

Defining the Core Archive Data Standards of the International Planetary Data Alliance (IPDA)

J. Steven Hughes ⁽¹⁾, Dan Crichton ⁽¹⁾, Reta Beebe ⁽²⁾, Ed Guinness ⁽³⁾, David Heather ⁽⁴⁾, Joe Zender ⁽⁵⁾

⁽¹⁾ *Jet Propulsion Laboratory (JPL)*

Pasadena, California, USA

Email: {Steve.Hughes, Dan.Crichton}@jpl.nasa.gov

⁽²⁾ *New Mexico State University*

Las Cruces, New Mexico, USA

Email: rbeebe@nmsu.edu

⁽³⁾ *Washington University*

St. Louis, Missouri, USA

Email: guinness@wunder.wustl.edu

⁽⁴⁾ *ESA-ESTEC*

Noordwijk, The Netherlands

Email: dheather@rssd.esa.int

⁽⁵⁾ *ESA-ESTEC*

Noordwijk, The Netherlands

Email: jzender@rssd.esa.int

ABSTRACT

A goal of the International Planetary Data Alliance (IPDA) is to develop a set of archive data standards that enable the sharing of scientific data across international agencies and missions. To help achieve this goal, the IPDA steering committee initiated a six month project to write requirements for and draft an information model based on the Planetary Data System (PDS) archive data standards. The project had a special emphasis on data formats.

A set of use case scenarios were first developed from which a set of requirements were derived for the IPDA archive data standards. The special emphasis on data formats was addressed by identifying data formats that have been used by PDS nodes and other agencies in the creation of successful data sets for the Planetary Data System (PDS).

The dependency of the IPDA information model on the PDS archive standards required the compilation of a formal specification of the archive standards currently in use by the PDS. An ontology modelling tool was chosen to capture the information model from various sources including the Planetary Science Data Dictionary [1] and the PDS Standards Reference [2]. Exports of the modelling information from the tool database were used to produce the information model document using an object-oriented notation for presenting the model. The tool exports can also be used for software development and are directly accessible by semantic web applications.

Information Model, Object Model, Data Dictionary, Science Archive, Data Standards, Ontology

INTRODUCTION

The International Planetary Data Alliance (IPDA) [3] is a joint effort by national space exploration agencies, research institutions, and universities to enable global access and exchange of high quality planetary science data. A goal of the IPDA is to develop a set of archive data standards that guide the development of IPDA-compliant data systems. To help achieve this goal, the IPDA steering committee initiated a six month project to write the requirements for and draft an information model based on the Planetary Data System (PDS) archive data standards. This effort has focused on the development of a draft IPDA information model¹ consisting of a several object models and a data dictionary, with special emphasis on data formats.

A set of use cases for the IPDA archive data standards was compiled to document the possible ways in which users might reference and apply the standards to build compliant² local archives and share science data within the global planetary science community. The special emphasis on data formats was addressed by identifying the data formats that have been used by PDS nodes and other agencies in the creation of successful data sets for the Planetary Data System (PDS). The captured use case scenarios were the basis for the derived requirements.

The dependency of the IPDA information model on the PDS required the development of a formal specification of the PDS standards currently in use. This specification was captured in an ontology modeling tool to ease the model management and documentation effort for the project. In addition, exports of the specification from this tool into languages such as the XML Metadata Interchange language (XMI), Resource Description Framework (RDF), Web Ontology Language (OWL) were used to generate the information model document. These exports can also be used to guide information system implementation, generated code, and are directly accessible by semantic web applications.

This paper describes the process used to develop the draft IPDA Information Model, the final deliverable of the IPDA Archive Data Standards Requirements Identification project. The IPDA Reference System Architecture, outlined in Figure 1, suggests set of best practice specifications to be used for guiding the implementation of IPDA compliant archive data systems. Of the three core components, namely the process, data, and technology architecture standards, this project focused on the Information Model namely the object models and a data dictionary as highlighted in Figure 1.

¹ An information model is a type of data model that emphasizes adding meaning to data through relationships. Information models are also broader in scope than data models and typically involve large and complex domains. In general a data model is defined as an abstract model that describes how data is represented and used. To help differentiate, an example of a data model would be the Flexible Image Transport System (FITS) specification. An Astrophysics model that included the FITS specification and its relationships to other domain classes could be considered an information model.

² To enable sharing, a distinction is made between the characteristic of being compliant that suggests conformance to a common standard so that no mediation or translation is needed for interaction to occur and the characteristic of compatible that suggests the need for mediation or translation for interaction to occur.

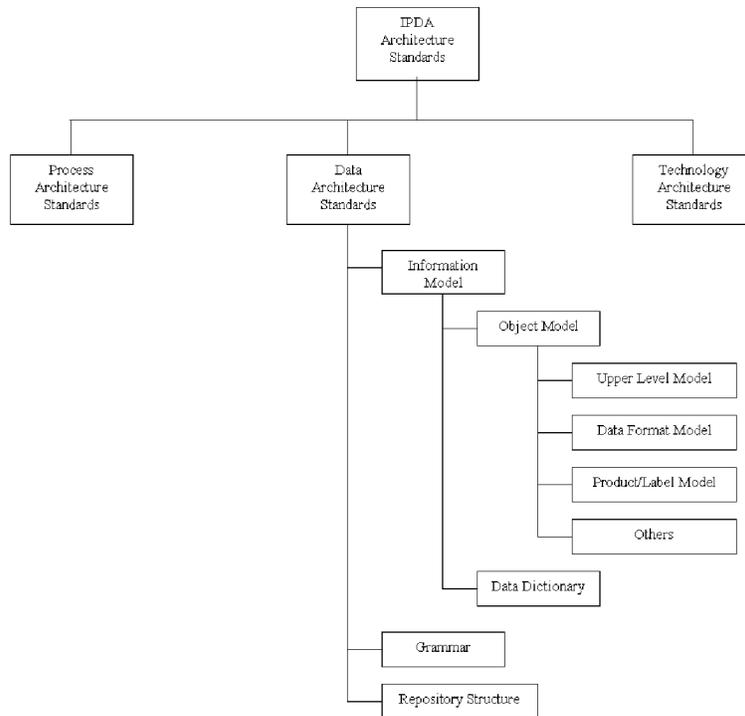


Figure 1 – IPDA Reference System Architecture

Scope

The IPDA Information Model consists of four related object models³ and the data dictionary. The first object model, the upper level object model, defines the key object classes that exist in the planetary science community. These include object classes such as mission, instrument, and data set and provide the science and programmatic context within which data products are collected and archived. The second object model, the data format object model, defines the object classes that describe the logical and physical structure of the digital data to be archived and include such commonly used object classes as Image and Table. The third or data product object model includes the object classes that are used to package the data and its metadata. For example an image data product is a package that contains a digital image and an instance of the Image object class that describes the structure of the digital image. The fourth object model includes all other object classes and will be more formally defined in future projects. Finally, the data dictionary is the set of attributes that have been used in any of the object class definitions.

Approach

The project's approach is illustrated in Figure 2, namely documenting the current PDS data model and then identifying core elements for consideration as archive data standards for the IPDA.

³ An object model is a type of model where object-oriented concepts, namely object classes are used as the key modeling construct.

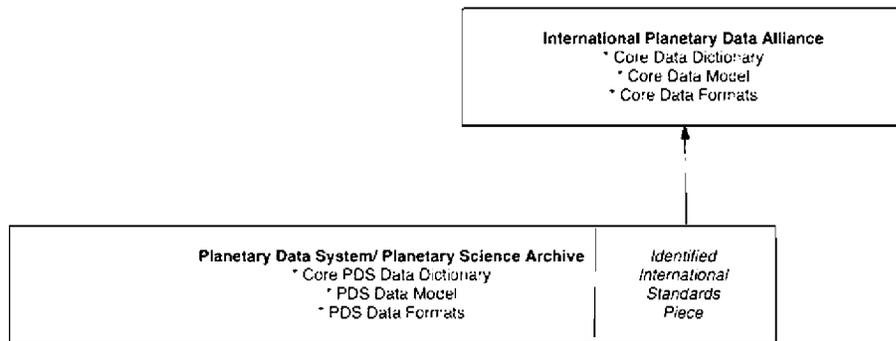


Figure 2 – Approach for Capturing Core Data Standards

As previously mentioned, the current PDS data model has been captured in an ontology modeling tool. Several sources have been used, including the Planetary Science Data Dictionary (PSDD), elements of the planetary science archive repository, and design documents produced during the PDS design and implementation phases from 1988 through 1990. In addition, each PDS discipline node has submitted successful data products from their local archive repositories to be considered for identifying the set of core data formats. The resulting database is the IPDA Information Model.

The IPDA Information Model is partitioned into four sections as mentioned above. The upper level object model is the first section. An abstract view of the PDS data model is shown in Figure 3 and the upper level object classes appear in the upper left hand part of the figure. It includes object classes such as Data Set, Instrument, and Target.

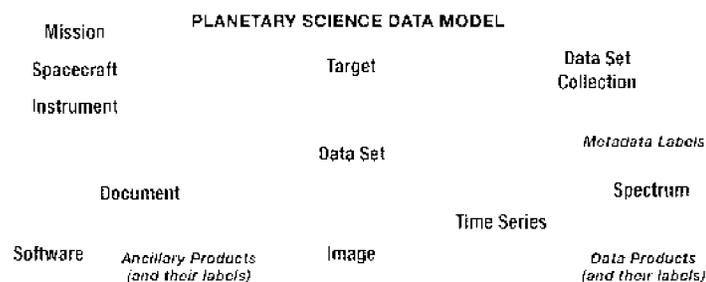


Figure 3 - PDS Data Model

The next section of the information model consists of the data format object classes. Data format object classes appear in the bottom right of Figure 3 and include such classes as Image and Time Series. Figure 4 shows a portion of a PDS image product label as an example of the how the PDS describes the data format of a digital camera image in the Object Description Language (ODL).

There are three groups of data format object classes defined in this section. The first group includes the generic data formats as currently defined in the PDS archive data standards. Generic data formats are defined as having both required and optional attributes, allowing for a great deal of flexibility in building descriptions for data objects. For example, in Figure 4, only the required attributes (e.g. LINES) have been used. However the generic definition of the IMAGE object class allows for many optional

attributes. (e.g. SCALING_FACTOR). The second group of object classes is derived from the generic data formats by using only the required attributes. These form the basis or parent classes for the object classes in the next group. Figure 4 is an example of a use of the parent IMAGE data format. The final group consists of the proposed core data formats. This group is derived from parent classes and the data formats extracted from the example data products submitted by the PDS Nodes. For example, since scaling_factor had been used extensively in successful imagin data sets, it is included as an attribute in the core data format for images.

The third section of the information model includes the object classes defined for the data formats extracted from the example data products. As the “raw” data formats submitted by the PDS Nodes, this section is primarily for informational purposes.

```

OBJECT = IMAGE
  LINES = 800
  LINE_SAMPLES = 800
  SAMPLE_TYPE = UNSIGNED_INTEGER
  SAMPLE_BITS = 8
END_OBJECT = IMAGE
  
```

Figure 4 - PDS Image Object

Section 4 of the information model focuses on the object classes needed to define the components of a data product label. As is evident in Figure 5, many of the components of an ODL label currently in use by the PDS are not formally defined as separate object classes but are simply groups of attributes delimited by comments. In the information model, these groups have been defined as object classes. These object classes are subsequently used to derive a very general set of proposed data product classes.

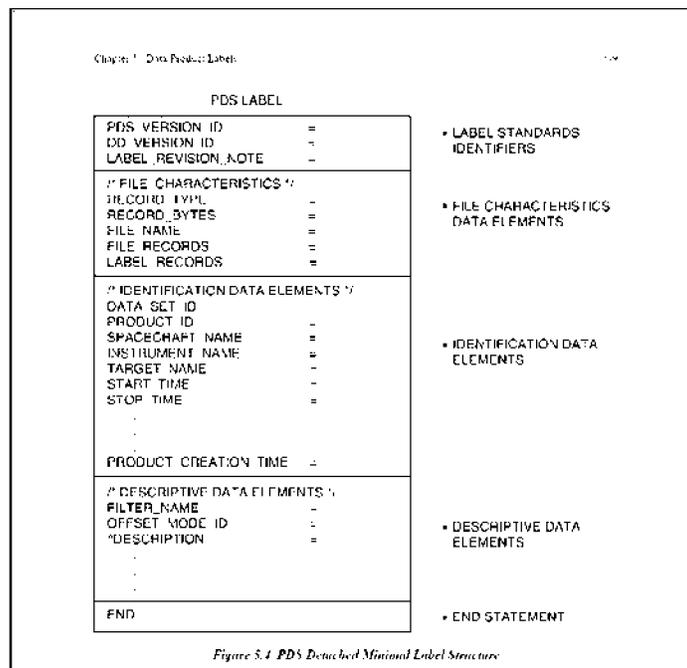


Figure 5 - PDS Label

The final section of the information model consists of the data dictionary. It includes only those attributes used in the object class definitions in the prior sections.

THE IPDA INFORMATION MODEL DOCUMENT

The IPDA Information Model has been captured in an ontology model tool. This tool provides a mechanism for comprehensively capturing an information model [4]. For example, at the minimum it provides the means to capture the information associated with a UML class diagram. It can also be used to capture data dictionary information for each item in the diagram.

The information model was extracted from the ontology modeling tool database and then used to generate the IPDA Information Model document. Each section of the document maps to a section of the information model described above. The information in each section is presented using several object-oriented notations where each of the notations was chosen for its ability to present an aspect of an object model. The “class hierarchy” and the Unified Modeling Language (UML) “class diagram” are two traditional notations and are used to “pictorially” illustrate the object classes and their relationships. A “table format” is used to “dump” a more complete class definition including the class hierarchy, class attributes, and their associations, both defined and inherited. Note that other notations can be generated and that choice of notation is dependant on what is best suited for the intended audience. The notations chosen for this version of the document are more engineering in nature.

In the following, the upper level section will be used to illustrate and explain the three notations used for all the succeeding sections. As mentioned above the upper level model includes the object classes that exist in the planetary science community and that provide a context within which science data products are collected, located, and used. For example, the Mars Viking Digital Image Mosaic is a data set created from images that were collected by the two vidicon cameras that flew on the Viking Orbiters. The upper level object model provides object classes for planetary missions, instruments, and data sets that are subsequently used to create objects that describe the Viking mission, the two Vidicon cameras, and the resulting data set. These objects and their relationships provide the context for the digital images collected.

Upper Level Class Hierarchy

A portion of the upper level object class hierarchy is illustrated in the following diagram. This diagram presents the “subclass of” relation among object classes. It is presented in a hierarchical (tree) format and provides a visual representation of the object classes in relation to their parent classes. The key concept of a class hierarchy is that a child “inherits” the definition of its parent. For example, in the following, the Spacecraft class inherits the definition of the Instrument_Host class. So a Spacecraft is an Instrument_Host, typically “specialized” with some additional attributes or constraints.

```
. Upper_Level_Object_Description
.. Data_Set
.. Instrument
.. Instrument_Host
... Earth_Based
... Rover
... Spacecraft
```

Upper Level UML Class Diagram

The upper level UML Class Diagram defines the object classes using UML software engineering standard notation. These diagrams were generated by exporting the tool’s database to the XML Metadata Interchange (XMI) language. The resulting ASCII file was loaded into a UML modeling tool and the UML class diagram was generated using a tool wizard. Due to XMI versioning problems, not all relations were correctly transferred between the tools. The resulting class diagrams subsequently had to be fixed by hand. It is expected that the versioning problems will soon be fixed by tool developers. It

should be noted that unlike relational models, the relations between object classes are single directional. Inverse relations are defined when necessary. Note that relationship names were deleted from the UML diagram to make the diagram simpler but the names are listed in the next section.

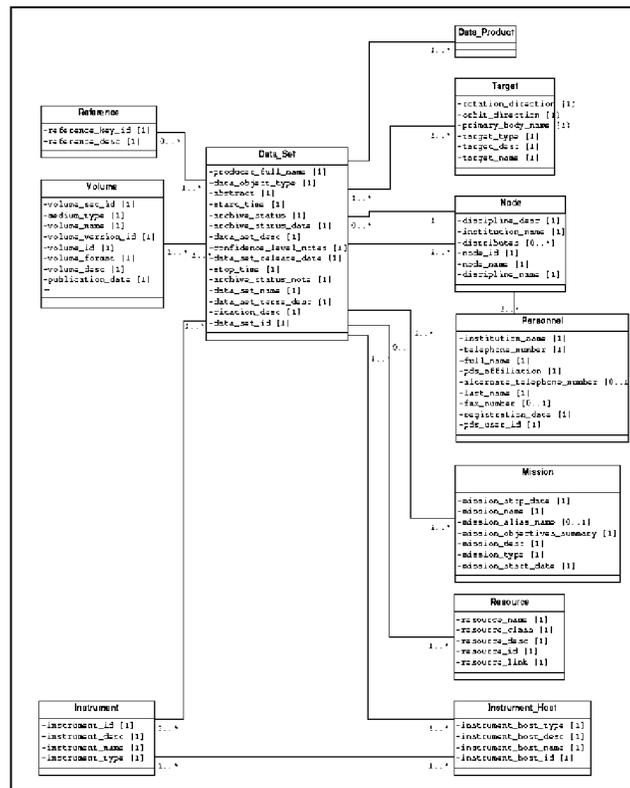


Figure 6 - Upper Level UML Class Diagram

Upper Level Object Model – Table Format

This section in the document presents the upper level object classes in a table format. The table includes the class hierarchy, class attributes, and class associations. The class attributes and associations listed include both those used to define the object class and those inherited from parent classes. Cardinalities are provided where appropriate. Figure 7 provide an example of the Data Set Object Class.

2.1 DATA_SET

Object Type: Upper_Level_Object_Description

Object Description: A collection of related data products

Relationship	Entity	Card	Value
Hierarchy	Upper_Level_Object_Description Data_Set		
Attribute	abstract_desc archive_status citation_desc confidence_level_note data_object_type data_set_desc data_set_id data_set_name data_set_release_date data_set_terse_desc producer_fullname start_time stop_time	1 1 1 1 1 1 1 1 1 1 1 1 1	
Inherited Attribute	none		
Association	curated_by distributed_by has_Host has_Instrument has_Mission has_Product Implicit has_Reference has_Resource has_Target has_Volume	1 1..* 1..4 1..* 1..* 1..* 0..* 1..* 1..4 1..*	Node Node Instrument_Host Instrument Mission Data_Product Reference Resource Target Volume
Inherited Association	none		

Figure 7 - Data Set Object Class

CONCLUSION

This paper describes the process used to develop the draft International Planetary Data Alliance (IPDA) Information Model, the final deliverable of the IPDA Archive Data Standards Requirements Identification project. This process is illustrated in Figure 8. The goal of the project was to identify a subset of the standards currently in use by NASA's Planetary Data System (PDS) that are appropriate for internationalization. The importance of this model can not be overstated since an information model is the foundation on which an information system is built. It guides the system's design and implementation by identifying and defining the items to be processed, the context for the items, and the relationships that provide meaning. In addition experience shows that the information model should be developed and maintained independent of the implementation choices since typically an information model evolves at a speed different from and outlasts any implementation technology choice.

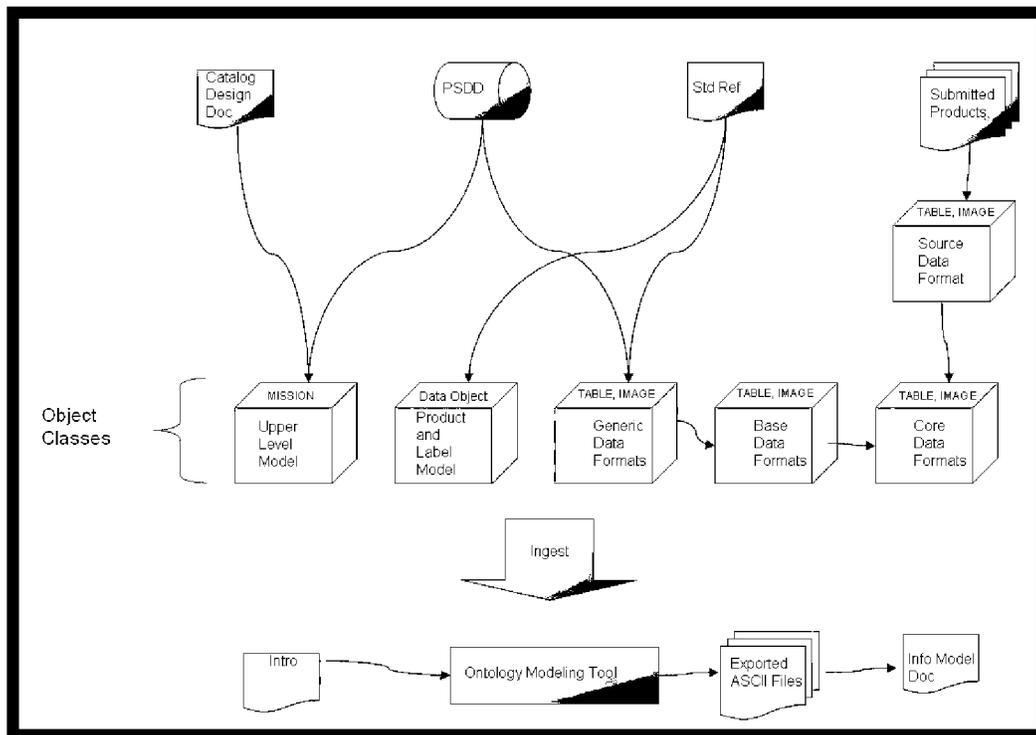


Figure 8 – Process for Producing the Information Model and its Document

For this project the information model was captured in an ontology modeling tool. This type of tool provides a comprehensive mechanism for capturing information models as was demonstrated by this project. Almost two decades of data modeling activities residing in various sources were compiled into a single database. The key goal of this project was to identify the core components of the PDS data model to be considered for adoption by the international community. The process for doing this has been successfully demonstrated by capturing examples of successful “core” data formats used by the PDS.

Finally, the information model is captured and maintain separate from its implementation. When requested, exports from the tool database, filtered as necessary, can be used to generate documentation customized to the audience. For example, the information model was documented using object-oriented notation for software engineers. The same exports can be used to generate software or accessed directly by semantic browsers.

Science information systems exist in complex and constantly evolving domains. The data in these domains range from binary data objects, to descriptions of the data objects, to descriptions of the descriptions, to descriptions of the instruments that collect the data objects. Experience shows that the development of a successful science information system is critically dependent on the existence of a comprehensive information model that formally defines the descriptions, relates them, and that exists independent of but drives the information system implementation. The second revolution of the web also dictates that information models be readily understood by the machine.

REFERENCES

- [1] - Planetary Science Data Dictionary Document, JPL D-7116, Rev. D., Jet Propulsion Laboratory, California Institute of Technology
- [2] - Planetary Data System, Standards Reference, March 20, 2006, Version 3.7, JPL D-7669, Part 2, Jet Propulsion Laboratory, California Institute of Technology
- [3] – Charter of the International Planetary Data Alliance, Ver.1.00, July 19, 2007
- [4] – Information Models as a Basis for Ontologies, Ed Barkmeyer, NIST, Ontolog Forum, April, 2007

End of File

