Planetary and Space Science xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Planetary and Space Science

journal homepage: www.elsevier.com/locate/pss

Enabling interoperability in planetary sciences and heliophysics: The case for an information model

J. Steven Hughes[a,*], Daniel J. Crichton[a], Anne C. Raugh[b], Baptiste Cecconi[c], Edward A. Guinness[d], Christopher E. Isbell[e], Joseph N. Mafi[f], Mitchell K. Gordon[g], Sean H. Hardman[a], Ronald S. Joyner[a]

[a] *Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA*
[b] *University of Maryland, College Park, MD, USA*
[c] *Baptiste Cecconi, Paris Observatory, Paris, France*
[d] *Washington University, St Louis, MO, USA*
[e] *USGS Astrogeology Science Center, Flagstaff, AZ, USA*
[f] *Institute of Geophysics and Planetary Physics, UCLA, Los Angeles, CA, USA*
[g] *SETI Institute, Mountain View, CA, USA*

ARTICLE INFO

Keywords:
Interoperability
Information architecture
Information model
Planetary science
Heliophysics
Semantic
Digital repository
Ontology

ABSTRACT

The Planetary Data System has developed the PDS4 Information Model to enable interoperability across diverse science disciplines. The Information Model is based on an integration of International Organization for Standardization (ISO) level standards for trusted digital archives, information model development, and metadata registries. Where controlled vocabularies provides a basic level of interoperability by providing a common set of terms for communication between both machines and humans the Information Model improves interoperability by means of an ontology that provides semantic information or additional related context for the terms. The information model was defined by team of computer scientists and science experts from each of the diverse disciplines in the Planetary Science community, including Atmospheres, Geosciences, Cartography and Imaging Sciences, Navigational and Ancillary Information, Planetary Plasma Interactions, Ring-Moon Systems, and Small Bodies. The model was designed to be extensible beyond the Planetary Science community, for example there are overlaps between certain PDS disciplines and the Heliophysics and Astrophysics disciplines. "Interoperability" can apply to many aspects of both the developer and the end-user experience, for example agency-to-agency, semantic level, and application level interoperability. We define these types of interoperability and focus on semantic level interoperability, the type of interoperability most directly enabled by an information model.

1. Introduction

The Planetary Data System (PDS) is NASA's planetary science data archive and has the mission to provide the near-term discoverability, long-term preservation, and usability of the data returned from all NASA supported missions to explore the solar system. The digital repository currently contains about two petabytes of data from across a diverse set of science disciplines. The data was collected from over 1200 instruments, including both remote and in situ measurements.

After its initial release in 1990 and two decades of operations the PDS initiated development of its next generation system, PDS4. Based on lessons-learned, it is a complete redesign of the PDS to develop a system that meets the demands of higher data volume and to leverage new information technologies. The PDS4 architecture has two primary components: the information architecture and the software/technical architecture. The PDS4 Information Architecture (Hughes et al., 2009, 2014; Crichton et al., 2011) defines the system's informational requirements and institutes a multi-governance scheme for management of the architecture components. The PDS4 Information Model (Hughes et al., 2009, 2014), the core component of the architecture, is based on the ISO 14721 (ISO 14721, 2003) and ISO/IEC 11179 standards (ISO/IEC 11179, 2008). It was designed to be extensible across science disciplines and to promote interoperability that facilitates data sharing internationally.

The PDS4 Information Model provides a hierarchical structure for data archiving with three types of products. The Bundle Product is a list of related collections. The Collection Product is a list of related basic products of similar type, for example the spectral cubes from a single instrument. The Basic Product is the smallest unit of data registered and tracked under PDS4. A Basic Product may consist of an image but may

---

* Corresponding author.

http://dx.doi.org/10.1016/j.pss.2017.04.005
Received 15 December 2016; Received in revised form 25 February 2017; Accepted 6 April 2017

Please cite this article as: Hughes, J.S., Planetary and Space Science (2017), http://dx.doi.org/10.1016/j.pss.2017.04.005

also include supplemental information such as a related engineering table. The model defines four fundamental data structures: Array - a homogeneous n-dimensional array of scalars (e.g., images or spectral cubes); Table – the traditional fixed-width structure based on a single record with heterogeneous binary or character fields; Parsable Byte Stream – a stream from which the data value can be extracted directly by applying parsing rules to the bytes (e.g., simple text files, XML files, CSV tables); Encoded Byte Stream – a stream in which the bytes must interpreted, transformed, or otherwise processed before the data can be extracted (e.g., PDF files, JPEG images, MPEG movies) (Hughes et al., 2015; Raugh and Hughes, 2015).

The software/technical Architecture is a distributed service-oriented architecture encompassing the individual PDS discipline nodes and the PDS's international partners. The architecture provides consistent protocols for access to the data and services and a federated registry infrastructure to track and manage the contents of the digital repository. The current distributed search infrastructure is based on metadata harvested from product labels and loaded into Apache Solr (Apache Solr), an open source enterprise search platform.

PDS4 is the first operational science information system resulting from an information model-driven development methodology (Crichton et al., 2014). It is being used to coordinate data archiving in both the national and international planetary science communities. With the system's information requirements captured in an ontology modeling tool significant but controlled change can occur as the science domains and implementation technologies change.

The PDS4 Information Model enables interoperability across the Planetary Science and related space science disciplines. In general the term interoperable is or relates to the ability to share data between different computer systems. In the following more specific aspects of interoperability are described.

1.1. Agency-to-Agency level

At the Agency-to-Agency level independent systems do not share a common infrastructure but are interoperable because of a mutual interest in the information products.

This type of interoperability is supported by the underlying standards. A good example is the interface between PDS and its deep archive, the NASA Space Science Data Coordinated Archive (NSSDCA). The "information package", the information stored by the archive (ISO 14721, 2003), provides the interoperability link needed to connect the two and support this vital relationship. Commonality of structure and metadata concepts shared by both institutions simplifies the transfer of information and the core operations of the target (NSSDCA) process.

Another good example of agency-to-agency level interoperability is the interface between the PDS and the European Space Agency (ESA) Planetary Science Archive (PSA) (Besse et al., 2016). This interface enables queries between the two systems using the Planetary Data Access Protocol (PDAP) (IPDA Planetary Access Protocol, 2013) and a common set of keywords to ensure mirroring of resources.

Interfacing with other archives built on the same standards is accommodated by the common terminology and structural skeleton defined by the standards.

1.2. Semantic level

At the semantic level systems interoperate based on the commonality of definitions of key concepts. These common definitions present an interface between the systems. The common definitions can also be viewed as shared knowledge.

The development of the PDS4 Information Model (IM) and its partition into discipline namespaces is an application of this. The model-driven design paradigm prevents unintentional bifurcation of meaning and supports partitioning of the model into namespaces that can be mapped directly to and managed as distinct contexts. A namespace provides a unified set of attributes to define something like display orientation in all product contexts in which the concept is applicable. The common definition provides the basis for programmatic interoperability by providing developers with a single reference point for display information. And that, in turn, enables applications and other namespaces to take advantage of the established terminology to, for example, describe target orientation within a displayed image.

1.3. Application level

At the application level, the systems support interactions between disparate systems and make the interactions look seamless from the end user's perspective.

The EuroPlaNet (EPN) Table Access Protocol (TAP) (Erard et al., 2014) interface and Virtual European Solar and Planetary Access (VESPA) (Erard et al., 2015) projects are good examples of this - adding a software layer between application and target archive that allows a user to treat products from disparate sources as computationally equivalent. The PDS4 service structure and its Application Program Interfaces (APIs) are designed to support this sort of interoperability, and the PDS4 Information Model can support the semantic translation mapping needed to interface the PDS4 named concepts to those in the target environment.

2. A brief history of interoperability in the space sciences

Since before the advent of the World Wide Web, interoperability across space science digital repositories has been a goal. In 1982 and 1986 the Committee on Data Management and Computing (CODMAC) issued reports that set guidelines for the development of science data archives (Bernstein et al., 1982; Arvidson et al., 1986). The committee recommended that sufficient ancillary and metadata be captured and archived with the data to ensure that future users of the data would be able to understand how to interpret the science data formats as well as understand the context under which the data was collected and processed.

The Planetary Data System (PDS) was established in 1989 based on CODMAC principles. In 1999, after the advent of the World Wide Web, PDS deployed the PDS Distributed Inventory System (DIS) (Hughes and McMahon) which harvest metadata from PDS product labels and provided a product location and retrieval services across the PDS's heterogeneous and distributed nodes. Also in 1999, the Interoperable Systems for Archival Information Access (ISAIA) team (Hanisch, 1999) was formed as a collaboration of several space science repositories, including the PDS, with the ambitious goal to provide an “interdisciplinary data location and integration service for space science” (Hanisch, 2002). The importance of metadata standards was highlighted.

In 2001, Uschold (Uschold and Gruninger, 2004) argued that a “single shared ontology” is critical for developing a digital library that enables semantic interoperability across disciplines. And in 2002, the report prepared by the National Virtual Observatory Science Definition Team (Hanisch, 2002) emphasized standards for metadata and data formats for accessing large astronomical data sets.

The Space Physics Archive Search and Extract (SPASE) (Thieman et al., 2010), an international consortium formed in 2001, is a community-based effort to define standards and services for the Space and Solar Physics community focusing on the Heliophysics data environment. The goals were to define a data model for Space Physics and enable interoperability in a distributed environment to allow resources to be easily registered, found, accessed, and used. The SPASE Metadata Model is an information model for describing the elements of the Heliophysics data environment.

Over the ensuing decade and a half, there have been many successful efforts where shared knowledge has been collected in support of data access and interoperability. The PDS is represents a