

# File formats for GISANS

GISANS Data Reduction and Analysis Upgrades Initiative

Discussion Point I:

Get a common data format for GISANS data reduction

2026-03-19

Brian Maranville (NIST Center for Neutron Research)

# GISANS data: raw and reduced

## Raw data

- Direct output of instrument
- Full instrument state
  - Sample Environment (temp, field, etc.)
  - Complete representation of measurement geometry
  - Detector outputs

## Reduced data

- From reduction of raw data
- Converted to “physical units”
- Instrument-independent (ideally)
- Used as direct input to modeling software

# Raw GISANS data

- NeXus format (HDF5)
- Geometry must be expressed in NeXus form
  - Is a new application definition required?  
(if data can't be expressed with existing NX classes)
- Must be readable by reduction software
- Universal reduction standard format / software:
  - Is it possible, desirable and worth the effort?
  - It has not been done in e.g. reflectometry and SANS

# Reduced GISANS data: content

- In physical units
  - $Q_x, Q_y, Q_z$
  - Normalized intensity
  - Incident  $Q$
  - Resolution, etc... (everything needed to do analysis)
- Metadata (provide context for later use: FAIR)
  - Sample (structure, composition, etc.)
  - Researchers
  - Experiment

# Reduced GISANS data: format

- Need binary structured file
  - Multidimensional data
  - Structured metadata
- Container choices:
  - HDF5 (NeXus compatible)
  - Xarray
  - Zarr

# Reduced GISANS data: NeXus?

- Provides useful classes
  - NXpolarization, etc.
- Provides a place to publish a standard
  - As application definition
- Not designed around sample-centric geometries
- Limited operational benefit for full application def
  - General NeXus tools will not help with analysis or reduction
- Can use NeXus metadata without defining a new standard
  - identify plottable data for viewers

# Example: ORSO reflectometry data format

- Stated goal:
  - Human-readable text format
  - Rich metadata
  - Standard layout of reflectivity data (Q, R, dR, dQ)
  - Extensible with additional columns
  - Sample-oriented geometry (incident angle, scattered angle wrt sample)
- Small group, single leader
- Initial proposal championed by Jochen Stahn
  - Wrote the first draft of the format
  - Led the discussion
- Adopting NeXus conventions was not a primary consideration

# ORSO standard: lessons learned

- Having a champion is critical to move things forward
- Agreeing on names is hard and necessary
- Versioning is necessary: plan for it
- Well-defined structure is more important than container
  - Initial implementation was text-based column data with YAML header
  - Mapping this to HDF5 was straightforward (mostly)
  - Adding zarr etc. containers would be straightforward  
(there are many hierarchical data containers available)

# Defining a new standard: first steps

- Identify stakeholders
  - Who will make the software that writes the format from raw data?
  - Who will make the software that reads the format for analysis?
- Identify champion
  - Who will organize the efforts
- Agree on scope
  - What are the most important design considerations (from stakeholders)