Enabling Interoperability in Planetary Sciences and Heliophysics: The Case for an Information Model

J. Steven Hughes[1], Daniel J. Crichton[1], Anne C. Raugh[2], Baptiste Cecconi[3],
Edward A. Guinness[4], Christopher Isbell[5], Joseph Mafi[6], Mitchell K. Gordon[7],
Sean Hardman[1], Ronald Joyner[1]

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

[2]University of Maryland, College Park, MD, USA

[3]Baptiste Cecconi, Paris Observatory, Paris, France

[4]Washington University, St Louis, MO, USA

[5]USGS Astrogeology Science Center, Flagstaff, AZ, USA

[6]Institute of Geophysics and Planetary Physics, UCLA, Los Angeles, CA, USA

[7]SETI Institute, Mountain View, CA, USA

Abstract

"Interoperability" can apply to many aspects of both the developer and the end-user experience. The Planetary Data System PDS4 redesign effort included aspects of interoperable design not only in its information model, but in the development of that model and in the design and implementation of the infrastructure that supports the archive holdings. We discuss how issues of interoperability were addressed in each stage of the design and development process, focusing primarily on semantic interoperability.

Keywords

Interoperability, Information Model, Planetary Science, Heliophysics, Semantic, Digital Repository

1.0 Introduction

Advances in the development of information architectures and supporting technologies now make interoperability across Planetary Sciences and Heliophysics digital repositories possible. However the level of interoperability that can be attained is directly related to the amount of knowledge about the digital repositories that is commonly accepted and that can be shared. The challenge is to formally capture and share the knowledge necessary to meet interoperability requirements.

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 [1], 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.

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) [2, 3, 4] 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 seemless from the end user's perspective.

The The EuroPlaNet (EPN) Table Access Protocol (TAP) [5] interface and Virtual European Solar and Planetary Access (VESPA) [6] 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 Semantic Information in the Space Sciences

Since before the advent of the World Wide Web, shared knowledge and interoperability have been community objectives in the space sciences. In 1982 and 1986 the Committee on Data Management and Computing (CODMAC) issued reports that set guidelines for the development of science data archives [28, 7]. 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) [8] 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 [9] 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" [10]. The importance of metadata standards was highlighted.

In 2001, Uschold [11] 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 [10] emphasized standards for metadata and data formats for accessing large astronomical data sets.

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, but these successes have typically been limited to single communities. The PDS is one of the few successful efforts for an interdisciplinary community. There have been many lessons learned [12] for example:

- There is never a definitive, exhaustive source; and it is not uncommon to find contextual nuance at work in the use of discipline terms that are considered 'well-known'.
- The shared knowledge is almost impossible to manage as a single monolithic unit because of the disparate sources but should be partitioned in order that a multi-level governance scheme can be applied.
- The knowledge to be shared must be collected in a formal language using accepted standardized modeling methodologies otherwise inconsistencies and ambiguities will over time significantly degrade the effectiveness of the knowledge.
- The stability of an information system is highly dependent on the stability of the shared knowledge. But at the same time, the shared knowledge must evolve to remain relevant as the science discipline evolves over time.

2.0 Overview of the PDS4 Information Model

As part of its information architecture, the Planetary Data System (PDS) has developed the PDS4 Information Model[1] [2, 3, 4]. This model captures the knowledge about the planetary science digital repository at several levels of specificity and provides a means by which both humans and machines can "communicate" about the digital content of the repository. The PDS4 Information Model is also leveraged as a set of information requirements that drives the PDS4 Information Architecture [13] and enables interoperability across the diverse science disciplines in the planetary science community. Multi-level governance Figure 1, instituted at the common, discipline, and mission levels, enables interoperability, helps manage the complexities of development, and allows the model to expeditiously evolve over time.

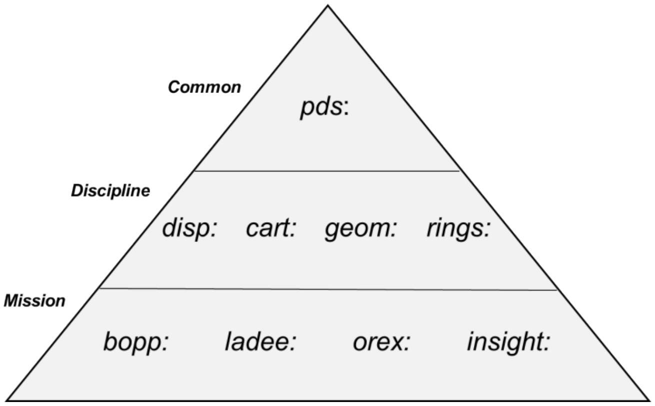

Figure 1- Mult-level Governance in the PDS4 Information Model

At the common or upper level of an information model resides the knowledge about what "things" (digital objects and products, in the case of the PDS archive) can be located and retrieved and how they are identified, referenced, and packaged. Digital objects in the repository must also have representation information provided in logical and well-defined terms so that they can be properly interpreted for scientific studies. At the next level, shared knowledge in specific disciplines must be available to understand and advance science, for example standard geometry models are needed to determine location and standard cartography models are needed for maps. Finally, a standard vocabulary is required within individual teams to communicate and effectively support the investigation.

---

[1] An information model in data engineering is a representation of concepts and the relationships, constraints, rules, and operations to specify data semantics for a chosen domain of discourse. It provides a sharable, stable, and organized structure of information requirements or knowledge for the domain context. [24]

3.0 Foundational Principles

To provide a stable foundation for the PDS4 Information Model and to keep it as broadly applicable as possible, several established meta-models were adopted as illustrated in Figure 2. The ebXML federated registry model [14] provided essential definitions for a federated registry, including the registry object and registry object typing, identification, and tracking. These definitions were incorporated into the model so that the model could in turn be used to configure a registry for specific types of registry objects. The Open Archival Information System (OAIS) Reference Model [1] provided core definitions for the model, namely the digital object, information object, information package, and a metadata classification scheme. The ISO/IEC 11179 [15] metadata registry reference model provided a data dictionary schema that is sufficient for defining science terms including their names, definitions and data types. When the data type is an enumerated list the definitions of each item in the list are provided. There is also the option to indicate units of measure and provide terminological entries for example, alternate-language definitions for any and all of the above.

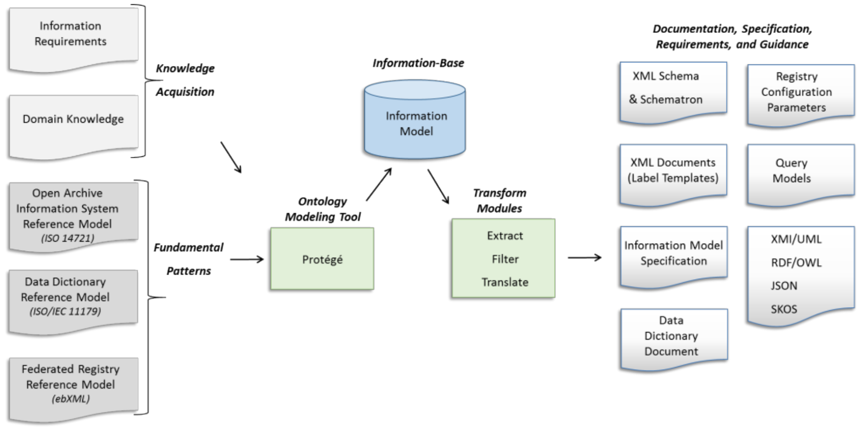

Figure 2- PDS4 Information Model

In the PDS4 Information Model each “thing of interest” to the community, for example a planetary image, is defined to the level necessary to meet the functional requirements of the system. This definition is the *Domain Knowledge* depicted in Figure 2. Sufficient representation information must be available for a digital object such as the planetary image to be interpreted and used in science research in the specific domain. In comparison, the information about the mission that collected the digital object may be limited to its name, a short description, and references to supporting documentation.

A fundamental principle used in the development of the PDS4 Information Model is that the model remains independent from its implementation. In a classic enterprise architecture such as that presented in the Zackman Framework of Enterprise Architecture [16] the architecture is partitioned into architectural elements, for example “why”, “how”, “what”, “who”, “where” and “when”. Each element is then modeled at five levels, from contextual to the detailed. The PDS4 Information Model encompasses the “what” element of the architecture, that is, the data being processed or archived. The model is agnostic to the other elements, especially the “how”, or the implementation element of the architecture. A model that is independent of the implementation is inherently more stable because it can more readily change as the “what” element in the science discipline changes. Concurrently it is shielded from changes in implementation technology which typically changes at more rapid pace.

To further manage complexities during the developmental and evolutionary phases of the PDS4 Information Model, the multi-level governance scheme is instituted at the common, discipline and mission levels. The common model is governed under a formal change control process where a change control board (CCB) decides whether to approve each change request based on the change’s potential impact on the overall PDS enterprise. At the discipline and mission levels similar governance but with contextually limited scope are instituted.

To promote stability an early PDS design policy was to limit the number of data structures for describing the digital objects in the repository. [17] These PDS4 fundamental data structures are the homogeneous n-dimensional array, fixed width binary and character tables, and the delimited table. It is expected these structures are sufficient for the majority of the digital objects in the repository.

For implementation, the contents of the PDS4 Information Model are extracted and translated to XML Schema [18, 19] and Schematron files [20] as illustrated in Figure 2. The XML Schema files are used to create XML [21] documents. Data providers subsequently populate the documents and validate the results using the XML Schema and Schematron files. The XML files are used to label the digital objects in the repository.

Version 1.0 of the PDS4 Information Model was released in 2012. A six month build cycle was established to provide a predictable and stable development schedule as the model continued to mature and discipline development began in earnest.

4.0 Discipline Level Models

To remain relevant in an ever changing science discipline; and in support of interoperability, the design process expected discipline extension to happen immediately, and the same development principles and tools are applied to the discipline-level and mission-level development as to the common part of the model. Having overarching discipline dictionaries to address themes like geometric metadata or display orientation metadata is a key way for PDS to ensure interoperability of tools on its own holdings. Several discipline level models have been or are currently under development:

4.1 Display model

The PDS4 display dictionary is a cross-discipline dictionary that is designed for use with image data that apply to many planetary science disciplines. The role of the PDS4 display dictionary is to provide a common method to map two spatial dimensions of an image array to the vertical and horizontal directions of a display device. The dictionary also contains attributes to specify which axis of a 3D image array should be used to as color bands such that a set of planes within that axis can be mapped to blue, green, and red channels of a color display.

4.2 Geometry model

The PDS4 geometry discipline model was developed to capture observational geometry metadata for planetary data. The geometry model provides a set of parameters that describe the conditions with which raw observational data are acquired and are thus important ancillary information for planetary data sets archived by the PDS. Accurate and well-defined geometry metadata is essential for processing the raw observational data into calibrated and derived data products, e.g., calibrated and map projected products. Geometry information captured in the PDS4 discipline model includes, for example, attributes for lighting and viewing angles, for position and velocity vectors of a spacecraft relative to the Sun and relative to the observing body at the time of an observation, and for the location and orientation of an observation projected onto the surface of a target body.

The content of the PDS4 geometry model has been developed from requirements gathered from domain experts in the planetary science community including researchers and data producers. The PDS4 geometry model provides consistency in geometry metadata incorporated in PDS4 data products across the wide range of planetary science disciplines and data collected by instruments observing many types of solar system bodies such as planets, ring systems, moons, comets, and asteroids. It also standardizes the definitions for the geometry attributes across PDS4 archives. Where there is overlap in concepts and terminology, the geometry model is consistent with usage in the PDS4 cartography model.

4.3 Cartography model

Generation and use of cartographic products are essential in support of scientific exploration and research. The PDS Cartography and Imaging Sciences node has lead a coordinated effort toward development of a discipline level Cartographic model in compliance with the primary PDS4 Information Model. The initial cartographic implementation utilized an existing terrestrial Federal Geographic Data Committee (FGDC) geospatial standard [22]. For PDS4, the FGDC standard has been extended and adapted in satisfying planetary requirements. For example, extension of existing and creation of new attributes and elements are required in order to describe tri-axial (implemented) and irregular shaped bodies (in progress, in coordination with geometry efforts), define map projection coordinate offset and origin parameters, and specify other unique parameters specific to planetary mapping needs. Implementation of these standards enables the scientific community the means of describing cartographic products used within planetary mapping and research, and satisfies short and long term usability and preservation requirements. Current and future utilization of the cartographic model across Mission and PDS archive activities will likely reveal additional requirements, and by design, influence intentional evolution of the model.

4.4 Planetary Plasma Interactions models

The PDS/PPI node has developed a number of separate discipline models in order to enhance the information that PDS metadata are able to provide for CDFformatted data files, which are the primary archival format for many of the MAVEN instruments. The data in a CDF file is stored in the form of single- or multi-dimensional arrays. While it is possible to describe these structures using the PDS core dictionary, there were no means of describing the logical relationships which exist between the various data arrays. In order to provide this information the PPI node created the following discipline level discipline models:

- Particle discipline model – This dictionary defines a series of the three classes which describe the relationship between the data array, and other supplemental arrays within a CDF file. Axis_Values identifies a 1-D array containing data associated with a single axis of the data array. Face_Values identifies a multi-dimensional array containing data associated with a "face" (i.e. multiple axes) of the data array. Aligned_Values identifies an array with the same dimensionality and axes as the data array which contains data which are supplemental to the data array values (e.g. uncertainties, etc.).
- Alt discipline model – This dictionary defines the Alternate_Values class which identifies arrays with the dimensionality that are equivalent in function and may be used interchangeably. An example would be multiple time columns in a single data file.

CDF allows for scalar values to be stored as data in form of a 1-D array with a single value. These values are captured in the PDS metadata using the "Parameters" class, which is defined in the MAVEN Mission discipline models.

4.5 Ring-Moon Systems models

The PDS4 rings discipline model was developed to capture ring specific supplemental metadata, and observational geometric metadata additional to that provided by the geometry discipline model. Fully describing ring observations is complicated. Each ring is composed of individual particles whose orbital velocities vary based on their radial separation from the central body. A ring may be inclined with respect to the equator plane of the central body and/or with respect to the other rings within the system. The local surface density, optical depth, and light scattering properties within a ring vary due to interactions at specific radial locations with satellites in the system, and ephemeral sub-kilometer structures (wakes) within the rings which are the result of gravitational interactions between ring particles.

The rings local dictionary currently contains more than 70 attributes, subsets of which support general ring system observations and observations of occultations of various sources by ring systems. There are three types of ring occultation observations, stellar – when the ring system passes between a star and the observer; solar – when the ring system passes between the sun and the observer; radio – when a spacecraft broadcasts a narrow radio signal through the rings to a receiver on the Earth. All three are supported by classes which can describe an occultation observation as a time series (how the data is typically captured), or as a derived radial profile of the ring or ring system.

A representative subset of the additional geometric parameters in the rings local dictionary includes a parameter which defines an inertial ring longitude that incorporates the inclination of the ring, a parameter which identifies the extent to which the rings are opened to the observer (from fully open on the illuminated side, through edge on, to fully open on the unilluminated side), a parameter which enables determining the orientation of the observation with respect to the orientation of the self-gravity wakes within the ring, and a parameter which enables determining the location of the observed portion of the ring with respect to the shadow of the parent body projected on the ring.

Since many of the occultation parameters defined in this model also are applicable to atmospheric occultations, a future iteration of the rings local dictionary will include a class or classes supporting atmospheric occultations.

5.0 The PDS4 Application

The PDS4 service structure and its Application Program Interfaces (APIs) are designed to support interactions between disparate systems and make the interactions appear seamless from the end user's perspective, as mentioned above. The PDS4 software is comprised of two main services that support this concept, the Registry Service which is based on the ebXML federated registry model [14] and the Search Service which is based on the Apache Solr [29] open source software. The architecture, depicted in Figure 3, shows metadata harvested from multiple sources into the Registry Service and then indexed and posted to the Search Service.

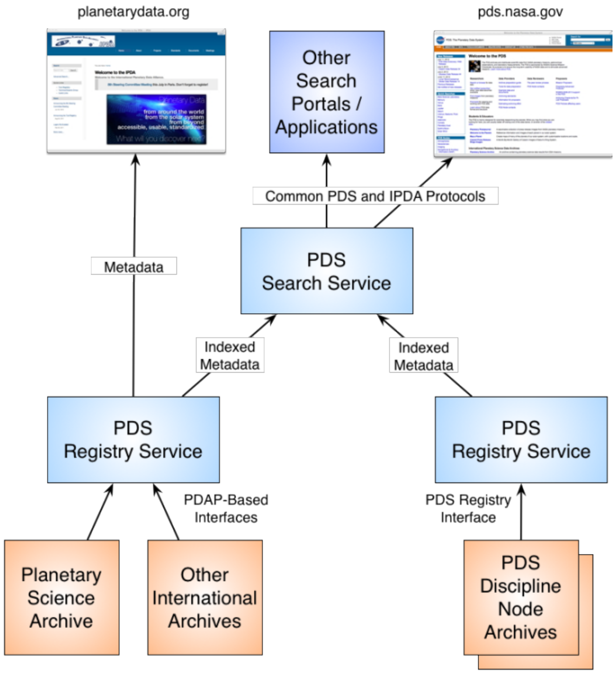

Figure 3 - PDS4 Service Architecture

The Search Service offers a Representational State Transfer (REST)-based interface supporting two protocols for product discovery. The first is a PDS homegrown protocol focusing on common search terms found in the PDS4 Information Model. The second protocol is the International Planetary Data Alliance's Planetary Data Access Protocol (PDAP) [30]. The previously mentioned EPN-TAP protocol is also under consideration for support by the Search Service.

Version 1.0 of the PDS4 Information Model was released in 2012 [23] and has been successfully used by three planetary science missions to archive their data, Lunar Atmosphere and Dust Environment Explorer (LADEE), Mars Atmosphere and Volatile EvolutioN (MAVEN), and Balloon Observation Platform for Planetary Science (BOPPS). It has also been recommended as the planetary science data standard by the International Planetary Data Alliance (IPDA).

The PDS Planetary Plasma Interactions (PPI) science discipline node, concurrently a member of the Heliophysics community, has had and continues to have an integral role in the development of the PDS4 Information Model. The PPI node has developed Heliophysics specific models and definitions for digital objects from the MAVEN mission.

6.0 Conclusion

Over the last decade or more there have been many attempts to enable interoperability between digital repositories in the Space Sciences. These attempts have typically focused on interoperability at the application layer to meet their goals, relying on system software, services, and protocols. However many have understood that a semantic level, or shared knowledge in the form of vocabularies, models, and/or ontologies, was essential to meet interoperability goals. The evidence is that in each attempt, the level of success was directly related to the level of effort applied developing knowledge about the targeted repositories, that could be shared. [25].

The PDS4 Information Model provides necessary and sufficient shared knowledge about digital objects in the PDS repository to drive and manage the PDS4 Information System. The PDS4 Information Model is implementation independent and remains relevant within the disparate set of evolving science disciplines through a multi-level governance scheme that provides a stable common level coupled with model development as needed at each and all levels.

In 2002, the National Virtual Observatory Science Definition Team said, "It is probably safe to say that no other professional community has reached the level of data interchange standards (both syntax and semantics) that we have reached in astronomy." [25] The PDS4 Information Model has enabled a new level of semantic interoperability across the diverse science disciplines of the Planetary Science community, including a Heliophysics sub-discipline. This shared knowledge can now be leveraged using software, services, and protocols to application level interoperability across the Planetary and Heliophysics science communities.

References

[1] ISO 14721:2003: Reference Model for an Open Archival Information System (OAIS), ISO, 2003.

[2] Hughes, J.S., Crichton, D. J., Mattmann, C.A., "Ontology-Based Information Model Development for Science Information Reuse and Integration", 10.1109/IRI.2009.5211603, IEEE International Conference on Information Reuse & Integration, 2009.

[3] Hughes, J.S.; Crichton, D.; Hardman, S.; Law, E.; Joyner, R.; Ramirez, P., "PDS4: A Model-Driven Planetary Science Data Architecture for Long-Term

Preservation", Data Engineering Workshops (ICDEW), 2014 IEEE 30th International Conference on, vol., no., pp.134,141, March 31 2014-April 4 2014.

[4] PDS4 Information Model Specification, PDS4 Information Model Specification Team, Version 1.7.0.0, Sep 28, 2016, https://pds.jpl.nasa.gov/pds4/doc/im/current/index_1700.html.

[5] Erard, S., Cecconi, B., et. al., The EPN-TAP protocol for the Planetary Science Virtual Observatory, arXiv:1407.5738 [astro-ph.IM], 2014.

[6] Erard, S., Cecconi, B., et. al., VESPA: Developing the Planetary Science Virtual Observatory in H2020, EPSC Abstracts, Vol. 10, EPSC2015-270, European Planetary Science Congress 2015.

[7] Arvidson, R.A., et al., Issues and Recommendations Associated with Distributed Computation and Data Management Systems for the Space Sciences, National Academy Press, 1986.

[8] Hughes, J.S., McMahon, The Planetary Data System Distributed Inventory System - An Experimental Interface to a Distributed Scientific Digital Library.

[9] Hanisch, Robert J., Distributed Data Systems and Services for Astronomy and the Space Sciences, Space Telescope Science Institute Baltimore, MD ADASS '99, Kohala Coast, Hawaii, 3-6 October 1999.

[10] Hanisch, Robert J., ISAIA: Interoperable Systems for Archival Information Access, NASA Technical Reports Service (NTRS), Sep 23, 2002.

[11] M. Uschold and Gruninger. M., "Ontologies and Semantics for Seamless Connectivity," SIGMOD Record, vol. 33, 2004.

[12] Hughes, J.S., Crichton, D., Simpson, R., Gordon, M., Joyner, R., Raugh, A., Guinness, E., Martin, M., The PDS4 Information Model Design Principles - How Well Did They Work?, Proceedings for the 2015 PV Conference, 3-5 November 2015, Darmstadt, Germany.

[13] Crichton, D., Hughes, J.S., Sean Hardman, S., Law, E., Beebe, R., Morgan, T., Grayzeck, E., "Scalable Planetary Science Information Architecture for Big Science Data", 2014 IEEE 10th International Conference on e-Science, Volume 2, 20-24 October 2014.

[14] Organization for the Advancement of Structured Information Standards, "OASIS/ebXML Registry Information Model V3.0", 2005.

[15] ISO/IEC 11179: Information Technology -- Metadata registries (MDR), ISO/IEC, 2008.

[16] Zachman, J. A., "A framework for information systems architecture," IBM Syst. J., vol. 26, pp. 276-292, 1987.

[17] Raugh, A. C., and Hughes, J. S. (2015) An Overview of PDS4 Archival File Format, Astronomical Data Analysis Software and Systems XXIV (ADASS XXIV), Proceedings of a conference held 5-9 October 2014 at Calgary, Alberta Canada. Edited by A. R. Taylor and E. Rosolowsky. San Francisco: Astronomical Society of the Pacific, 2015, p.551.

[18] XML Schema Part 1: Structures Second Edition, W3C Recommendation, 28 October 2004.

[19] XML Schema Part 2: Datatypes Second Edition, W3C Recommendation, 28 October 2004.

[20] ISO/IEC 19757-3 Information technology -- Document Schema Definition Languages (DSDL) -- Part 3: Rule-based validation -- Schematron

[21] Extensible Markup Language (XML) 1.0 (Fifth Edition), W3C Recommendation, 26 November 2008.

[22] Content Standard for Digital Geospatial Metadata, Metadata Ad Hoc Working Group, Federal Geographic Data Committee, FGDC-STD-001-1998, https://www.fgdc.gov/standards/projects/metadata/base-metadata/v2_0698.pdf.

[23] Hughes J.S., R. Beebe, D. Crichton, E. Grayzeck, L. Huber, L. Neakrase (2012) The Release of the PDS4 Data Standards for First Use by a Planetary Mission, 39th COSPAR Scientific Assembly, Abstract PSB.1-18-12, p. 782.

[24] Lee, Y. T. (1999). "Information modeling from design to implementation" National Institute of Standards and Technology.

[25] NVO Science Definition Team, “Towards the National Virtual Observatory”, A Report Prepared by the National Virtual Observatory Science Definition Team, 2002.

[26] W3C RDF/XML Syntax Specification (Revised), W3C Recommendation, 10 February 2004.

[27] OWL Web Ontology Language Reference, W3C Recommendation 10 February 2004

[28] Bernstein, R., et al., "Data Management and Computation, Volume 1: Issues and Recommendations", National Academy Press, 1982.

[29] Apache Solr, Apache Lucene, http://lucene.apache.org/solr/.

[30] IPDA Planetary Access Protocol, V1.0, April 16, 2013, https://planetarydata.org/standards/IPDA_PDAP_v1.0.pdf.