A Reference Model for Science Data Archives

J. Steven Hughes
steve.hughes@jpl.nasa.gov
NASA Jet Propulsion Laboratory (JPL)
California Institute of Technology

Earth Science Information Partners (ESIP)
Winter Meeting 2018
January 9-11th, 2018 – Bethesda, MD

2018 Geosemantics Symposium
Monday January 8, 2018
The Consultative Committee for Space Data Systems (CCSDS) Data Archive Interoperability (DAI) working group has developed the core elements for the underlying processes for digital preservation.

- **Reference Model for an Open Archival Information System (OAIS)** – ISO 14721, CCSDS 650.0-M-2
- **Information Preparation to Enable Long Term Use (IPELTU)**

The DAI working group is now addressing interoperable protocols and interface specifications.

- **Enable the access, merging and interoperable re-use of the data**
- **Support for the fundamental scientific technique of checking reproducibility of results.**
Abstract Protocol Architecture

Data Archive Ingest (DAI) WG Report to the CCSDS Management Council (CMC), Figure 2: Notional Data Archive Architecture, March 2017
The Data Archive Architecture Reference Model (DAARM) is an implementable model for trusted digital repositories¹.

- **Trusted Digital Repository** - A repository whose mission is to provide usable, long-term access to digital resources for a designated community.

- **The model is an integration of concepts and standards from:**
  - Open Archival Information System (OAIS) Reference Model¹
  - ISO/IEC 11179 Metadata Registry (MDR) standard
  - Three decades of digital repository development for science research.

- **The intended scope of the reference model is for digital archives in general.**

¹ ISO 14721:2012 (CCSDSS 650.0-P-1.1) Open archival information system (OAIS) -- Reference model
Figure 4-10: Information Object

1 ISO 14721:2012 (CCSDSS 650.0-P-1.1) Open archival information system (OAIS) -- Reference model
Information Categories

1. Identification - Provides a unique and immutable identifier for each data object.

2. Representation/Format - Provides meaning for a data object and allows it to be interpreted.

3. Integrity - Ensures the data object has not been unintentionally altered.

4. Provenance - Provides the history of the data object and is essential for authenticity and reproducibility.

5. Context - Describes the environment in which the data object was created.

6. Reference - Allows the data object to be referenced.

7. Access Rights - Defines the access restrictions pertaining to the data object, including the legal framework, licensing terms, and access control.

8. Quality* - Provides a scheme for assessing and assigning a quality measure to the data object.
Figure 4-1: OAIS Functional Entities

ISO 14721:2012 (CCSDSS 650.0-P-1.1) Open archival information system (OAIS) -- Reference model
Functional Entities

- **Ingest** - Accept information objects from producers, prepares them for storage, and ensures that they become established.

- **Archival** – Store and retrieve Information Objects.

- **Data Management** – Maintaining administrative information, for example consumer access statistics.

- **Access** - Make the archival information holdings and related services visible to Consumers.

- **Administration** - Control the operation of the other functional entities.

- **Preservation Planning** - Monitoring the environment to ensure that the information stored remains usable by the Designated Community.
Archival
Store and Retrieve Information Objects

Repository Service

operations
+putInformationObject()
+getDigitalObject()
+getServiceStatus()
...

Registry Service

operations
+RegisterInformationObject()
+ValidateInformationObjectMetadata()
+ValidateInformationObjectData()
+getInformationObjectMetadata()
+getInformationObjectData()
+getServiceStatus()
+getInformationObject()
Information Model Definitions

• “An information model is a representation of concepts, 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.

• Information Modeling is an essential discipline within Data Science

---

Information Model (IM)

Knowledge Acquisition

Information Base

Information Model

Ontology Modeling Tool

Protégé

Extract
Filter
Translate

Transform Modules

Documentation, Specification, Requirements, and Guidance

XML Schema
& Schematron

Registry Configuration Parameters

XML Documents
(Label Templates)

Query Models

Information Model Specification

XMI/UML
RDF/OWL
JSON
SKOS

Fundamental Patterns

Domain Knowledge

Information Requirements

Open Archive Information System Reference Model (ISO 14721)

Data Dictionary Reference Model (ISO/IEC 11179)

Federated Registry Reference Model (ebXML)
Information Model Roles

- **Requirements**: The IM is the primary source for information requirements.

- **Governance**: A multi-level governance scheme reduces the impact of change as the science community grows and evolves.

- **Semantics**: The IM provides named relationships to support semantic technologies.

- **Usability**: The IM provides the metadata needed to interpret and use the data.

- **Interoperability**: The IM is designed by discipline experts to provide interoperability, at multiple levels.

- **Configuration**: Extracts from the IM are used to configure tools and services.
Multi-level Governance

• Registration Authority
• Steward

Common

Discipline

Mission

1 ISO 14721:2003 - Open Archival Information System (OAIS) Reference Model
## Model Components
### Common, Discipline and Mission Dictionaries

<table>
<thead>
<tr>
<th>Registration Authority</th>
<th>Steward Id</th>
<th>Namespace Id*</th>
<th>XML Schema Namespace</th>
<th>Logical Identifier Prefix</th>
<th>Governance Level</th>
<th>Steward</th>
<th>Oversight</th>
</tr>
</thead>
<tbody>
<tr>
<td>0001_NASA_PDS_1</td>
<td>pds</td>
<td>pds</td>
<td><a href="http://pds.nasa.gov/pds4/pds/v1">http://pds.nasa.gov/pds4/pds/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Common</td>
<td>PDS EN Node*</td>
<td>CCB</td>
</tr>
<tr>
<td>0001_NASA_PDS_1</td>
<td>atm</td>
<td>atm</td>
<td><a href="http://pds.nasa.gov/pds4/atm/v1">http://pds.nasa.gov/pds4/atm/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Discipline</td>
<td>PDS ATM Node</td>
<td></td>
</tr>
<tr>
<td>0001_JAXA_DARTS_1</td>
<td>darts</td>
<td>darts</td>
<td><a href="http://pds.nasa.gov/pds4/darts/v1">http://pds.nasa.gov/pds4/darts/v1</a></td>
<td>urn:jaxa:darts:</td>
<td>Discipline</td>
<td>DARTS (JAXA)</td>
<td></td>
</tr>
<tr>
<td>0001_NASA_PDS_1</td>
<td>en</td>
<td>dph</td>
<td><a href="http://pds.nasa.gov/pds4/dph/v1">http://pds.nasa.gov/pds4/dph/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Discipline</td>
<td>PDS EN Node</td>
<td></td>
</tr>
<tr>
<td>0001_NASA_PDS_1</td>
<td>geo</td>
<td>geom</td>
<td><a href="http://pds.nasa.gov/pds4/geom/v1">http://pds.nasa.gov/pds4/geom/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Discipline</td>
<td>PDS GEO Node</td>
<td></td>
</tr>
<tr>
<td>0001_NASA_PDS_1</td>
<td>img</td>
<td>cart</td>
<td><a href="http://pds.nasa.gov/pds4/cart/v1">http://pds.nasa.gov/pds4/cart/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Discipline</td>
<td>PDS IMG Node</td>
<td></td>
</tr>
<tr>
<td>0001_NASA_PDS_1</td>
<td>img</td>
<td>disp</td>
<td><a href="http://pds.nasa.gov/pds4/disp/v1">http://pds.nasa.gov/pds4/disp/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Discipline</td>
<td>PDS IMG Node</td>
<td></td>
</tr>
<tr>
<td>0001_NASA_PDS_1</td>
<td>img</td>
<td>img</td>
<td><a href="http://pds.nasa.gov/pds4/img/v1">http://pds.nasa.gov/pds4/img/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Discipline</td>
<td>PDS IMG Node</td>
<td></td>
</tr>
<tr>
<td>0001_NASA_PDS_1</td>
<td>naif</td>
<td>naif</td>
<td><a href="http://pds.nasa.gov/pds4/naif/v1">http://pds.nasa.gov/pds4/naif/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Discipline</td>
<td>PDS NAIF Node</td>
<td></td>
</tr>
<tr>
<td>0001_NASA_PDS_1</td>
<td>ops</td>
<td>pds</td>
<td><a href="http://pds.nasa.gov/pds4/pds/v1">http://pds.nasa.gov/pds4/pds/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Discipline</td>
<td>PDS EN Node</td>
<td></td>
</tr>
<tr>
<td>0001_NASA_PDS_1</td>
<td>ppi</td>
<td>alt</td>
<td><a href="http://pds.nasa.gov/pds4/alt/v1">http://pds.nasa.gov/pds4/alt/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Discipline</td>
<td>PDS PPI Node</td>
<td></td>
</tr>
<tr>
<td>0001_NASA_PDS_1</td>
<td>ppi</td>
<td>particle</td>
<td><a href="http://pds.nasa.gov/pds4/particle/v1">http://pds.nasa.gov/pds4/particle/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Discipline</td>
<td>PDS PPI Node</td>
<td></td>
</tr>
<tr>
<td>0001_NASA_PDS_1</td>
<td>ppi</td>
<td>ppi</td>
<td><a href="http://pds.nasa.gov/pds4/ppi/v1">http://pds.nasa.gov/pds4/ppi/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Discipline</td>
<td>PDS PPI Node</td>
<td></td>
</tr>
<tr>
<td>0001_NASA_PDS_1</td>
<td>ppi</td>
<td>wave</td>
<td><a href="http://pds.nasa.gov/pds4/wave/v1">http://pds.nasa.gov/pds4/wave/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Discipline</td>
<td>PDS PPI Node</td>
<td></td>
</tr>
<tr>
<td>0001_ESA_PSA_1</td>
<td>psa</td>
<td>psa</td>
<td><a href="http://psa.esa.int/psa/v1">http://psa.esa.int/psa/v1</a></td>
<td>urn:psa:esa:</td>
<td>Discipline</td>
<td>ESA PSA</td>
<td></td>
</tr>
<tr>
<td>0001_NASA_PDS_1</td>
<td>rings</td>
<td>rings</td>
<td><a href="http://pds.nasa.gov/pds4/rings/v1">http://pds.nasa.gov/pds4/rings/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Discipline</td>
<td>PDS Rings Node</td>
<td></td>
</tr>
<tr>
<td>0001_NASA_PDS_1</td>
<td>rs</td>
<td>rs</td>
<td><a href="http://pds.nasa.gov/pds4/rs/v1">http://pds.nasa.gov/pds4/rs/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Discipline</td>
<td>PDS RS Node</td>
<td></td>
</tr>
<tr>
<td>0001_ROS_RSSA_1</td>
<td>rssa</td>
<td>rssa</td>
<td><a href="http://pds.nasa.gov/pds4/rssa/v1">http://pds.nasa.gov/pds4/rssa/v1</a></td>
<td>urn:ros:rssa:</td>
<td>Discipline</td>
<td>RSSA (IKI)</td>
<td></td>
</tr>
<tr>
<td>0001_NASA_PDS_1</td>
<td>sbn</td>
<td>sbn</td>
<td><a href="http://pds.nasa.gov/pds4/sbn/v1">http://pds.nasa.gov/pds4/sbn/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Discipline</td>
<td>PDS SBN</td>
<td></td>
</tr>
<tr>
<td>0001_NASA_PDS_1</td>
<td>sbn</td>
<td>sp</td>
<td><a href="http://pds.nasa.gov/pds4/sp/v1">http://pds.nasa.gov/pds4/sp/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Discipline</td>
<td>PDS SBN</td>
<td></td>
</tr>
<tr>
<td>0001_NASA_PDS_1</td>
<td>atm</td>
<td>ladee</td>
<td><a href="http://pds.nasa.gov/pds4/ladee/v1">http://pds.nasa.gov/pds4/ladee/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Mission</td>
<td>PDS ATM Node</td>
<td></td>
</tr>
<tr>
<td>0001_NASA_PDS_1</td>
<td>ppi</td>
<td>mvn</td>
<td><a href="http://pds.nasa.gov/pds4/mission/mvn/v1">http://pds.nasa.gov/pds4/mission/mvn/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Mission</td>
<td>PDS PPI Node</td>
<td></td>
</tr>
<tr>
<td>0001_NASA_PDS_1</td>
<td>ppi</td>
<td>mvn</td>
<td><a href="http://pds.nasa.gov/pds4/mvn/v1">http://pds.nasa.gov/pds4/mvn/v1</a></td>
<td>urn:nasa:pds:</td>
<td>Mission</td>
<td>PDS PPI Node</td>
<td></td>
</tr>
</tbody>
</table>
Information Requirements
Generated from the Dictionaries

Lines of XML Schema and Schematron

Common

Discipline

- Common
- Display
- Rings
- Cartography
- Geometry
- Imaging
Usability

• A desk assessment of PDS4 against ISO 16363\(^1\), the instrument for assessing a repository against the OAIS Reference Model\(^2\), found that 92% of the metrics of the ISO 16363 standard were satisfied

  • Governance and Organizational Viability

  • Digital Object Management

  • Infrastructure and Security Risk Management.

  • *Maintain the value of the data over time*

---

1 ISO 16363:2012 (CCSDS 652.0-R-1) Audit and certification of trustworthy digital repositories
2 ISO 14721:2012 (CCSDSS 650.0-P-1.1) Open archival information system (OAIS) -- Reference model
Next Steps

- **CCSDS Data Archive Interoperability (DAI) Working Group**
  - Write and review *CCSDS Blue Book*
  - CCSDS Fall 2018 Technical Meetings (Spring and Fall)
  - Develop two working prototypes
  - Reference Model Review
    - JPL's Center for Data Science and Technology – D. Crichton
    - NASA Planetary Data System (PDS) System Development – S. Hardman
    - CCSDS Systems Architecture (SAWG) Chair – P. Shames
    - JPL’s Multimission Ground System and Services (MGSS) Project – C. Radulescu
    - Life Storage of Mission Data (LSMD) task – M. McAuley
    - FernUniversität in Hagen - M. Hemmje
    - Engineering Data Management (EDM) task – L. Jewell
    - Information Retrieval and Data Science Group – C. Mattmann
Acknowledgements

• CCSDS Data Archive Interoperability (DAI) Working Group
  Bruce Ambacher
  Robert Downs
  John Garrett
  David Giaretta
  Matthias Hemmje
  Mike Kearney
  Terry Longstreth
  Don Sawyer

• JPL's Center for Data Science and Technology – Dan Crichton
• NASA Planetary Data System (PDS) – Sean Hardman, Ronald Joyner
• JPL Principal Data Scientist, USC Adjunct Associate Professor – Chris Mattmann
• JPL's Multimission Ground System and Services (MGSS) Project – Costin Radulescu
• CCSDS Systems Architecture (SAWG) Chair – Peter Shames
• Life Storage of Mission Data (LSMD) task – Mike McAuley
• Engineering Data Management (EDM) task – Laura Jewell

This research was carried out by the Jet Propulsion Laboratory, managed by the California Institute of Technology under a contract with the National Aeronautics and Space Administration.
References

- Data Archive Ingest (DAI) WG Report to the CCSDS Management Council (CMC), Figure 2: Notional Data Archive Architecture, March 2017
- DAI Architecture Analysis, SEA System Architecture WG, Slide 16, Alternative Standardized Archive System Architecture Deployment Option (3), May 2017
- CCSDS Reference Architecture for Space Information Management (RASIM) CCSDS 311.0-M-1
Thank You

Questions and Answers

PDS homepage: https://pds.nasa.gov/
• Utilize the Cornerstone Framework (NPO-49832) for model capture and management.
  – **Cornerstone is the framework used to capture and manage the PDS4 Information Model.**
  – **Provides a framework for model-driven information system development**
  – **Maintains Information Model independence.**
View Points

Community’s View

Information Modeler’s View

Repository View

Product

Tagged Data Object
(Information Object)

<local_identifier>MPFL_M_IMP_IMAGE</local_identifier>
<offset unit="byte">0</offset>
<axes>2</axes>
<axis_index_order>Last_Index_Fastest</axis_index_order>
<encoding_type>Binary</encoding_type>
<Element_Array>
<data_type>SignedMSB4</data_type>
<unit>pixel</unit>
</Element_Array>
<Axis_Array>
<axis_name>Line</axis_name>
<elements>248</elements>
<sequence_number>1</sequence_number>
</Axis_Array>
<Axis_Array>
<axis_name>Sample</axis_name>
<elements>256</elements>
<sequence_number>2</sequence_number>
</Axis_Array>
</Array_2D_Image>

Describes

Data Object
Conceptual Architecture

ISO 14721:2012 (CCSDSS 650.0-P-1.1) Open archival information system (OAIS) -- Reference model
• All registry objects are first class products.
  – *All products have a Persistent Identifier (PID)*
  – *Named relationships are used to relate objects* (semantic)
    • data, documents, people, software, and contextual objects
  – *Supports Linked Open Data.*