

International Planetary Data Alliance

**IPDA Information Model and Data
Dictionary**

Requirements

23 January 2009
Version 0.10090123

CHANGELOG

Revision	Date	Description	Author
Start Draft	2007-01-24	First Draft	S. Hughes
0.10070619	2007-06-19	Updated with review comments	S. Hughes
0.10070716	2007-07-16	Made suggested updates from splinter session.	S. Hughes
0.10081223	2008-12-23	Applied Assessment Project Recommendations	S. Hughes
0.10090123	2009-01-23	Applied comments from E.Guinness	S. Hughes

Table of Contents

1	INTRODUCTION	4
2	DEFINITIONS	4
3	NOTATION	6
4	APPLICABLE DOCUMENTS	6
5	REQUIREMENTS	7
5.1	Data Model	7
5.2	Data/Archive Producer	9
5.3	Interoperability	9
5.4	Software Development	10
5.5	Implementation	10
6	SOURCE REQUIREMENTS	11
	APPENDIX A – ASSESSMENT RECOMMENDATIONS	13

1 Introduction

The multi-disciplinary nature of planetary science and the increasing number of national space agencies involved in planetary exploration suggest the need for common data standards to improve access to and exchange of high quality planetary science data products across international boundaries. The purpose of this document is to initiate the development of a set of international science data standards by capturing requirements for the Data Model component of the archive data standards. [1]

It is accepted that inter-operability between archive repositories is critically dependent on compliance to a common Data Model with a data dictionary of terms, standard data formats, and a model of objects and their relationships.

The requirements captured in this document have been derived from use case scenarios for the IPDA data archive standards. These use cases are documented in the IPDA archive data standards use case document. [2]

Following the completion of the IPDA Data Model Requirements Identification Project on 16-July-2007, the IPDA Data Standards Assessment project was formed to review the requirements and draft Information Model and make recommendations. These recommendations are reflected in this update. [3,4,5]

2 Definitions

The following terms are used in this document.

1. **Actors** - An actor is a person, organization, or external system that plays a role in one or more interactions with the system.
2. **Archive** – An archive is an organization of people and systems that preserves information and makes it available for a designated community. It typically includes a data repository.
3. **Archive Package** – An archive package is a collection of science data and ancillary data that is being managed and preserved in an archive.
4. **Compatible** – Compatible is a characteristic that suggests the need for mediation or translation for interaction to occur.
5. **Compliant** – Compliant is a characteristic that suggests conformance to a common standard so that no mediation or translation is needed for interaction to occur.
6. **Consumer** – Consumer is the role played by those persons, or client systems, that interact with the services of a local instance of an IPDA compliant science data archive to find and acquire science data of interest.

7. **Data Dictionary** - A data dictionary is a set of metadata that contains definitions and representations for data elements.
8. **Data Element** – A data element is an atomic unit of data that has a identifier, a definition, one or more representation terms, optional enumerated values, and list of possible synonyms.
9. **Data Format** – A data format is a particular way to encode information in computer storage. It also represents a classification of data.
10. **Data Model** – A data model is a representation of the entities, properties, and relationships in an area of interest.
11. **Data Product** – A data product is a collection of one or more data files that contains science digital data and information about the data.
12. **Developer** – Developer is the role played by those persons that interact with the IPDA to develop software and systems for an IPDA compliant science data archive.
13. **Distribution Package** – A distribution package is a collection of data that has been prepared for distribution from an archive.
14. **Knowledge Base** - A knowledge base is a special kind of database for identifying, creating, representing, and distributing knowledge for reuse and learning across an organization.
15. **Manager** – Manager is the role played by one who sets overall policy for and is involved in day-to-day operations of a local instance of an IPDA compliant science data archive.
16. **Producer** – Producer is the role played by those persons, or client systems, which provide the information to be preserved by a local instance of an IPDA compliant science data archive.
17. **Repository** – A repository is a central place where data is stored and maintained.
18. **Standards Coordinator** – Standards Coordinator is the role played by an IPDA staff member to develop, document, maintain, and distribute the IPDA archive data standards.
19. **Submission Package** – A submission package is a collection of science data and ancillary data submitted to an archive with the intent that the package will be accepted as placed in the archive as an archive package.
20. **Scenario** - An imagined or projected sequence of events, esp. any of several detailed plans or possibilities.
21. **Use cases** - A use case describes a sequence of actions that provide something of measurable value to an actor.

3 Notation

The numbering of the requirements in this document will be formatted as **ADS.DM.N**, where:

- **ADS** is an acronym representing Archive Data Standards.
- **DM** is an acronym representing Data Model requirements.
- **N** is a unique number for the type of requirement.

Following the text of a requirement may be a reference to the requirement from which it was derived. The reference will be in parenthesis.

A paragraph following a requirement, which is indented and has a reduced font size, represents a comment providing additional insight for the requirement that it follows. This comment should not be considered part of the requirement for development or testing purposes.

4 Applicable Documents

Developing a Core Set of Data Standards for the IPDA, Concept White Paper, January 2007, <http://pds-engineering.jpl.nasa.gov/projects/ipda/ipda-whitepaper-20070207.pdf>.

IPDA Archive Data Standards Requirements Identification Project, Use Cases, Version 0.10070108, 08 January 2007, http://planetarydata.org/projects/copy_of_archive-standards/documents/use-cases-and-requirements/IPDA_SCWG_UseCases_060927.pdf.

International Planetary Data Alliance (IPDA), Information Model, June 29, 2007, http://planetarydata.org/projects/copy_of_archive-standards/documents/data-model/information-model/international-planetary-data-alliance-ipda-information-model/view.

IPDA Archive Data Standards, Requirements Identification Project, Draft Final Report, 16 July 2007, http://planetarydata.org/projects/copy_of_archive-standards/documents/data-model/information-model/ipda-archive-data-standards-requirements-identification-project-draft-final-report.

Assessment of the IPDA Data Standards, August 1, 2007.

Planetary Data System (PDS) Standards Reference, March 20, 2006, Version 3.7, JPL D-7669, Part 2.

Planetary Science Data Dictionary Document, August 28, 2002, Planetary Data System (PDS), JPL D-7116, Rev E.

PDS Information Model Specification, Version 0.070916t, September 8, 2008, http://pds-engineering.jpl.nasa.gov/system_eng/InfoModel/PDSInfoModel.pdf.

5 Requirements

The purpose of the following requirements is to guide the development of a set of core archive data standards for the IPDA. Specifically these requirements focus on a Data Model that will define the terminology, entities, relationships and formats needed to enable interoperability between local IPDA compliant science data archives. Local archive repositories that are compliant or compatible with these standards will promote global access to and the exchange of high quality planetary science data across international boundaries. In the following, the term Model will mean the IPDA Data Model.

5.1 Data Model

ADS.DM.1 - The Model shall consist of formal definitions of the terms, objects, data formats and their definitions in the Planetary Science domain. (3)

ADS.DM.1.1 - The Model shall have the planetary science domain as its scope. (ADS.DM.1)

Note: The Archive Data Standards splinter session held on 16 July 2007 at Caltech in association with the 2nd IPDA international meeting, agreed that the scope of the Model needed to be more clearly specified. The following classification hierarchy suggests a scope of what might be considered as the Model is assessed and continues to be developed.

Note2: The IPDA Data Standards Assessment Task has also proposed a data object hierarchy. This hierarchy is provided in the assessment recommendations appendix below.

Draft Classification Hierarchy

- Science
 - Upper Level Model (Data Set, Mission, Instrument, Data Product,...)
 - Data Format Model (Image, Table)
 - Interpretive Model
 - Map_Projection

- Engineering
 - SPICE
 - Packaging
 - Product Labels
 - Data Set Organization
 - Archive Package
- Administrative
 - Institution/Agency/Site
 - Personnel

ADS.DM.1.2 - The Model shall formally define a set of common data elements. (i.e. IPDA Data Dictionary). (ADS.DM.1)

ADS.DM.1.3 - The Model shall formally define a set of common data formats. (ADS.DM.1)

ADS.DM.1.4 - The Model shall formally define a set of planetary science entities and their relationships. (ADS.DM.1)

ADS.DM.1.5 - The Model shall be maintained by the IPDA standards coordinator and made accessible as hardcopy documents, from an IPDA website, and from a machine accessible knowledge base. (ADS.DM.1)

ADS.DM.1.6 - The Model shall be periodically reviewed by the IPDA standards coordinator to address new requirements. (ADS.DM.1)

ADS.DM.1.7 - The Model shall be periodically updated in accordance with the established procedures that specify the process for both identifying and implementing changes to the Model structure and/or content. (ADS.DM.1)

ADS.DM.1.8 - The Model will be designed to permit the addition or removal of entities.

ADS.DM.1.9 - The Model (or some supporting documentation for the Model) will establish nomenclature rules.

Note: As a simple example of a nomenclature conflict between the Model in this document and PDS is the construction of compound entities. PDS prepends modifiers (e.g., FITS_HEADER) while in this document they tend to be appended (e.g., HEADER_FITS), probably to facilitate grouping quasi-functionally when listed alphabetically. IPDA does not need to automatically conform to the PDS rules, but it does need to establish rules, or explicitly state allowed options.

5.2 Data/Archive Producer

ADS.DM.2 The Model shall provide the specification necessary for Data Producers to design and generate data products. (3)

ADS.DM.2.1 - The Model specification shall be available to any Data Producer desiring to place data into a local IPDA compliant archive. (ADS.DM.2)

ADS.DM.2.1.1 - The Model shall provide a set of common data formats that a Data Producer can use for structuring data. (ADS.DM.2.1)

ADS.DM.2.1.2 - The Model shall provide a set of standard objects and data elements that a Data Producer can use for describing data. (ADS.DM.2.1)

ADS.GR.n.n - The Archive Data Standards shall provide a specification for the use of the Object Description Language (ODL) in annotating the data using the Data Model. (ADS.DM.2)

This requirement is not a requirement on the Model but is included to point to the fact that a language (grammar) is needed to implement the Model.. It is expected to be a requirement on the Archive Data Standards Grammar.

ADS.DM.2.2 - The Model shall provide the necessary specification for the development of IPDA compliant archives. (3)

ADS.DM.2.2.1 - The Model shall be available to anyone desiring to create a new IPDA compliant archive or to make an existing archive IPDA compliant. (ADS.DM.2.2)

5.3 Interoperability

ADS.DM.3 - The Model shall support interoperability between distributed IPDA data archives. (3)

ADS.DM.3.1 - The Model shall provide the specifications necessary for supporting query and retrieval protocols between distributed IPDA compliant archives or IPDA compatible archives. (ADS.DM.3)

ADS.DM.3.1.1 - The Model shall provide a set of common terms to be used as query constraints for finding data and ancillary data from across distributed IPDA data archives. (ADS.DM.3.1)

ADS.DM.3.1.2 - The Model shall provide a set of common data formats that will be returned from distributed IPDA data archives. (ADS.DM.3.1)

ADS.DM.3.1.3 - The Model shall provide a set of common entities and their relationships to be used as query constraints for finding data and ancillary data from across distributed IPDA data archives. (ADS.DM.3.1)

ADS.DM.3.1.5 - The Model shall support the user interfaces of distributed IPDA data archives by providing information such as available data types, query parameters, entity descriptions, and data dictionary information. (ADS.DM.3.1)

5.4 Software Development

ADS.DM.4 - The Model shall provide the necessary specifications for Software Developers to develop software that is compliant with the IPDA Data Model. (3)

ADS.DM.4.1 - The Model shall be available for online access through an API from a knowledge base. (ADS.DM.4)

ADS.DM.4.2 - The Model shall provide the necessary specifications for developers of validation software for validating that data and data submission packages are compliant with the IPDA Data Model. (ADS.DM.4)

ADS.DM.4.3 - The Model shall provide the necessary specifications for developers of user interfaces for a) the identification of the types and formats of data available, b) the parameters available for query constraints, and c) searching, locating and retrieving data from distributed IPDA compliant repositories. (ADS.DM.4)

ADS.DM.4.4 - The Model shall provide the necessary specifications for developers of visualization, data process, and data analysis software that wish to be compliant with the IPDA Data Model. (ADS.DM.4)

ADS.DM.4.5 - The Model shall provide the necessary specifications for developers of a set of IPDA protocols for interoperating with both IPDA compliant and compatible archives. (ADS.DM.4)

5.5 Implementation

ADS.DM.5 - Procedures will be established to identify and implement changes to the Model structure and/or content.

6 Source Requirements

The requirements in this document have been derived from the IPDA Requirements, released on January 22, 2008.

1. IPDA will form an international alliance that will actively work with data providers who use its standards for archiving science data from planetary science missions

- 1.1 IPDA members will represent the interests of archiving activities at their respective international space agency or institution
- 1.2 IPDA will provide guidelines for data archiving functions including planning, implementing and operating planetary archive systems.
- 1.3 IPDA will provide guidelines and examples for designing, organizing and including data products and metadata in an archive
- 1.4 IPDA will provide guidelines for preparing and including documentation and reduction algorithms or software in an archive

2. IPDA will facilitate global access to international planetary science data archives

- 2.1 IPDA will develop recommendations for interoperability within a federation of international planetary data archive systems
- 2.2 IPDA will develop recommendations to support owners of international planetary science data archives in making their data available online
- 2.3 IPDA will encourage international planetary data archives to share and exchange data using IPDA data standards
- 2.4 IPDA will maintain a website to help planetary data providers and users to use IPDA standards

3. IPDA will develop, maintain and publish standards for archiving and sharing planetary science data among international archive systems

- 3.1 IPDA will provide standards for archiving of science data produced during planetary science research including related metadata, calibration data, ancillary data, documented reduction algorithms and processing software
- 3.2 IPDA will develop, maintain, and publish processes for maintaining IPDA data standards
- 3.3 IPDA will maintain a structured data dictionary containing definitions of data elements, their relations, and their scopes in aim to enable standardized descriptions of planetary science data

- 3.4 IPDA will maintain an Information Model of object classes, their attributes, and relationships to support the archive, search, and management of planetary science data
- 3.5 IPDA will define a standard grammar for describing planetary science data
- 3.6 IPDA will establish minimum required content for a planetary science dataset including both primary and ancillary data
- 3.7 IPDA will structure its data standards to allow planetary data systems to develop their own profiles, i.e. to adopt and extend the standards for local agency, mission and data provider uses
- 3.8 IPDA will develop and publish protocols for sharing data between planetary data systems
- 3.9 IPDA will publish standards for querying planetary data system catalogs including standard query models, protocols, and templates of user interfaces

4. IPDA will promote use of shared tools and services across archive systems in order to support scientific collaboration

- 4.1 IPDA will adopt existing international standards, where necessary, to ensure interoperability and reuse of existing scientific tools
- 4.2 IPDA will encourage member agencies to share, exchange and reuse tools as allowed by their local institutional policies

Appendix A – Assessment Recommendations

The following recommendations provided by the IPDA Data Standards Assessment task are included with their proposed resolution.

1. A revised version of the proposed IPDA Information Model, restricted to the proposed Model without reference to progenitors, should be produced. Users wishing to determine the progenitor associations should be referred to an updated version of this document.

Resolution: To differentiate between classes to be used as progenitors and those to be used by developers and producers to create instances, the Data Model shall indicate which classes are abstract classes. Abstract classes can not be instantiated.

2. The IPDA Information Model should be endorsed in principle and work should continue on this Model.

Resolution: During FY 2008 significant work has continued on the Information Model in association with the PDS3 Information Model Specification task and the follow-on PDS 2010 Data Architecture and Design tasks. Working groups for each task have been formed using PDS science discipline node and engineering node staff members. These working groups have addressed the majority of the recommendations listed in this section. The Information Model resulting from the PDS 2010 tasks will be submitted to the IPDA Information Model and Data Dictionary Working Group (IMDDWG) for review. At completion of its task, the IMDDWG will submit the Information Model to the IPDA for assessment.

3. An IPDA Information Model which includes all of the above Data Object Classes should be sufficient to address all of the data currently being archived by the planetary science community. However, the number of object classes can be reduced and still support the planetary science community. In particular the ASCII/Binary bifurcations can be removed for Table and the Series subclasses. The bifurcation was recognized as generating de facto subclasses by the ontology tool, and such explicit subclasses do offer some potential advantages. However the reduced structure obtained without those subclasses will make achieving archive compatibility simpler. Further simplification can be achieved by elevating the two File sub classes and removing the File Object Class. My recommended hierarchy is given at the end of this section. The result is 16 object classes, 10 subclasses and 1 sub-subclass.

Resolution: See Item 2.

4. Additional comments for specific Data Object Classes

- a. Alias – while this is a necessary object class, its use should be restricted to archiving agencies (not data providers). Its function should be to correct naming errors and to assist with achieving archive compliance with the IPDA Model.
- b. Bit_Column – recommend that this be deprecated in the IPDA Information Model. Bit_Column allows information to be stored in columns with bit rather than byte boundaries. It has been available for a very long time in the PDS but has seen only limited use. It offers very few advantages, and these are outweighed by the increased difficulty it imposes on general users of the data. I expect it to be deprecated in the next major revision of PDS.

Resolution: See Item 2.

- 5. At least one and possibly two additional data object classes will need to be added in the near future:
 - a. a movie object – currently there is no broadly accepted standard for this. IPDA should wait until the member agencies move closer to defining such an object before incorporating it in the Model. The movie object may prove to be vehicle for establishing procedures for joint develop of additional entities within the Model.
 - b. n-dimensional array data structure – I suspect that during the coming year, PDS will attempt to develop a new, simple, n-dimensional array data structure to replace the Array, Qube, and Spectral_Qube objects. Again IPDA should wait to include such an object and in the meantime should support the existing objects.

Resolution: See Item 2.

- 6. Sensibly, the Model is presented in this document at a very high level pending at least approval in principle for this approach and structure. assuming that approval, which I recommend, The next phase should provide:
 - c. detailed definitions of the various objects.
 - i. The current document identifies objects, but does not define them.
 - d. identify required and optional attributes (metadata) to be associated with each object.
 - ii. The attributes listed in the document describe the data (e.g., provide the information necessary to display an image) but do not describe the content of the data (e.g., attribute to identify the content of an image).
 - e. a draft IPDA specific data dictionary.
 - iii. The current document contains a limited set of items extracted directly from the PDS data dictionary, many of these need to be revised to become IPDA rather than PDS specific and additional items need to be included as the Model develops.

Resolution: See Item 2.

7. Recommended Data Object Classes in hierarchical form are:

- Data_Object_Description
 - Alias_Core
 - Array_Core
 - Bit_Column_Core
 - Column_Core
 - Container_Core
 - Element_Core
 - Field_Core
 - Explicit_File_Core
 - Implicit_File_Core
 - Header_Core
 - Header_FITS_Core
 - Header_VICAR_Core
 - Histogram_Core
 - History_Core
 - Image_Core
 - Banded_Image_Core
 - Simple_Image_Core
 - Qube_Core
 - Spectral_Qube_Core
 - Table_Core
 - Gazetter_Core
 - Index_Table_Core
 - Palette_Core
 - Spreadsheet_Core
 - Series_Core
 - Time_Series_Core
 - Spectrum_Core

Resolution: See Item 2.