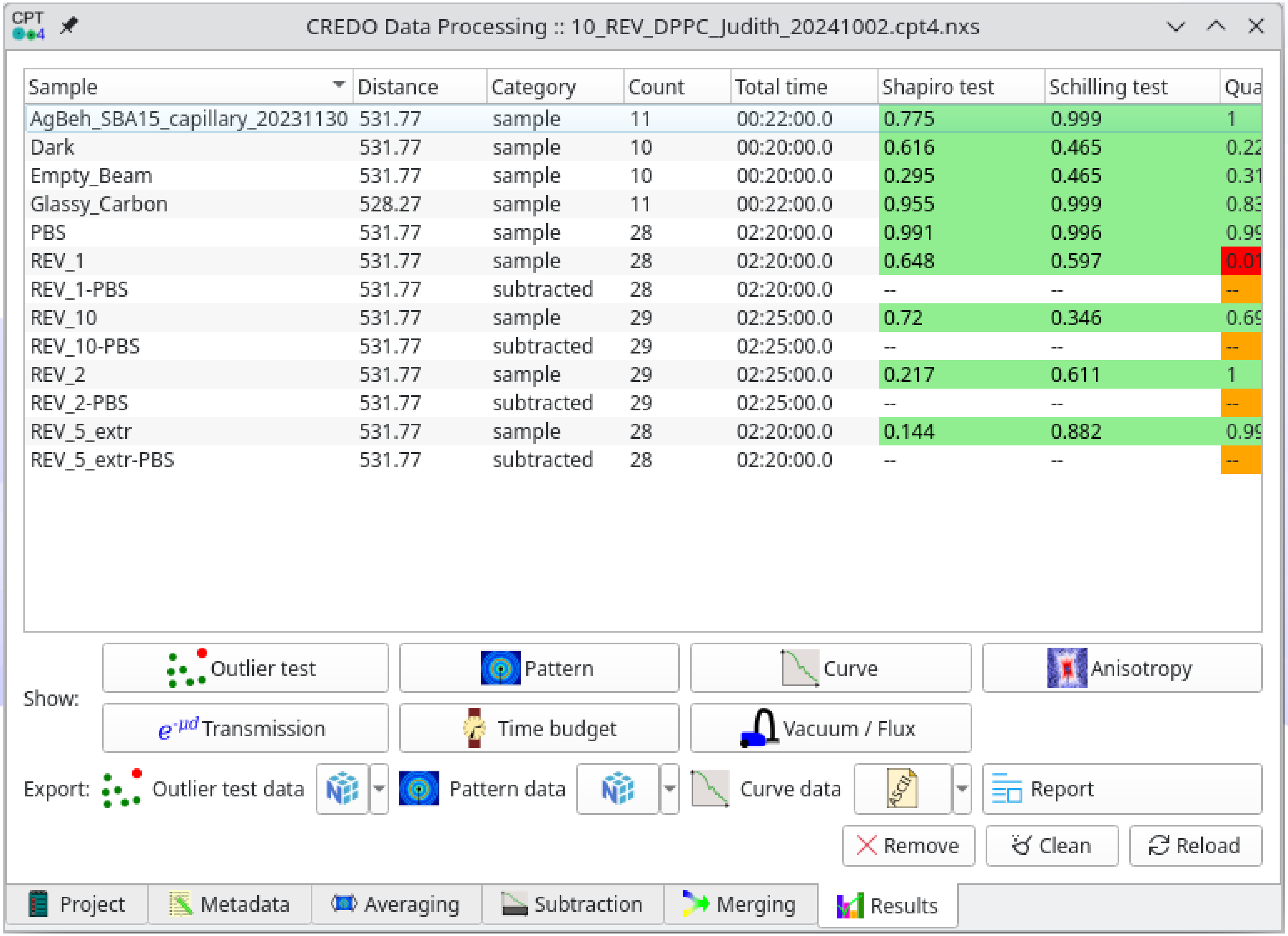


CREDO Data Processing :: 02_amphiphilic_conetworks_20240909.cpt4.nxs

Sample/Distance	qmin	qmax
AgBeh_SBA...		
DD-5-30		
DD-5-40		
DD-5-50		
539.67	0.3794	6.0637
1303.52	0.0700	2.2447
DD-5-60		
DD-5-70		
DD_M-5-40		
DD_M-5-50		
DD_M-5-60		
FD210125_...		
FD210125_...		
HD-5-36		
HD-5-74		
ND-5-30		
ND-5-40		
ND-5-50		
ND-5-60		
ND-5-70		
ND_M-5-40		
ND_M-5-50		
ND_M-5-60		
SD36-22.2-...		
SD36-22.2-...		

Data post-processing GUI

- Present results: draw curves, patterns, analyze anisotropy, export to various formats



CREDO Data Processing :: 10_REV_DPPC_Judith_20241002.cpt4.nxs

Sample	Distance	Category	Count	Total time	Shapiro test	Schilling test	Qua
AgBeh_SBA15_capillary_20231130	531.77	sample	11	00:22:00.0	0.775	0.999	1
Dark	531.77	sample	10	00:20:00.0	0.616	0.465	0.22
Empty_Beam	531.77	sample	10	00:20:00.0	0.295	0.465	0.31
Glassy_Carbon	528.27	sample	11	00:22:00.0	0.955	0.999	0.83
PBS	531.77	sample	28	02:20:00.0	0.991	0.996	0.99
REV_1	531.77	sample	28	02:20:00.0	0.648	0.597	0.01
REV_1-PBS	531.77	subtracted	28	02:20:00.0	--	--	--
REV_10	531.77	sample	29	02:25:00.0	0.72	0.346	0.69
REV_10-PBS	531.77	subtracted	29	02:25:00.0	--	--	--
REV_2	531.77	sample	29	02:25:00.0	0.217	0.611	1
REV_2-PBS	531.77	subtracted	29	02:25:00.0	--	--	--
REV_5_extr	531.77	sample	28	02:20:00.0	0.144	0.882	0.99
REV_5_extr-PBS	531.77	subtracted	28	02:20:00.0	--	--	--

Conclusions

- Storage format for in-house
  - Home-brewn might be okay on the short term
  - Standardized format on the long term
- User facility → requirement of interoperability (and F+A+R, too)
- Completeness vs. simplicity
- Find the most agreed-upon data format for your domain
  - SAXS: NeXus, canSAS
  - XAS: NeXus?
  - HDF5 should be a good general choice
- Exporting processed data to other formats
  - “No matter how good a data storage format you adapt, most users will ask for ASCII text files”
  - “… or Excel workbooks.”

Thank you for your attention (again)!

- Research Group for Biological Nanochemistry, HUN-REN Research Centre for Natural Sciences (https://bionano.ttk.hu/biological-nanochemistry)
- CREDO SAXS laboratory (https://credo.ttk.hu)