International Planetary Data Alliance

IPDA and PDS Information Model Project Report

IPDA Information Model Task Group

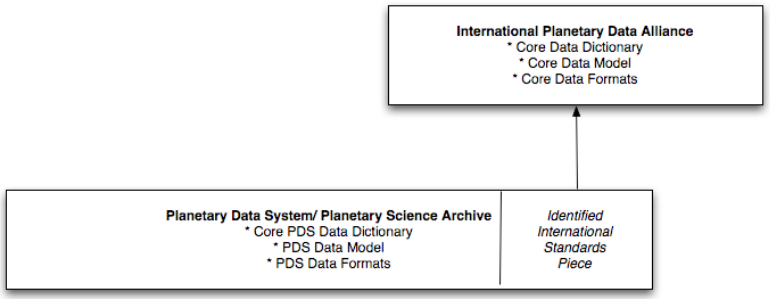
IPDA Steering Committee Meeting
July 2, 2009
Rome, Italy

Topics

- **Past and Current IPDA and PDS Tasks**
- **Proposed IPDA Task**
- **PDS4 Task**

The IDPA and the PDS4 Data Design Task Background

"The data standards within the IPDA, including the data models and derived dictionaries, are based on the NASA Planetary Data System (PDS) standard that is the de-facto standard for all planetary data at the time of the IPDA founding".
Charter of the International Planetary Data Alliance,
3rd Draft, May 24, 2007



Completed IPDA Data Standards Tasks

- **Archive Data Standards Requirements Identification Project**
  - IPDA preliminary model should be derived from NASA's Planetary Data System data standards - IPDA SC Meeting, Nov 2006
  - Deliverables:
    - IPDA Archive Data Standards Use Cases and Requirements
    - IPDA Common Data Model Document – Preliminary
- **Standard Data Model Assessment**
  - An independent assessment shall be executed on the preliminary model – IPDA SC Meeting, Jul 2007
  - Key Finding - The Information model should be endorsed in principle and work should continue on the model.
- **Data Dictionary Model**
  - The data dictionary model can be drafted in parallel with the ongoing data model assessment and modeling activities - IPDA SC Meeting, Jul 2007
  - Recommendation – Adopt the ISO/IEC 11179 Metadata Registry Standard for the IPDA data dictionary.

Current IPDA Data Standards Project

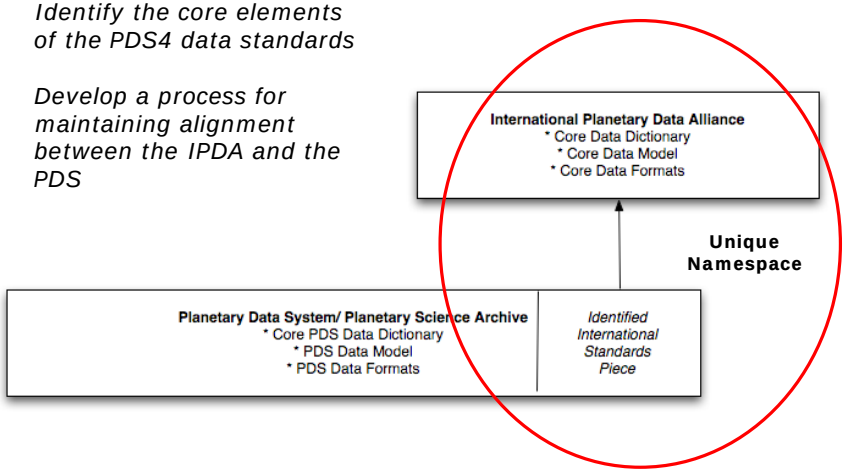
- **IPDA Information Model and Data Dictionary**
  - The IPDA Steering Committee directed that the IPDA Information Model and Data Dictionary development should continue. IPDA SC Meeting - July 2008.
  - The recommendations provided by the IPDA Core Standards Assessment Project were addressed.
    - Requirements document was updated. http://planetarydata.org/standards
    - The startup of the PDS4 data standards task impacts the adoption of PDS standards by the IPDA.
      - The IPDA project has been observing the PDS4 data standards task.
      - The PDS data standards task is six months into an 18 development cycle.
  - **Recommendations**
    - The IPDA and PDS4 data standards efforts should be aligned.
    - Propose that a new IPDA project be formed for this purpose.

Proposed IPDA Data Standards Project

- **IPDA and PDS Data Standards Alignment**
  - Review and comment on the preliminary PDS4 data standards.
    - Aligns with PDS4 Data Design Working Group (DDWG) request for outside reviewers.
    - Items to be reviewed include the Information Model, Data Dictionary, Grammar, and the Standards Reference
  - Identify the core elements of the PDS4 data standards
    - Propose as core data standards for the international planetary science community
    - Propose that a unique namespace be created for the core data standards.
  - Develop a process for maintaining alignment between the IPDA and the PDS
    - Work with PDS to allow IPDA representation in the PDS standards development process.

Proposed IPDA Data Standards Project

- *Identify the core elements of the PDS4 data standards*
- *Develop a process for maintaining alignment between the IPDA and the PDS*



Standards Governance Issues

- **Two distinct standing committees should exist, one for the IPDA and one for the PDS.**
  - Each has one representative from the other committee as a member with full voting rights.
    - Scope is limited to data standards that impact the international community.
- **Standards Documents**
  - Common
    - versus -
  - Distinct with references and replicated content*.
- **Website**
  - Distinct with references and replicated content.

* Replicated content will be registered and with notifications sent when changes occur.

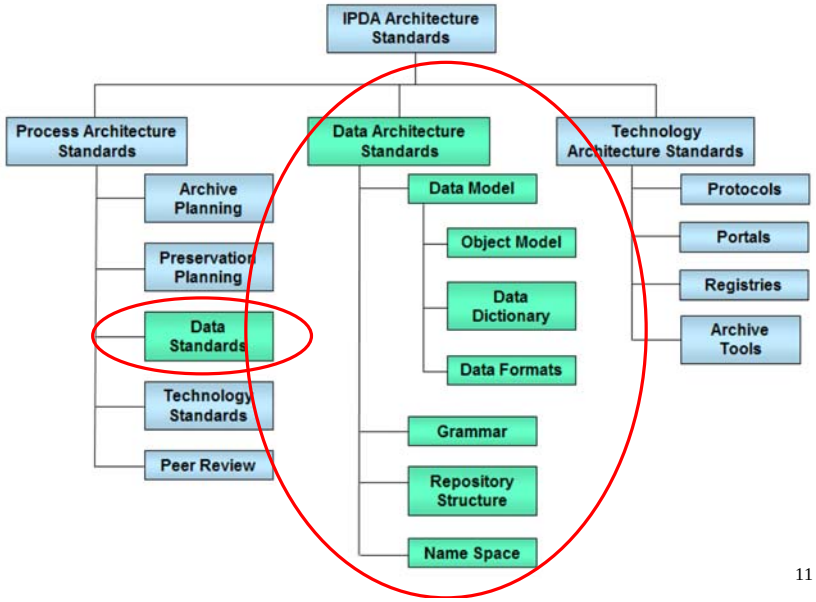
The PDS4 Data Standards Task

- **The PDS formed a working group to review the results of the IPDA Common Data Modeling task, IPDA Project #1.**
  - Identified core elements.
  - Identified over 60 problems, issues and anomalies associated with the standards.
- **The PDS decided to develop the PDS4 data standards based on the core elements and to address the problems found.**

PDS4 Data Design Working Group

- **Formed in December 2008**
- **Personnel**
  - Engineering and Science staff from each of the PDS Nodes.
  - Now requesting review by larger community.
- **Weekly Telecons**
- **Project Website**
  - http://pds-engineering.jpl.nasa.gov/system_eng/PDS4_Data_Design/index.html
  - The design documents and label schemas are re-generated after each update to the data model.

IPDA Data Architecture Standards



PDS4 Key Goals

- **Enable a stable and usable long-term archive.**
- **Enable more efficient archive preparation for data providers.**
- **Enable services for the data consumer to find the specific data they need and provide the formats they require.**

PDS4 Design Principles

- **The data model:**
  - is defined in a formal language
  - is independent of implementation
  - defines a few fundamental data structures that do not evolve over time
  - is extensible enabling it to handle more complex data formats
- **The archive data formats shall be designed independent of data provider and data consumer formats.**
- **The data architecture shall include a standard data dictionary model.**

DDWG Deliverables

- **PDS4 Information Model**
  - The Information Model defines PDS object classes. This includes data structures, formats, and products as well as data sets, documents, missions.
- **PDS4 Data Dictionary Model**
  - The Data Dictionary Model provides the schema for the PDS data dictionary. The data dictionary documents the data elements used in the PDS4 Information Model.
- **PDS Standards Reference V4.0**
  - The PDS Standards Reference V4.0 will be written in the format and tone of a standards reference document.
- **Grammar Options**
  - The Grammar is used to capture PDS archive metadata for product labels.

PDS4 Data Design Accomplishments - General

• Project and Project Members Defined • Principles and Drivers updated	**Done**
• General Data Model (Draft) • Product Data Model (Draft) • Data Dictionary Model (Draft) • Grammar Options • PDS Standards Reference (Outline)	**Substantial Progress**
• Data Dictionary Model (Final) • Grammar Decision • PDS Standards Reference (Draft)	**Next 3 Months**
• PDS and community wide review • General Data Model (Final) • Product Data Model (Final) • PDS Standards Reference (Final)	**Due 9/31/10**

PDS4 Data Design Accomplishments - Details

- **Four basic data structures**
  - Homogeneous N-dimensional array of scalars – Array_Base
  - Heterogeneous repeating record structure of scalars – Table_Base
  - Unencoded byte stream
  - Encoded byte stream

- **Array_Base**
  - Array_2D
    - Image_Grayscale
    - Spectrum_2D
  - Array_3D
    - Image_3D
    - Movie
- **Table_Base**
  - Table_Character
  - Table_Binary
  - Table_Binary_Grouped

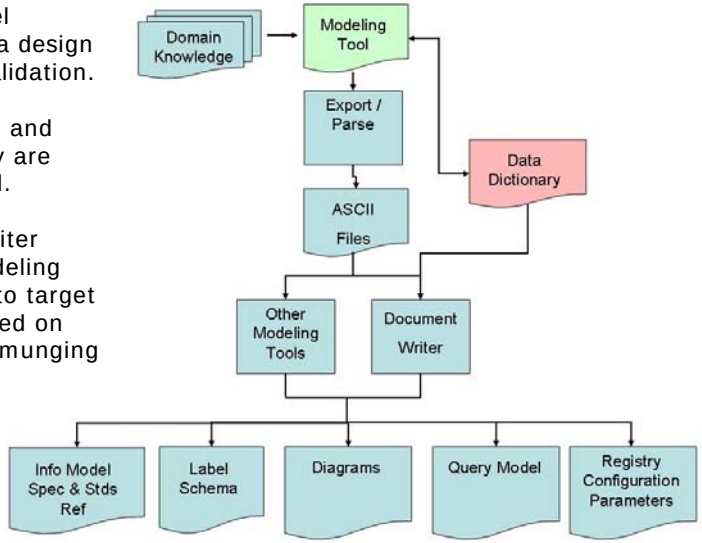
- **Identifiable**
  - Digital Product
    - Data Product
      - Product_Image_Grayscale
      - …Image_3D
      - …Movie
      - …Table_Character
      - …Table_Binary
      - …Table_Binary_Grouped
    - Document Set
    - Software_Set
  - *Non-Digital Product*
    - *Mission*
    - *Instrument*
    - *Resource*

- **Data Type**
  - Binary Data Type
    - Decimal Integer
    - …
  - Character Literals
    - Character Integer
    - …

- **Draft PDS Data Dictionary Model**
- **Grammar – ODL+, PVL, and XML Labels**

The Model Design Process

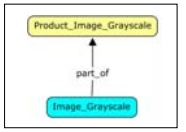
- Master model constrains data design and defines validation.
- Master model and data dictionary are tightly coupled.
- Document writer translates modeling information into target languages based on grammar and munging rules.
- Updates to master model are reflected quickly in the documents.



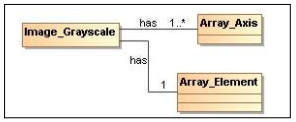
Example Results

Image _Grayscale

Concept Map



UML Class Diagram



XML Schema

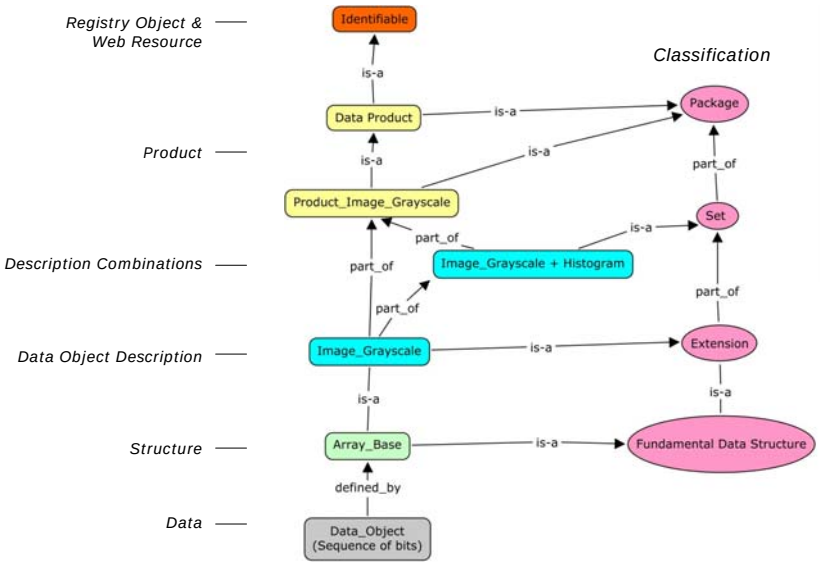
```
<!-- PDS4 XML/Schema for Product_Image_Grayscale -->
<xs:complexType name="Image_Grayscale_Type">
  <xs:sequence>
    <xs:element name="axes_order" type="axes_order_Type"
```

PVL Label Template

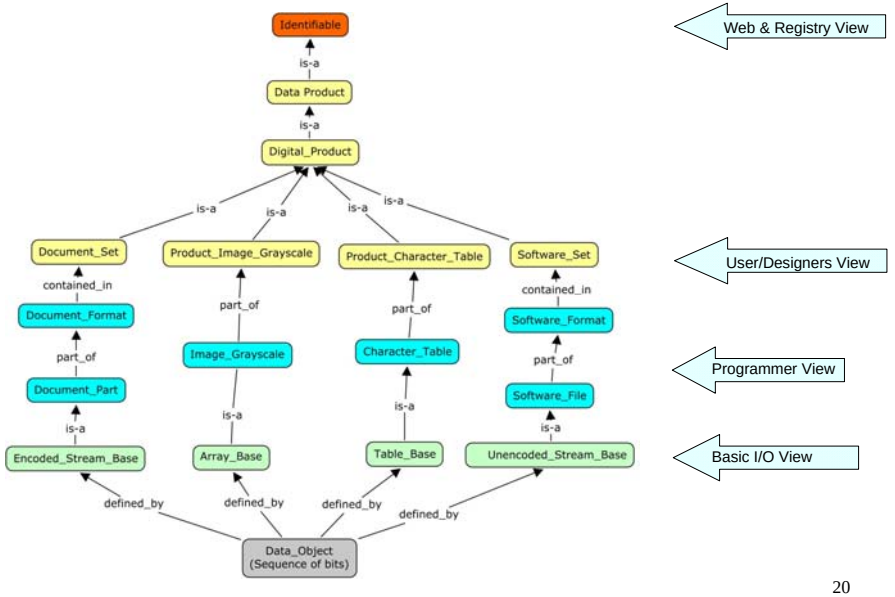
```
/* ******* Label Template - Product_Image_Grayscale
Object = Product_Image_Grayscale;
    Object = Image_Grayscale;
      local_identifier     = ${local_identifier};
```

Class Definition Table

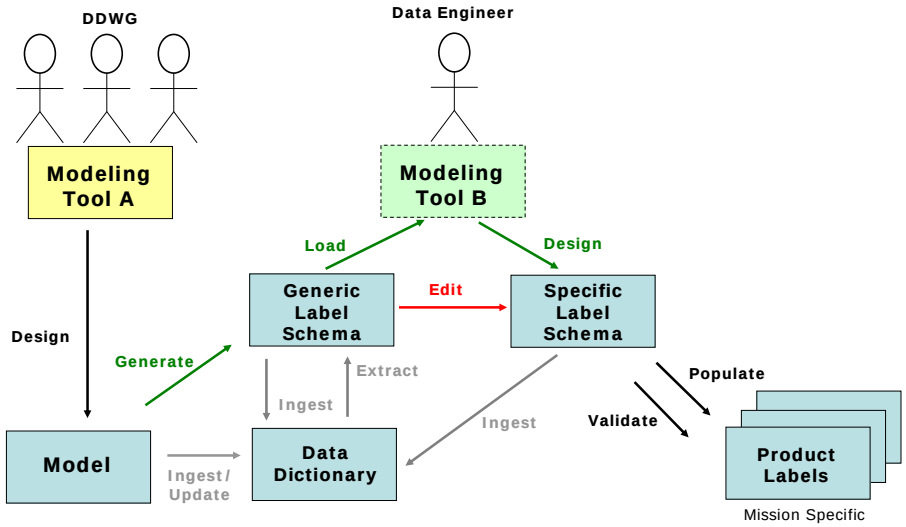
PDS4 Data Product Model Components



PDS4 Data Product Concept Map



PDS4 Product Label Creation



Current using editor
Current using Oxygen/Future Design Tool
Proposed

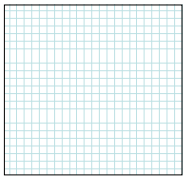
Issue - Ambiguity

- **A PDS3 Image object only requires the number of lines and line samples in a simple image.**
  - However the 2 dimensional structure becomes a 3 dimensional structure with the addition of the BANDS keyword.
  - A two dimensional structure is assumed by the omission of the BANDS keyword.

Image Description

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

Simple Image

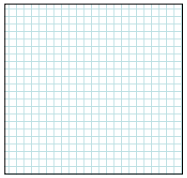


Problem
Data Structure is not Rigorously Defined

- **Current Image Description**
  - Where is the first logical pixel?
  - Are the pixels in row or column major order?
  - How many axes exist?

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

Simple Image



Solution
Rigorously Define Data Structures

- **Array_Base Structure**
  - Where is the first logical pixel? **TOPLEFT**
  - Are the pixels in row or column major order? **FAST2SLOW**
  - How many axes exist? **<#axes>**

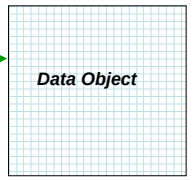
- **Ambiguity cleared up.**

Abstract Data Object Description

```
Object              = Array_Base
    Data_Location   = (filename, 255);
    first_element   = TOPLEFT
    min_index       = 0
    number_of_axes  = #axes
    axes_order      = FAST2SLOW
    byte_order      = MSBF
    element_bytes   = 1
    element_type    = DECIMAL_INTEGER
    axis_length     = (#first, #second, …)
    axis_name       = (first, second, …)
End_Object          = Array_Base
```

Defines

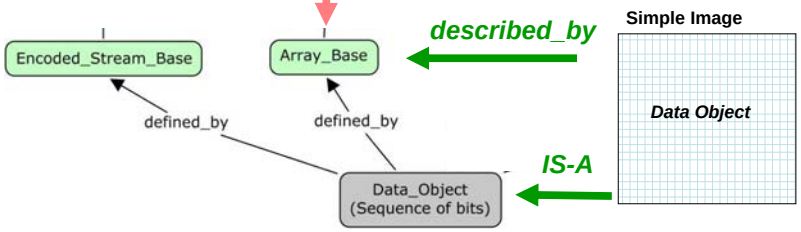
Simple Image



Extend a Data Structure

Data Object Description

```
Object              = Image_Grayscale;
    Data_Location   = (filename, 255);
    first_element   = TOPLEFT;
    min_index       = 0;
    number_of_axes  = 2;
    axes_order      = FAST2SLOW;
    byte_order      = MSBF;
    element_bytes   = 1;
    element_type    = DECIMAL_INTEGER;
    axis_length     = (800, 800);
    axis_name       = (ROW, COLUMN);
End_Object          = Image_Grayscale;
```



Combine Data Structures - Sets

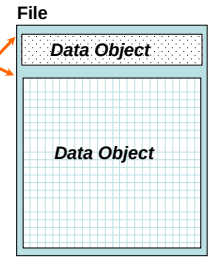
Data Object Descriptions

```
Object = Tagged_Image_Grayscale_Set;

  Object            = Image_Grayscale;
    Data_Location   = (filename, 255);
    first_element   = TOPLEFT;
    min_index       = 0;
    number_of_axes  = 2;
    axes_order      = FAST2SLOW;
    byte_order      = MSBF;
    element_bytes   = 1;
    element_type    = DECIMAL_INTEGER;
    axis_length     = (800, 800);
    axis_name       = (ROW, COLUMN);
  End_Object        = Image_Grayscale;

  Object = Header;
    Data_Location   = (filename, 0);
    comment         = "This Header…";
    bytes           = 255;
  End_Object = Header;

End_Object = Tagged_Image_Grayscale_Set;
```



Package a Product

Product

```
Object = Product_Image_Grayscale
```

Identification Metadata

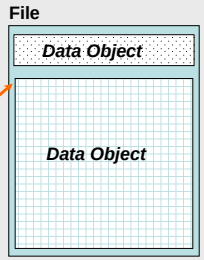
```
Object = Identification_Section;
  Identifier            = "PDS4_IMG_IMAGE_...
  URN                   = "http://URN:MPFL-M-IMP-2...
  Title                 = "MARS PATHFINDER LANDER ...
  Version               = "1.0";
  Label_Revision_Note   = "20090101:1.0 – initial ...
  DD_Version_Id         = "DD_Version_Id";
  PDS_Version_Id        = "PDS4.0";
  Product_Creation_Time = 1998-07-14T00:36:08.000;
End_Object = Identification_Section;
```

Descriptive Metadata

```
Object = Circumstances_Of_Observation_Section;
  Spacecraft_Clock_Start_Count = "1246943630";
    ...
```

Structural Metadata

```
Object = Tagged_Image_Grayscale_Set;
  Object            = Image_Grayscale;
    Data_Location   = (filename, 255);
    first_element   = TOPLEFT;
    ...
```



Query Model

- A formal model consisting of classes, attributes, and relations that are appropriate for use as search constraints
  - Types of query models include
    - Data Set
    - Product
      - Data Product
      - Document Product
      - Software Product
    - Any PDS4 Identifiable
- The query models are subsets of the archive model
  - Can be augmented with external metadata (e.g. LDD)
- Example query constraints
  - general parameters - time, target, mission, instrument host, instrument
  - any metadata defined in the archive model
  - any association between two classes
    - e.g. documents associated with data sets
  - class type and hierarchy
  - geometry - (lat, lon), (az, el), (ra, dec), (x,y,z).

Benefits of the PDS4 Data Model

- **The data model is managed in a data modeling tool.**
  - The model is formally defined.
  - The model can be validated and tested.
- **Defines a few simple fundamental data structures.**
  - Fundamental data structures may be extended and combined to form more complex data formats
- **The overall architecture is model driven.**
  - Disentangles the model from its implementation.
  - Model can evolve over time as research domain changes.
  - Drives the generation of documentation, label schema, and other model dependent artifacts.
- **The data dictionary uses a standard data dictionary model.**

Questions and Answers