MS051.O04

Creating a robust,extensible XAS data standard

James Reginald Hester[1]
[1]*Acns, Ansto, Kirrawee Dc, Australia*
E-mail: jxh@ansto.gov.au

The ability to verify results given the raw data is fundamental to the integrity of research. However, the absence of computer-friendly standards for expressing increasingly complex XAS data processing and analysis pipelines means that assessing published XAS results is a laborious process. Attempts to create a suitable standard must cope with the dispersed nature of the XAS community, a variety of popular software packages and input formats, and differences in experimental, data reduction and analysis approaches.

The robust, inclusive and low-maintenance solution proposed here is based on a curated collection of plain text names, together with definitions. XAS practitioners contribute to this "dataname commons", and then interested groups refer to some subset of the datanames when formulating and publishing their own format- and application- specific standards. One of many advantages of this scheme is that datafiles based on these individual standards can be losslessly and unambiguously interchanged [1]. The approach itself is a straightforward application of the "olog" concept [2].

The utility of this dataname commons for computational work relies on the definitions meeting three simple criteria: each definition must include (i) a list of other ("input") datanames whose values affect the given dataname's value; (ii) a description of the range of possible values taken by the given dataname; (iii) a description of the link between the input datanames and the given dataname that allows unambiguous identification of a corresponding value for the given dataname given the values of the input datanames. In other words, the definition must describe a mapping.

Management of the dataname collection is simplified by introducing naming conventions: (i) all datanames that depend on the same set of input datanames begin with the same text string; (ii) a versioning system to signal addition of a new input dataname to a definition. These rules allow variations in XAS model and data reduction processes to be captured with minimal proliferation of datanames.

[1] Hester, J.R. (2016) "A Robust, Format-Agnostic Scientific Data Transfer Framework" The Data Science Journal 15:12 http://doi.org/10.5334/dsj-2016-012

[2] Spivak, D. I. and Kent, R. E. (2012) "Ologs: A Categorical Framework for Knowledge Representation" PLoS ONE 7(1) http://doi.org/10.1371/journal.pone.0024274

Keywords: Data standards, XAS