

# Designing format-agnostic, robust metadata specifications

James R. Hester

Bragg Institute, Australian Nuclear Science and Technology Organisation (ANSTO)  
Locked Bag 2001, Kirrawee DC, NSW 2232  
Australia  
e-mail: [jxh@ansto.gov.au](mailto:jxh@ansto.gov.au)

## ABSTRACT

This talk proposes a framework for analyzing and organising large, diverse collections of metadata specifications. Current and legacy standards can be reduced to essential concepts and relationships, after which they are reassembled into a uniform metadata specification that is easily expandable while maintaining compatibility with the previous standards. No distinction is made (or needs to be made) between data and metadata, and non-programmer domain experts can easily contribute to, or manage, the specification.

The two key ideas behind this approach are: (a) the "olog" concept, which is a lightly disguised presentation of category theory [1,2] and (b) that scientific facts must be independent of the format used to transmit them. (a) provides a precise and simple set of requirements for constructing an extensible, well-defined, modular set of concepts (ontology), and (b) allows us to construct our ontology independent of any particular storage or transmission format.

Encapsulation and extraction of the data and metadata into/from any one particular storage format is achieved by a minimal "format adapter" that should contain no scientific knowledge and has a simple, uniform interface. Application software using this format adapter specifies only the metadata name, type and units. If metadata relationships are provided in machine readable form within the ontology, the format adapter can be wrapped into a generic object which will use the ontology to automatically transform between metadata items, transparently expanding the range of metadata immediately available from a given file format while effectively isolating the application programmer from format-related issues.

The talk will include a description of a freely-available working implementation of these ideas written in Python and using a machine-readable ontology for images. Practical principles and lessons learnt from this exercise will be shared.

## REFERENCES

- [1] D. I Spivak and R. E. Kent, *Ologs: A Categorical Framework for Knowledge Representation*, PLoS ONE 7(1): e24274. doi:10.1371/journal.pone.0024274 (2012)
- [2] D. I. Spivak, *Category Theory for the Sciences*, The MIT Press, <https://mitpress.mit.edu/books/category-theory-sciences> (2014).