canSAS working group report: a common data format

Andrew Jackson (ESS)

Pete Jemian (APS), Tobias Richter (now ESS, was Diamond), Ron Ghosh

Data Formats – Background

- canSAS started (1998) on the idea of a common data format to aid the nomadic scatterer

- sasCIF developed and proposed – did not get ratified by IUCr …

- canSAS V at NIST in 2007 – revisited this goal and agreed an XML based format for 1D data (canSAS1D)
  - Produced at ISIS, NIST, ANSTO, APS, Diamond (sometimes), ILL
  - Mantid capable of outputting
  - Read by SasView and IGOR (NIST macros and Irena)
  - Online converter for 3 column data
  - Code for reading/writing provided in various languages (Fortran, Java, Python …)

Data Formats – Goals at canSAS 2012

Format for 1D and 2D ***REDUCED*** data

Simplicity
and
Extensibility

but

Well Defined
Limited in Scope

Simplicity of Creation
Simplicity of Use

Appropriate Metadata
Traceability/provenance
Deposition

Address issues with canSAS1D XML

Outcomes from canSAS 2012 and Current Status

- v1.1 of 1D xml format (incorporating wavelength dependent transmissions) : http://www.cansas.org/formats/canSAS1d/1.1/doc/index.html
- Proposal for nD format (implementation agnostic) : http://www.cansas.org/formats/canSAS2012/1.0/doc/index.html
- nD format presented at SAS 2012 by P Jemian: "*1105 - The canSAS Format for Storage and Interchange of Reduced MultiDimensional Small-Angle Scattering Data*"
- canSAS2012 format contributed to NeXus as NXcanSAS class definition in December 2014: http://download.nexusformat.org/doc/html/classes/contributed_definitions/NXcanSAS.html

canSAS2012 format

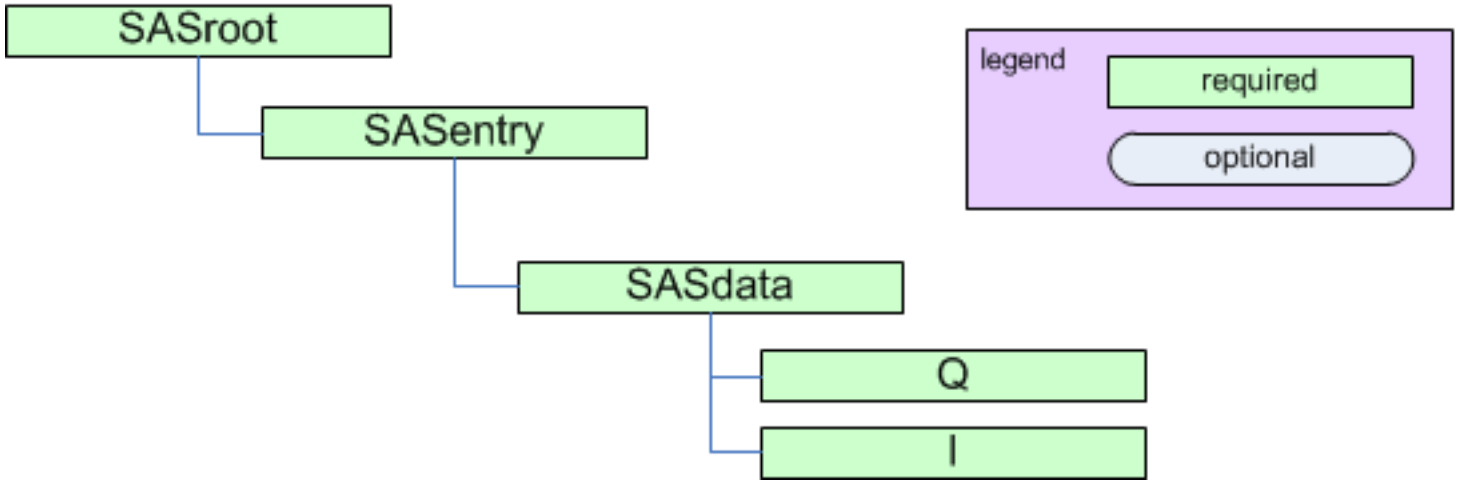
Requirements

The absolute minimum information required for the standard analysis of small-angle scattering measurements is intensity as a function of scattering vector,

- Allow for representation of reduced data of any dimensionality
  - 1D SAS data
  - 2D SAS data from detectors
  - additional dimensions for complex experiments
  - Q can be either a vector (***Q***) or magnitude |**Q**|
- Identify and associate scanning axes ("self describing data")
- Provide (when possible)
  - uncertainties and their constituents
  - masking information
- Allow for
  - complex experiments with multiple detectors
  - easy plotting of the data in close to their "raw" form
- Maintain the original dimensionality of the data if at all possible
- Use existing standards where possible or practical

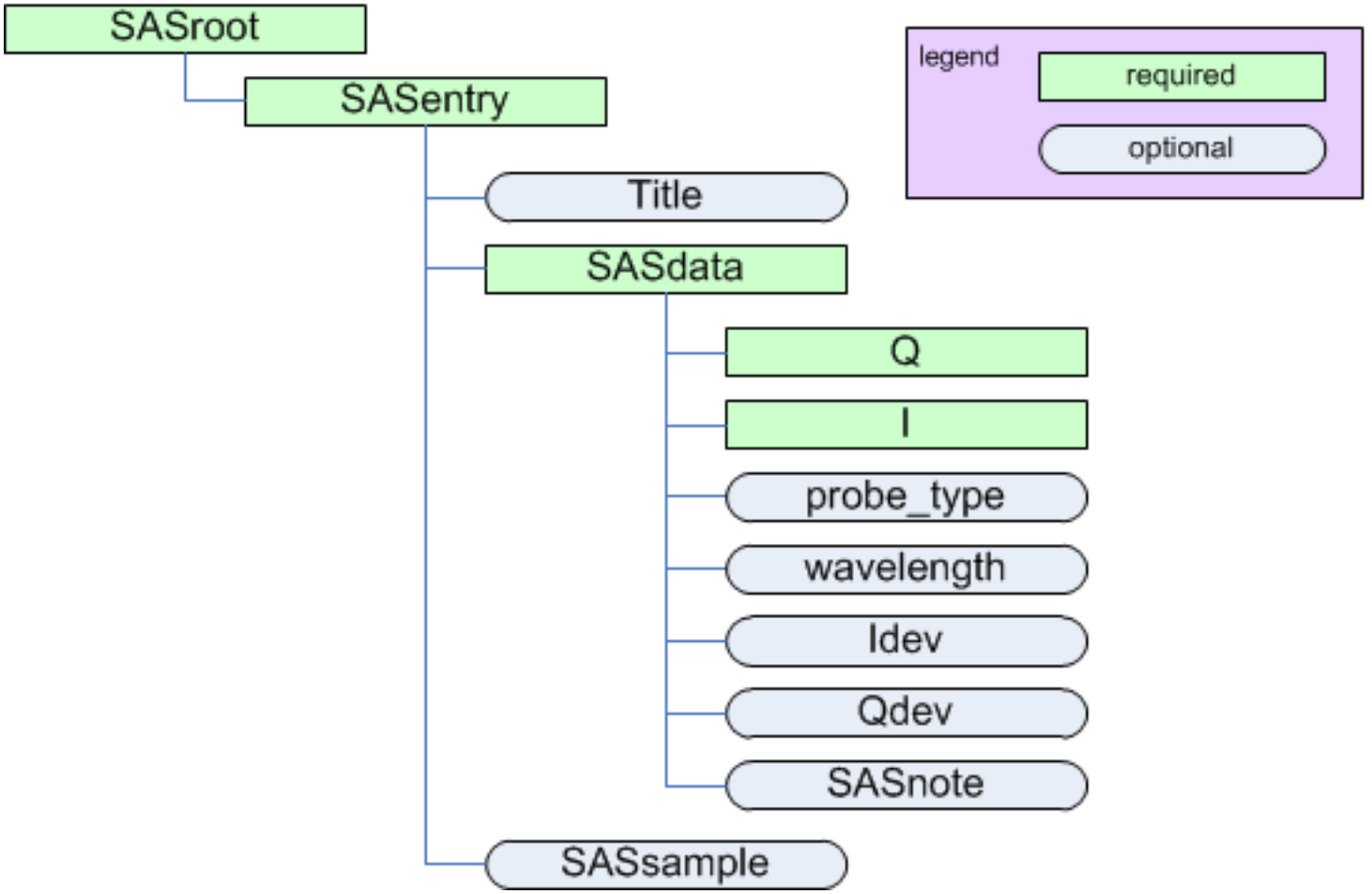
canSAS2012 format

Absolute minimum requirement for analysis of SAS data



canSAS2012 format

Recommended minimum content for reduced SAS data



canSAS2012 format

SASdata:

different use from original 1D format, refers to a single reduced data set that can be represented thus (such as from one detector)

SASdata groups have several attributes:

@name: Text that describes this group. Must be unique within a SASentry group. Such as:

```
@name="sasdata01"
```

@I_axes: Comma-separated list that describes the names of the data objects that correspond to the `I` data object. Such as:

```
@I_axes="Temperature,Time,Pressure,Q,Q"
```

@Q_indices: Array that describes which indices (of the I data object) are used to reference the `Q` data object. The items in this array use zero-based indexing. Such as:

```
@Q_indices=1,3,4
```

which indicates that Q requires three indices from the I data object: one for time and two for Q position.

@Mask_indices: Array that describes which indices (of the I data object) are used to reference the `Mask` data object. The items in this array use zero-based indexing. Such as:

```
@Mask_indices=3,4
```

which indicates that Q requires two indices from the I data object for Q position.

To indicate the dependency relationships of other varied parameters, use attributes similar to `@Mask_indices` (such as `@Temperature_indices` or `@Pressure_indices`).

canSAS2012 format

1-D $I(Q)$

Examples: HDF5 XML

```
SASroot
  SASentry
    SASdata
      @Q_indices=0
      @I_axes="Q"
      I: float[100]
      Q: float[100]
```

2-D (image) $I(|Q|) \pm \sigma(|Q|)$

```
SASroot
  SASentry
    SASdata
      @Q_indices=0,1
      @I_axes="Q,Q"
      I: float[300, 300]
        @uncertainty=Idev
      Q: float[300, 300]
      Idev: float[300, 300]
```

canSAS2012 format

1-D $I(t, Q(t))$

This example is slightly more complex, showing data where Q is also time-dependent.

```
SASroot
  SASentry
    SASdata
      @Q_indices=0,1
      @I_axes="Time,Q"
      I: float[nTime,100]
      Q: float[nTime,100]
      Time: float[nTime]
```

canSAS2012 format

2-D $I(t, T, P, Q(t, T, P))$

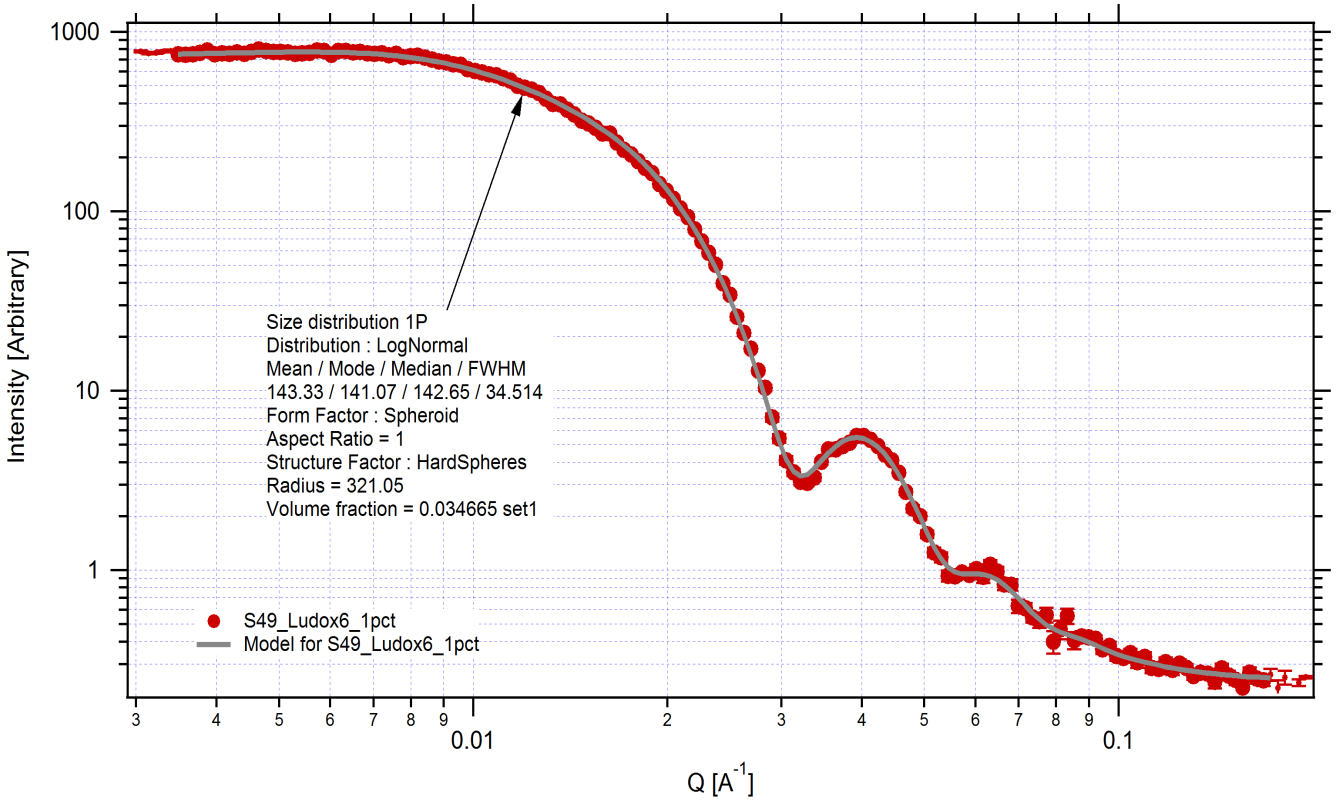
Complex case of $I(t, T, P, Q(t, T, P))$ where all Q values are different for each combination of time, temperature, and pressure.

Examples: HDF5 XML

```
SASroot
  SASentry
    SASdata
      @Q_indices=0,1,2,3
      @I_axes="Time,Temperature,Pressure,Q"
      I: float[nTime,nTemperature,nPressure,100*512]
      Qx: float[nTime,nTemperature,nPressure,100*512]
      Qy: float[nTime,nTemperature,nPressure,100*512]
      Qz: float[nTime,nTemperature,nPressure,100*512]
      Time: float[nTime]
      T: float[nTemperature]
      P: float[nPressure]
```

NXcanSAS Example

NXcanSAS
Example from APS SAXS



@canSAS_class = SASroot
  entry:NXentry
    @NX_class = NXentry
    @canSAS_class = SASentry
    title:char[15] = S49_Ludox6_1pct
    data:NXdata
      @NX_class = NXdata
      @signal = I
      @axes = Q
      @canSAS_class = SASdata
      I:float64[258] = [1067702.0569415579, '...', 0.25364629360280838]
        @units = a.u.
        @uncertainty = Idev
      Idev:float64[258] = [61701.129367048801, '...', 0.0039049644734773402]
        @units = a.u.
      Q:float64[258] = [0.00012556253793213361, '...', 0.19507868408793791]
        @units = 1/A