Geophysical Research Abstracts
Vol. 13, EGU2011-9509, 2011
EGU General Assembly 2011

An Ontology Driven Information Architecture for Interoperable Disparate Data Sources

John S. Hughes (1), Daniel Crichton (2), Sean Hardman (3), Ronald Joyner (4), Chris Mattmann (5), Paul Ramirez (6), and Sean Kelly (7)

(1) Jet Propulsion Laboratory, Caltech, Pasadena, CA, USA (Steve.Hughes@jpl.nasa.gov), (2) Jet Propulsion Laboratory, Caltech, Pasadena, CA, USA (Daniel.J.Crichton@jpl.nasa.gov), (3) Jet Propulsion Laboratory, Caltech, Pasadena, CA, USA (Sean.H.Hardman@jpl.nasa.gov), (4) Jet Propulsion Laboratory, Caltech, Pasadena, CA, USA (Ronald.S.Joyner@jpl.nasa.gov), (5) Jet Propulsion Laboratory, Caltech, Pasadena, CA, USA (Chris.A.Mattmann@jpl.nasa.gov), (6) Jet Propulsion Laboratory, Caltech, Pasadena, CA, USA (Paul.M.Ramirez@jpl.nasa.gov,), (7) Jet Propulsion Laboratory, Caltech, Pasadena, CA, USA (Sean.C.Kelly@jpl.nasa.gov)

An information architecture that enables interoperability between disparate data sources is described focusing on three components that are pertinent to the topic. The components are a federated registry, the Open Archive Information System (OAIS) Information Object, and domain ontologies. An OAIS Information Object consists of a data object and its representation information, for example an image and its descriptive metadata. These information objects are ingested into a federated registry as extrinsic registry objects, a type of object whose content is opaque to the registry. The use of extrinsic registry objects enables a single registry to manage more than one type of registry object. Unique characteristics that are required for registry object indexing, classification, and association are provided to the registry on ingestion. The domain ontologies are used to define, collect, and validate a data object's representation information, for example an image's data structure and its relationship to calibration information. Finally the use of existing standards for the federated registry, ontologies, and archive information objects provide the necessary foundation for successful interoperable data systems.

One or more domain ontologies are used to define the data structures, the metadata for the science interpretation of the data, the context within which the data was captured, processed, and archived, and the relationships between data. These definitions are used to capture and validate all metadata to be registered. In addition, these definitions can be used to configure a newly deployed registry, generate schemas for data labeling and validation, and be exported for user documentation.

Federated registries provide a distributed architecture with local governance and basic functionality for object tracking, versioning, locating, and retrieval. At ingest, registry objects are uniquely identified and versioned, classified, and cataloged automatically. Federated registries also provide notification of registry events to subscribers and federated query, linking, and replication of registry objects.

Experience also suggests that no single search paradigm will meet all requirements especially in situations where swift and appropriate responses are required. Forms-based, text-base, facet-based, and semantic search might all be required to meet user needs in situation such as natural disasters or for climate research. Using the architecture described previously, with registry object management independent of user search, new search applications can be quickly and efficiently deployed, often for ad-hoc purposes, by harvesting and filtering the metadata from the registries.

The Planetary Data System (PDS) has adopted this architecture for it next generation information system, PDS 2010. The authors will report on progress, briefly describe the components in more detail and illustrate how the new system will be phased into operations to handle both legacy and new science data. The resulting information system will help meet the expectations of modern scientists by providing more data interconnectedness, correlative science, and system interoperability across diverse data sources.