

Erica Yang, Brian Matthews, Michael Wilson, *Enhancing the Core Scientific Metadata Model to Incorporate Derived Data*

Abstract

The Core Scientific MetaData model (CSMD) is used by large scientific facilities to catalogue scientific data. The current version provides support to experimental scientists to access their raw data, facility managers for accounting for facility usage and other scientists who wish to re-use raw experimental data. Much of the value in scientific data is provided not only in the raw data but through the analysis of that data to derive published results. An analysis of the raw data analysis process for structural science has shown that various data sets derived from the raw data are of use to scientists and should be stored with the raw data. Extensions to the CSMD are presented to describe the analysis process so that the provenance of the derived data can be captured. A pilot implementation incorporating derived data through this extended CSMD model has been trialled with experimental scientists. Remaining challenges to the adoption of CSMD and tools it supports are considered.