Geophysical Research Abstracts
Vol. 19, EGU2017-9572, 2017
EGU General Assembly 2017

Reference architecture and interoperability model for data mining and fusion in scientific cross-domain infrastructures

Rainer Haener (1), Joachim Waechter (1), Sylvain Grellet (2), and Francois Robida (2)

(1) Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, CeGIT, Potsdam, Germany (rainer.haener@gfz-potsdam.de), (2) The French Geological Survey, Orleans, France

Interoperability is the key factor in establishing scientific research environments and infrastructures, as well as in bringing together heterogeneous, geographically distributed risk management, monitoring, and early warning systems. Based on developments within the European Plate Observing System (EPOS), a reference architecture has been devised that comprises architectural blue-prints and interoperability models regarding the specification of business processes and logic as well as the encoding of data, metadata, and semantics.
The architectural blueprint is developed on the basis of the so called service-oriented architecture (SOA) 2.0 paradigm, which combines intelligence and proactiveness of event-driven with service-oriented architectures. SOA 2.0 supports analysing (Data Mining) both, static and real-time data in order to find correlations of disparate information that do not at first appear to be intuitively obvious: Analysed data (e.g., seismological monitoring) can be enhanced with relationships discovered by associating them (Data Fusion) with other data (e.g., creepmeter monitoring), with digital models of geological structures, or with the simulation of geological processes.
The interoperability model describes the information, communication (conversations) and the interactions (choreographies) of all participants involved as well as the processes for registering, providing, and retrieving information. It is based on the principles of functional integration, implemented via dedicated services, communicating via service-oriented and message-driven infrastructures. The services provide their functionality via standardised interfaces: Instead of requesting data directly, users share data via services that are built upon specific adapters. This approach replaces the tight coupling at data level by a flexible dependency on loosely coupled services.
The main component of the interoperability model is the comprehensive semantic description of the information, business logic and processes on the basis of a minimal set of well-known, established standards. It implements the representation of knowledge with the application of domain-controlled vocabularies to statements about resources, information, facts, and complex matters (ontologies). Seismic experts for example, would be interested in geological models or borehole measurements at a certain depth, based on which it is possible to correlate and verify seismic profiles. The entire model is built upon standards from the Open Geospatial Consortium (Dictionaries, Service Layer), the International Organisation for Standardisation (Registries, Metadata), and the World Wide Web Consortium (Resource Description Framework, Spatial Data on the Web Best Practices). It has to be emphasised that this approach is scalable to the greatest possible extent: All information, necessary in the context of cross-domain infrastructures is referenced via vocabularies and knowledge bases containing statements that provide either the information itself or resources (service-endpoints), the information can be retrieved from.
The entire infrastructure communication is subject to a broker-based business logic integration platform where the information exchanged between involved participants, is managed on the basis of standardised dictionaries, repositories, and registries. This approach also enables the development of Systems-of-Systems (SoS), which allow the collaboration of autonomous, large scale concurrent, and distributed systems, yet cooperatively interacting as a collective in a common environment.