An Ontology Driven Information Architecture for Big Data and Diverse Domains

John S. Hughes, Dan Crichton, Sean Hardman, Ron Joyner, and Paul Ramirez
United States (steve.hughes@jpl.nasa.gov)

The Planetary Data System's has just released the PDS4 system for first use. Its architecture is comprised of three principle parts, an ontology that captures knowledge from the planetary science domain, a federated registry/repository system for product identification, versioning, tracking, and storage, and a REST-based service layer for search, retrieval, and distribution. An ontology modeling tool is used to prescriptively capture product definitions that adhere to object-oriented principles and that are compliant with specific registry, archive, and data dictionary reference models. The resulting information model is product centric, allowing all information to be packaged into products and tracked in the registry. The flexibility required in a diverse domain is provided through the use of object-oriented extensions and a hierarchical governance scheme with common, discipline, and mission levels. Finally all PDS4 data standards are generated or derived from the information model. The federated registry provides identification, versioning, and tracking functionality across federated repositories and is configured for deployment using configuration files generated from the ontology. Finally a REST-based service layer provides for metadata harvest, product transformation, packaging, and search, and portal hosting.

A model driven architecture allows the data and software engineering teams to develop in parallel with minimal team interaction. The resulting software remains relatively stable as the domain evolves. Finally the development of a single shared ontology promotes interoperability and data correlation and helps meet the expectations of modern scientists for science data discovery, access and use.

This presentation will provide an overview of PDS4 focusing on the data standards, how they were developed, how they are now being used, and will present some of the lessons learned while developing in a diverse scientific community.