Mechanisms for integration of information models across related domains

Rob Atkinson

CSIRO Land & Water (rob.atkinson@csiro.au)

It is well recognised that there are opportunities and challenges in cross-disciplinary data integration. A significant barrier, however, is creating a conceptual model of the combined domains and the area of integration. For example, a groundwater domain application may require information from several related domains: geology, hydrology, water policy, etc. Each domain may have its own data holdings and conceptual models, but these will share various common concepts (eg. The concept of an aquifer).

These areas of semantic overlap present significant challenges, firstly to choose a single representation (model) of a concept that appears in multiple disparate models,, then to harmonise these other models with the single representation. In addition, models may exist at different levels of abstraction depending on how closely aligned they are with a particular implementation. This makes it hard for modellers in one domain to introduce elements from another domain without either introducing a specific style of implementation, or conversely dealing with a set of abstract patterns that are hard to integrate with existing implementations.

Models are easier to integrate if they are broken down into small units, with common concepts implemented using common models from well-known, and predictably managed shared libraries. This vision however requires development of a set of mechanisms (tools and procedures) for implementing and exploiting libraries of model components. These mechanisms need to handle publication, discovery, subscription, versioning and implementation of models in different forms.

In this presentation a coherent suite of such mechanisms is proposed, using a scenario based on re-use of geosciences models. This approach forms the basis of a comprehensive strategy to empower domain modellers to create more interoperable systems. The strategy address a range of concerns and practice, and includes methodologies, an accessible toolkit, improvements to available modelling software, a community of practice and design of model registries.

These mechanisms have been used to decouple the generation of simplified data products from a data and metadata maintenance environment, where the simplified products conform to implementation styles, and the data maintenance environment is a modular, extensible implementation of a more complete set of related domain models. Another case study is the provisioning of authoritative place names (a gazetteer) from more complex multi-lingual and historical archives of related place name usage.