



## **Standardization and Interoperability across the Geosciences: Insights from GEON**

C. Baru

San Diego Supercomputer Center, University of California, San Diego (baru@sgsc.edu)

The Geosciences Network (GEON, [www.geongrid.org](http://www.geongrid.org)) is a project in the United States that is developing cyberinfrastructure for data discovery and integration across extremely heterogeneous Earth Sciences data. A number of lessons have been learnt in GEON in terms of supporting interoperability across widely varying, independently developed and managed data resources. This talk will present some of GEON's solutions and also discuss existing challenges in creating a truly interoperable framework for sharing Earth Science data.

GEON employs a service-oriented architecture and uses Grid certificates for user authentication. The Grid Account Management Architecture (GAMA) system is used to manage Grid certificates and to authenticate distributed service invocations. Resources (data, tools, Web services) that are registered in GEON are tagged with metadata. We have adopted the ADN (Alexandria DLESE NASA) metadata schema, which includes descriptive fields, spatial and temporal coverage fields, as well as metadata for education purposes. Data in a wide variety of formats can be registered with the system, including shapefiles, relational databases, Excel spreadsheets, delimited ASCII, and geoTIFFS and PDFs. GEON also provides the ability to "crawl" for WMS services and auto-register such services. Registration of netCDF data is currently in development. We are also developing a WMS service for accessing resources once they are discovered via GEON, and registration of WCS and publication of resources via WCS is also under consideration. While resources may be registered with GEON, it is important to note that GEON supports a distributed architecture. For example, relational databases may be registered with the system but can reside in a remote location, independently administered in a different "administrative domain". Access from the GEON Portal is via a "GEON-JDBC" interface which implements the Grid certificate-based authentication. GEON also provides mechanisms for knowledge

representation—for registering ontologies as well as to tie schema elements and data to ontologies. The OWL language is used as the standard for knowledge representation. Based on this, GEON has developed ODAL (Ontology Database Annotation Language) and SOQL (Simple Ontology-based Query Language) as the standards for data annotation and ontology-based querying.

While many challenges remain for interoperability, we highlight two important ones. At the systems level, there is a need for a common authentication system. Currently, the websites for each major project or database provider have their own user account management system. Thus, users cannot easily “surf” from one site to the next since they need to login using separate account information at each site. Adherence to an authentication standard and the use of a common system like GAMA can be a great benefit to the end user. Second, in a system like GEON, once a user discovers a set of data resources of interest, there is a need to “integrate” these data, e.g. using a common 3-D data model. The development of such a model and the corresponding mappings from existing data to the model remains a challenge.