Geophysical Research Abstracts,
Vol. 10, EGU2008-A-02998, 2008
SRef-ID: 1607-7962/gra/EGU2008-A-02998
EGU General Assembly 2008

Standards-based methodology for developing a geoscience markup language

S. J. D. Cox

CSIRO Exploration and Mining, Bentley, WA, Australia

(Simon.Cox@csiro.au / Phone: +61-8-6436-8639)

Markup languages have been developed for data transfer in a variety of earth science disciplines. Most of these have been developed using an informal methodology – typically guided by a data model implicitly defined in some existing document or database, but with the XML schema often designed directly using ad-hoc patterns, or sometimes created automatically by some proprietary toolkit. This often leads to a language that is efficient for a single application or within a workgroup or community, but with limited scope for interoperability across domain boundaries. The latter is a serious constraint to the use of data from diverse sources in cross-disciplinary investigations.

A uniform methodology, based on standards published by Open Geospatial Consortium and ISO, has been developed and applied in the design of GeoSciML. The method is based on the Object Management Group's Model Driven Architecture (MDA), with model design in UML using the General Feature Model from ISO 19109, the use of components from other standards in the ISO 19100 series, and production of the XML schema following the encoding rules specified in ISO 19136. The resultant encoding shows a literal and explicit relationship to the UML model. This is unlikely to be as compact as hand-coded special cases, but is consistently structured across similar models. Full structure and meaning is preserved, and compactness is easily dealt with using standard compression techniques. Furthermore, the use of standard components for elements that are common across domains ensures maximum interoperability.

To assist in the use of this methodology, we have developed two tools:

1. “HollowWorld” – a UML template with ISO 19100 components, stereotypes, and tags pre-loaded, plus some other cross-domain components;
2. “FullMoon” – a UML processing framework, based on application of sets of rules against the XMI representation of a model.

We use a UML design tool that allows direct binding to one or more SVN repositories. These host the various UML packages that are under separate governance arrangements. This overcomes an important limitation of most UML-based methodologies, which effectively treat the entire model as a single artefact.

Three rule-sets are available For FullMoon:

1. validating the UML model with respect to the profile described in ISO 19136
2. generating GML-conformant XML schema according to the rules in ISO 19136
3. generating human-readable documentation of the model, in the form of an HTML frameset.

Use of these tools has allowed the GeoSciML team to develop and maintain the model as a single normative artefact (XMI). Implementation views in XML Schema and HTML documentation are generated automatically at significant release points. This addresses two key issues with ad-hoc approaches: ensuring normative and descriptive content are consistent across maintenance activities, and the ability to support convenient cross-reference between the conceptual model and the XML encoding.