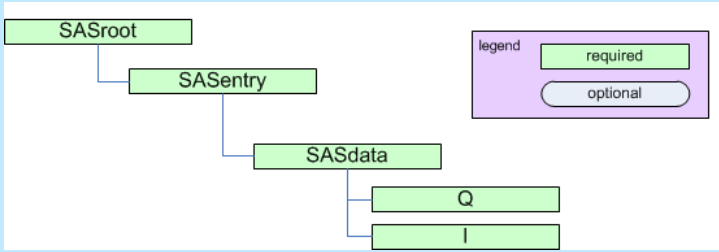
http://download.nexusformat.org/doc/html/classes/applications/NXcanSAS.html

NXcanSAS: standard to store reduced SAS data of any dimension



Pete R. Jemian
Advanced Photon Source
Argonne National Laboratory
Argonne, IL USA

2017-06-05

This work was supported in part by the U.S. Department of Energy, Office of Science, under Contract No. DE-AC-02-06CH11357.

Outline

- Background
- Goals
- Status
- Requirements
- Examples
- Who uses
- Next steps

Reduced data

Data presented for analysis after all instrument-specific artifacts and corrections have been applied

Summary

NXcanSAS is a standard to store reduced small-angle scattering data of any dimension. It is sufficiently general that it may be used for any form of reduced SAS data in different scientific applications. The hierarchical structure of the canSAS standard has been designed to be compatible with the NeXus data format.

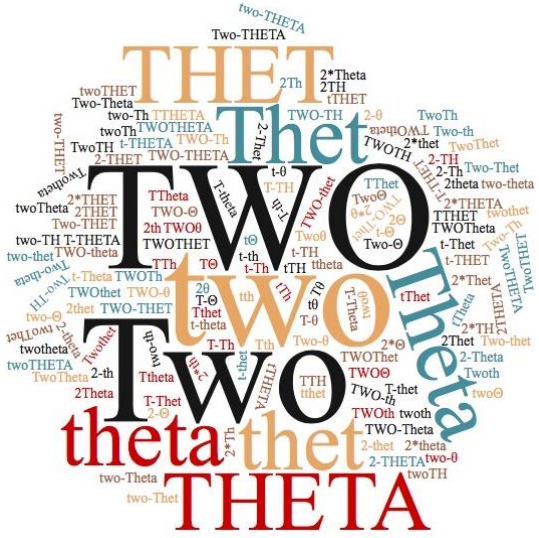
The canSAS format is intended for use in Data Analysis and Data Deposition

Background

… *very* brief history here

- *NXcanSAS* evolved from canSAS1d/1.1 XML standard
- Extend 1-D standard to include multi-dimensional data
- Note: standard for solution bioSAS adopted by IUCr:
  Acta Cryst. D Biol. Cryst. **68**.6 (2012) 620–626
- Basic plans set out at 2012 canSAS Uppsala meeting
- Additional requirements from 2015 canSAS
- Ratified by NeXus late 2016, initial release 2017-01

- Metadata dictionary and structured data support automated workflows



How many variations of this common term?

Goals

- Facilitate better sharing of SAS data analysis software
- Common data formats allow the easy use of different analysis software packages
- Generalize to describe simple experiments and complex experiments (such as with multiple detectors or multimodal experiments)
- Store reduced SAS data of any dimension
- Q can be either a vector (***Q***) or magnitude |**Q**|
- Identify and associate scanning axes (“self describing data”)
- Easy plotting of the data
- Maintain the original dimensionality of the data if at all possible
- Use existing standards where possible or practical
- Address the SAS community not reached by 2012 bioSAS standard
- Open source repositories

Status

NXcanSAS is a NeXus application definition

- What this means:
  - standard is defined in an XML file using NXDL (NeXus definition language)
  - syntax and content can be validated with NeXus schema
  - NeXus provides rich metadata dictionary and structure in HDF5 file
- Where is the source of the documentation?
  - Automatically generated from XML definition
- How do we modify *NXcanSAS* as we want?
  - Make a GitHub pull request with proposed changes
  - Describe changes to someone who can make a pull request
- Can data files be verified?
  - Work in progress: *punx* (http://punx.readthedocs.io)

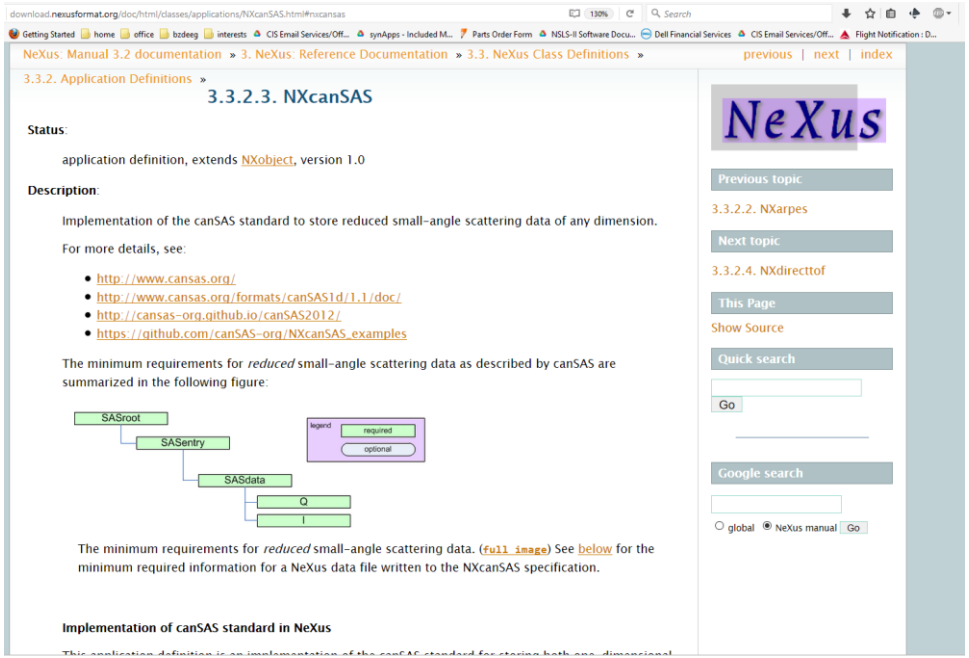
NXcanSAS is part of NeXus

Source is on GitHub: https://github.com/nexusformat/definitions

NXDL source

```
<group type="NXdata">
  <doc>
    A *SASData* group contains a single reduced small-angle scattering data set
    that can be represented as :math:`I(\vec{Q})` or :math:`I(|\vec{Q}|)`.

    *Q* can be either a vector (:math:`\vec{Q}`) or a vector magnitude (:math:`|\vec{Q}|`)

    The name of each *SASdata* group must be unique within a SASentry group.
    Suggest using names such as ``sasdata01``.

    NOTE:  For the first *SASdata* group, be sure to write the chosen name
    into the `SASentry/@default` attribute, as in::

        SASentry/@default="sasdata01"

    A *SASdata* group has several attributes:

    * I_axes
    * Q_indices
    * Mask_indices

    To indicate the dependency relationships of other varied parameters,
    use attributes similar to ``@Mask_indices`` (such as ``@Temperature_indices``
    or ``@Pressure_indices``).
  </doc>
  <attribute name="canSAS_class">
    <doc>Official canSAS group: :index:`NXcanSAS (applications); SASdata`</doc>
    <enumeration>
      <item value="SASdata" />
    </enumeration>
  </attribute>
  <!-- attributes, see: http://www.cansas.org/formats/canSAS2012/1.0/doc/framework.html#terms -->
  <attribute name="signal" type="NX_CHAR" >
    <doc>
      Name of the default data field.
    </doc>
    <enumeration>
      <item value="I"><doc>For canSAS **SASdata**, this is always "I".</doc></item>
    </enumeration>
  </attribute>
```

documentation



Argonne NATIONAL LABORATORY

Requirements

Extend canSAS1d/1.1 XML standard

- include multi-dimensional data

from 2015 meeting in Japan

- Describe multiple types of uncertainty
- Distinguish between uncertainty and resolution
- Minimize required contents

Other Information to be stored (when possible)

- additional dimensions for complex experiments (λ, T, t, P, …)
- uncertainties and their constituents
- masking information
- metadata (title, wavelength, radiation type and source, sample info, thickness, raw data reference, owner contact info…)
- analytical results
- complementary data

Units

- NeXus uses the UDUNITS standard which is very flexible
- canSAS requires specific units for *Q* and *I*:
  - *Q*: "`1/nm`" (preferred), "`1/m`", or "`1/angstrom`"
  - *I*: three cases
    - absolute units: *dΣ/dΩ(Q)*
      - differential cross-section per unit volume per unit solid angle (such as: `1/cm/sr` or `1/m/sr`)
    - absolute units: *dσ/dΩ(Q)*
      - differential cross-section per unit atom per unit solid angle (such as: `cm^2` or `m^2`)
    - arbitrary units: *I(Q)*
      - usually a ratio of two detectors but units are meaningless (such as: `a.u.` or `counts`)
- Data expressed in other units will generate a warning from validation software and may not be processed by some analysis software packages.

Examples: on GitHub

https://github.com/canSAS-org/NXcanSAS_examples

- these example files have been written as NXcanSAS (NeXus/HDF5) files:
  - 1d_standard
  - canSAS2012_examples (random number data)

- Examples are provided as
  - NeXus HDF5 files
  - Python code to write the file
  - Text file describing structure of HDF5 file

Examples: How to describe *uncertainty*

- These values are the estimates of uncertainty.
- By default, this will be interpreted to be the estimated standard deviation.
- In special cases, when a standard deviation cannot possibly be used, its value can specify another measure of distribution width.
- Can be added to any dataset (NeXus calls this a *field*) with an attribute:
  - Exact name of uncertainty dataset is flexible

```
I:NX_FLOAT64[91]
  @units = 1/cm
  @uncertainties = Iesd
Iesd:NX_FLOAT64[91]
  @units = 1/cm
Q:NX_FLOAT64[91]
  @units = 1/nm
  @resolutions = Qdev
Qdev:NX_FLOAT64[91]
  @units = 1/nm
```

```
I : float[m,n]
  @uncertainties=Idev
Idev : float[m,n]
 @components=I_uncertainties
I_uncertainties : (group)
 electronic : float[m,n]
  @basis="Johnson noise"
 counting_statistics : float[m,n]
  @basis="shot noise"
 secondary_standard : float[m,n]
  @basis="esd"
```

Example: metadata

```
sassample:NXsample
     @NX_class = NXsample
     @canSAS_name = sassample
     @canSAS_class = SASsample
     ID:NX_CHAR[34] = AF1410-10 (AF1410 steel aged 10 h)
     details:NX_CHAR[128] =
       transverse saturation magnetic field (1.6 T) applied in
       horizontal direction to clear magnetic domain scattering
```

```
sasinstrument:NXinstrument
     sascollimation:NXcollimator
     sasdetector:NXdetector
     sassource:NXsource
       incident_wavelength:NX_FLOAT64 = 0.85
       radiation:NX_CHAR[7] = neutron
       wavelength_spread:NX_FLOAT64 = 25.0
         @units = percent
```

Example: multimodal data

- NeXus structure is flexible to describe many sets of measurements
- Use data groups under a single entry group if from the same sample
- (Sample and wavelength data not shown for brevity)

2-D SAS/WAS images

```
SASroot
  SASentry
    SASdata
      @name="sasdata"
      @Q_indices=0,1
      @I_axes=Q,Q
      I: float[100, 512]
      Qx: float[100, 512]
      Qy: float[100, 512]
      Qz: float[100, 512]
    SASdata
      @name="wasdata"
      @Q_indices=0,1
      @I_axes=Q,Q
      I: float[256, 256]
      Qx: float[256, 256]
      Qy: float[256, 256]
      Qz: float[256, 256]
```

Example:

many samples, two (NIST) SANS each

- Steel sample, aging series
- Area detector image at each aging time
- Binned in horizontal & vertical sectors
- 1-D I(Q) have different lengths
- (lots of other metadata in example file)
- file: cs_af1410.h5
  https://github.com/canSAS-org/NXcanSAS_examples/tree/master/1d_standard

```
cs_af1410.h5 : NeXus data file
  @creator = xml2hdf5.py
  @default = AF1410_10
  AF1410_10:NXentry
    AF1410_a10:NXdata
      I:NX_FLOAT64[77]
      Idev:NX_FLOAT64[77]
      Q:NX_FLOAT64[77]
    AF1410_b10:NXdata
      I:NX_FLOAT64[76]
      Idev:NX_FLOAT64[76]
      Q:NX_FLOAT64[76]
    sasinstrument:NXinstrument
      sascollimation:NXcollimator
      sasdetector:NXdetector
      sassource:NXsource
    sassample:NXsample

  AF1410_1h:NXentry
    AF1410_a1h:NXdata
    AF1410_b1h:NXdata
  AF1410_20:NXentry
  AF1410_2h:NXentry
  AF1410_50:NXentry
  AF1410_5h:NXentry
  AF1410_8h:NXentry
  AF1410_cc:NXentry
  AF1410_hf:NXentry
  AF1410_qu:NXentry
```

Example:
citation

- NeXus provides the *NXnote* to describe any additional freeform information not covered by the other base classes

- These tags and attributes were supplied *ad hoc*

```
sasnote:NXnote
     citation:NXcollection
       @NX_class = NXcollection
       @canSAS_name = citation
       @tag = citation
       journal:NX_CHAR[11] = Acta Metall
         @tag = journal
       pages:NX_CHAR[9] = 1869-1884
         @tag = pages
       title:NX_CHAR[92] = Small-Angle Neutron Scattering
         Studies of Carbide Precipitation in
         Ultrahigh-Strength Steels
         @tag = title
       volume:NX_CHAR[2] = 41
         @tag = volume
       year:NX_CHAR[4] = 1993
         @tag = year
       authors:NXcollection
         @NX_class = NXcollection
         @canSAS_name = authors
         @tag = authors
         author_0:NX_CHAR[10] = A.J. Allen
           @tag = author
         author_1:NX_CHAR[11] = D. Gavillet
           @tag = author
         author_2:NX_CHAR[13] = J.R. Weertman
           @tag = author
```

Software to write, read, & view NXcanSAS

- *Anything* that can read or write HDF5

- Viewing (as NeXus data file): NeXpy, PyMCA
- reading: h5py, SASview, Mantid, IgorPro, Irena/Nika (planned)
- writing: h5py, SASview, Mantid, Irena/Nika (planned)
- Validation: *punx*

Next steps

- Provide example code to read and write
- Developers: obtain acceptance as storage format
- IUCr CSAS: obtain recognition as deposition format
- Present at 2017 IUCR meeting

Acknowledgements

- Andrew Jackson (ESS)
- Tobias Richter (ESS)
- Paul Butler (NIST)
- Steve King (ISIS)
- Adrian Rennie (Uppsala Univ.)
- Brian Pauw (BAM)

Thank you for your attention