

Design of an Ontology-based Metadata Translator

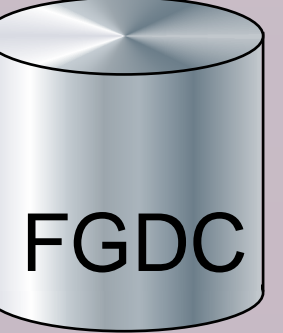
Thomas Russ and Hans Chalupsky

Information Sciences Institute, University of Southern California

Metadata and Interoperability: The Babel Problem

One goal of the SCEC/CME development is to make it easier for researchers to share the data products produced through research or the running of simulation codes. In order to share data products, it must be possible to locate the appropriate items. Descriptions of the data are encoded using “metadata”. Managing metadata across different implementations and institutions within the SCEC/CME community is one of the challenges facing our project. This challenge grows even greater when we wish to make data and code resources available and searchable outside the SCEC community.

One major component of this challenge is that different groups within and outside SCEC use different metadata schemas to describe their resources and data products. To a certain extent, this problem can be simplified by the adoption of standardized metadata schema. In fact, there are currently some efforts underway. But our experience indicates that standardization is rarely complete, and there are often other, external groups who develop their own standards. That makes an approach based on translation between metadata schema look attractive. We are now working on the design and implementation of just such a translator.

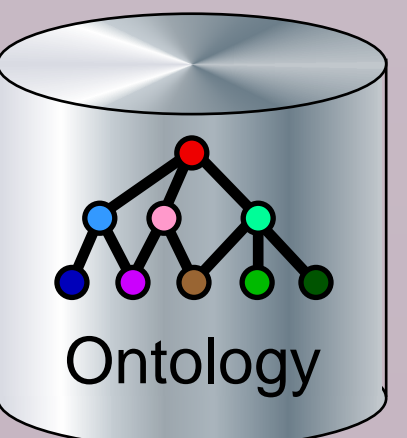


Ontology-based Translation

Our design will involve constructing a knowledge model (ontology) to describe the data products. This model will be encoded and manipulated by a Knowledge Representation and Reasoning (KR&R) system called PowerLoom. PowerLoom provides a logical deductive engine that can reason about the data model. This data model will be used to describe the relevant terms in a domain. Rules and links to transformation procedures will enable us to perform semantic mappings that go beyond purely syntactic transformations.

For example, locations can be specified using UTM coordinates or latitude and longitude pairs. Latitude and longitude can be expressed in either decimal degrees or degree-minute-second form. Transformation procedures linking these alternate representations allow translation. Furthermore, by understanding the hierarchical relationship between general and specific terms, such a reasoning system can support information expressed at differing levels of detail. For example, a particular metadata set may not have a data item which corresponds directly to the term “Fault.” Perhaps all of the information is encoded using the type of fault

instead (i.e., Strike-slip fault, normal fault, reverse fault, etc.) Semantic translation support backed by an ontology that encodes these relationships allows interpretation and bridging of differences in specificity.



The semantic translator will have a common model of terms relevant to datasets of interest. Each metadata schema will have a mapping into and out of this common model. Adding support for a particular metadata schema involves providing a mapping between the metadata schemata and this core model. Using a common model and mapping to and from that model means we do not have to generate an exhaustive collection of pair-wise schema translators. With translation one can build search engines for SCEC data products that can use multiple metadata “languages.”

Initial development will concentrate on supporting the USC and SDSC/SDSU metadata as well as the Federal Geographic Data Committee (FGDC) metadata standard. In the example below, we assume a user who is more familiar with the FGDC metadata (or more likely, having a tool for formulating FGDC queries) wishes to find velocity information for part of the Los Angeles basin. The query is sent to the translator where the FGDC terms are mapped to a logical model of the data. A separate mapping from the logical data model to USC metadata terms is then made and the search can proceed using its own native search mechanisms.

Semantic Translation Example

