International Planetary Data Alliance

IPDA Archive Data Standards

IPDA Data Model Requirements Identification Project

Requirements

DRAFT

19 June 2007
Version 0.10070619

CHANGE LOG

Revision	Date	Description	Author
Start Draft	2007–01–24	First Draft	S. Hughes
0.10070619	2007-06-19	Updated with review comments	S. Hughes

Table of Contents

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.

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 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. **Data Dictionary -** A data dictionary is a set of metadata that contains definitions and representations for data elements.
7. **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.
8. **Data Format** – A data format is a particular way to encode information in computer storage. It also represents a classification of data.
9. **Data Model** – A data model is a representation of the entities, properties, and relationships in an area of interest.
10. **Data Product** – A data product is a collection of one or more data files that contains science digital data and information about the data.
11. **Distribution Package** – A distribution package is a collection of data that has been prepared for distribution from an archive.
12. **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.
13. **Repository** – A repository is a central place where data is stored and maintained.
14. **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.
15. **Scenario** - An imagined or projected sequence of events, esp. any of several detailed plans or possibilities.
16. **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 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.

4.1 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. (1,5 – See section 5)

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

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 and updated to address new requirements. (ADS.DM.1)

4.2 Data/Archive Producer

ADS.DM.2 The Model shall provide the specification necessary for Data Producers to design and generate data products. (1,2)

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. (1,2)

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)

4.3 Interoperability

ADS.DM.3 - The Model shall support interoperability between distributed IPDA data archives. (1,2,3,4)

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)

4.4 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. (1,2)

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 Requirements

The requirements in this document and the use cases from which they were derived have the following requirements as a basis. The following were extracted from the NASA-PDS/ESA-PSA Planetary Data Interoperability White Paper and the IPDA charter document. Another guiding reference is the OAIS Reference Model

1. The IPDA shall provide standards for planetary science archives that will improve the global access to and exchange of high quality planetary science data products across international boundaries.
2. A standard protocol that addresses query, data modeling, and data formatting aspects of the interaction will be adopted.
3. Respond with metadata corresponding to matches (including pointers to the real data).

4. Convert semantically varying "keywords" (that might depend on specific projects) to a set of "standardized" keywords.
5. The overall activity for developing an international standard must evolve over time.