The Planetary Science Ontology
A Case Study in an Ontology-Based Information Architecture

J. Steven Hughes, steve.hughes@jpl.nasa.gov
Daniel Crichton, dan.crichton@jpl.nasa.gov

Jet Propulsion Laboratory
Pasadena, CA 91016

The development of the Planetary Science Ontology started in 1985 as the key component of the Planetary Data System, NASA's official archive for the Planetary Science Community. With the primary requirement being the long-term archive of all science data collected from solar system exploration missions, it was recognized that a data model was required to guide the capture of the metadata needed to both catalog and describe the science products in the archive. The Entity-Relationship model was initially used to develop the data model. The data model was then realized in both the relational model and the Object Description Language (ODL), an object-based language developed to capture the metadata for the archive.

The planetary science data model includes all entities needed to identify, describe, and provide the context within which the science data were collected. These entities include target bodies such as planets and asteroids, spacecraft, instruments, missions, and personnel. The model also includes data product types that cover the range of science data structures such as images, time-series, spectra and qubes. Entities such as map projections and histograms are also available to provide additional semantic information. The Planetary Science Data Dictionary forms the basis of the model with about 1,200 data elements that range in scope from start_time, an element used across all planetary science domains, to elements that are specific to an individual instrument.

The Planetary Data System (PDS) archive contains science data collected from over 30 years of Solar System Exploration, including data from the Mariner, Viking, Voyager, Galileo, and Mars Pathfinder missions. In 2000, the total archive volume was about 5 terabytes of science data and metadata. The currently active 2001 Mars Odyssey mission will approximately doubled the archive volume by the end of 2004. The Mars Reconnaissance Orbiter scheduled to start returning data in 2007 will increase the archive by a factor of 10 and future missions are expected to increase the archive data volume by orders of magnitude. Typically a mission will create about 50 new product types. Some however, like the Cassini mission to Saturn will create over new 120 product types. Each product type is described using from 50 attributes to over 200 attributes. The number of products per product type range from a few hundred to a few million.

The PDS is a geographically distributed system, with the data curated by five science discipline nodes and three support nodes. The Geosciences, Atmospheres, Plasma Interactions, Rings, and Small Bodies nodes provide the science expertise needed for the development of discipline oriented data models and the production of archive quality data products. The Navigation and Imaging nodes are considered service nodes, providing navigational and image processing support respectively. The central node is responsible for overall management, system engineering, and data model and standards development. Science data to be included in the archive are essentially published, undergoing peer review of the data and metadata for scientific usefulness and validity. Ultimately, archived data can be cited in science articles. It is estimated that 25% of the PDS resources are budgeted toward maintaining and evolving the planetary science data model.

The unprecedented increase in the volume of data together with the advent of the Internet and user expectations for online access, forced the PDS to discontinue the use of optical media for the production and distribution of the archive. In October 2002, the PDS used the Object Oriented Data Technology (OODT) framework as the core infrastructure to successfully support the release and distribution of the science data for 2001 Mars Odyssey. This framework is characterized by the separation of the technology architecture from the data architecture and is based on the assumption that information technology will change at a much faster rate than the under-laying data model(s) within an enterprise. The PDS also wanted to take advantage of the wealth of metadata that is captured for the archive. The OODT framework now allows the PDS to leverage this metadata and provide seamless single-point-of-entry access to the geographically distributed data repositories and to support correlative searches across repositories.

Long recognized within the PDS for the critical role metadata plays in making the data scientifically useful for future users, there is now a renewed interest in formalizing the data model as an ontology and using it in a variety of new situations. For instance, the ability to do simple correlative searches across missions and instruments has prompted research into data mining within and across all product types. The unprecedented volume of data expected from MRO has suggested that ontology-based validation of the data product production pipeline be performed as early as possible in the mission life cycle to reduce costs. And finally, NASA requirements to support interoperability across the astrophysics, space physics, and planetary science archives will require the integration, and in some cases development, of domain ontologies.

This presentation will provide a brief overview of the PDS and the Planetary Science ontology including the early design of the data model, the data dictionary, and its use in creating the archive. Current work will also be described including the use of Protégé to capture a draft version of the PDS ontology, the development of an ISO/IEC 11179 based data element registry, and the use of the ontology for intelligent resource discovery across the distributed repositories. Future plans will also be briefly described including the use of ontology-based validation early in the mission life-cycle. Finally, the importance of ontology-based processing and its place in information architectures, such as that represented by the OODT framework will be emphasized with the PDS case-study as well as other deployments such as the National Cancer Institute's Early Detection Research Network.