PDS4 Data Dictionary
Tutorial

October 28, 2011

Introduction and Purpose

The Planetary Data System (PDS) PDS4 Data Dictionary (DD) is an adjunct to the PDS4 Information Model (IM); together they define the organization and components of PDS4 product labels. The components of a product label are description objects created from classes and their attributes. This document provides an overview of the Data Dictionary, its management, and its use.

The PDS4 Data Dictionary is actually a data base; it is distributed to users in two versions — abridged and unabridged. The unabridged version includes an entry for every attribute in each class it appears. Since many attributes are used with several classes and the meaning often doesn't change, there is considerable repetition. The abridged version has been abstracted from the unabridged version; it contains full definitions but fewer details and none of the repetition. 'Housekeeping' information is suppressed in both versions.

Related Documents

• Controlling Documents

- PDS4 Information Model Specification — Version 0.5.0.0.g (19 October 2011 20:31:53 PDT) — The source for PDS4 class, attribute, and data type definitions.

- ISO/IEC 11179:3 Registry Metamodel and Basic Attributes Specification, 2003 - The reference schema for the PDS4 data dictionary.

• Reference Documents

- Glossary of PDS4 Terms - The source for terms used across the Planetary Data System in its version 4 (PDS4).

- PDS3 Planetary Science Data Dictionary - The online version of the PDS3 data dictionary was used as the source for a few data entries carried over to the PDS4 system.

Terminology

The following are definitions of some important terms used in the Data Dictionary; these are taken verbatim from the PDS4 Glossary.

1. **attribute**: A property or characteristic that provides a unit of information. For example, 'color' and 'length' are possible attributes.

2. **class**: The set of attributes (including a name) which defines a family. A class is generic — a template from which individual members of the family may be constructed. If a class named 'rope' is defined by

attributes 'color' and 'length', we can construct a family of ropes — e.g., red and 3 m long, red and 4 m long, blue and 2 m long, ...

3. **association:** An attribute that establishes a unidirectional relationship between two classes. For example, a table has records; 'has record' is the relationship between one entity (Table_Base, the simplest table in PDS4 nomenclature) and another (Table_Record).

4. **object:** The realization of a single member of a family defined by a class. If the class 'rope' has attributes 'color' and 'length', we can construct a 'rope' family with three members — red and 3 m long, red and 4 m long, and blue and 2 m long. Each member is an object.

5. **conceptual object:** An object which is intangible (and, because it is intangible, does not appear in a digital archive). Examples of 'conceptual objects' include the Cassini mission and NASA's strategic plan for solar system exploration. Note that a PDF describing the Cassini mission is a digital object, not a conceptual object (nor a component of a conceptual object).

6. **digital object:** An object which is real data — for example, a binary image of a redwood tree or an ASCII table of atmospheric composition versus altitude.

7. **physical object:** An object which is physical or tangible (and, therefore, does not itself appear in a digital archive). Examples of 'physical objects' include the planet Saturn and the Venus Express magnetometer. Note that an ASCII file describing Saturn is a digital object, not a physical object (nor a component of a physical object).

8. **resource:** The target (referent) of any Uniform Resource Identifier; the thing to which a URI points.

Note that the term **data element** is sometimes used as a synonym for **attribute** (or **class** or both). In PDS3 it was used frequently as a synonym for keyword; it is not used in this tutorial. The term **keyword** is avoided since **keyword** is more closely associated with search terms, as in publication keywords. The term **object**, a synonym for **class** in the PDS3 data model, is not used in that sense here; an **object** is *created from* a **class**.

PDS4 Data Dictionary Structure

Meta-attributes (attributes of attributes) are used to define attributes. For example, the attribute `axes` is defined using the meta-attributes **name** and **definition** (among others). **name** provides a common name for the attribute ("axes") and **definition** provides a statement that describes the attribute ("a count of the axes").

A subset of the ISO/IEC 11179 Metadata Registry reference model was chosen for the PDS4 Data Dictionary meta-attributes and structure. This standard ensures data system stability and interoperability.

The PDS4 Data Dictionary focuses on attributes and, to a lesser extent, on classes. A model of an attribute definition is shown in Figure 1. A PDS4 attribute is defined by the structure within the dashed line: the four orange boxes labeled "Attributes" (which contain *meta*-attributes) and the four blue boxes (containing decimal numbers) which represent classes. Definitions of the meta-attributes are given below. The **Bibliographic_Reference** class provides optional background information for the attribute definition, and the structure *outside* the dashed line is important for DD management; neither will be discussed further here.

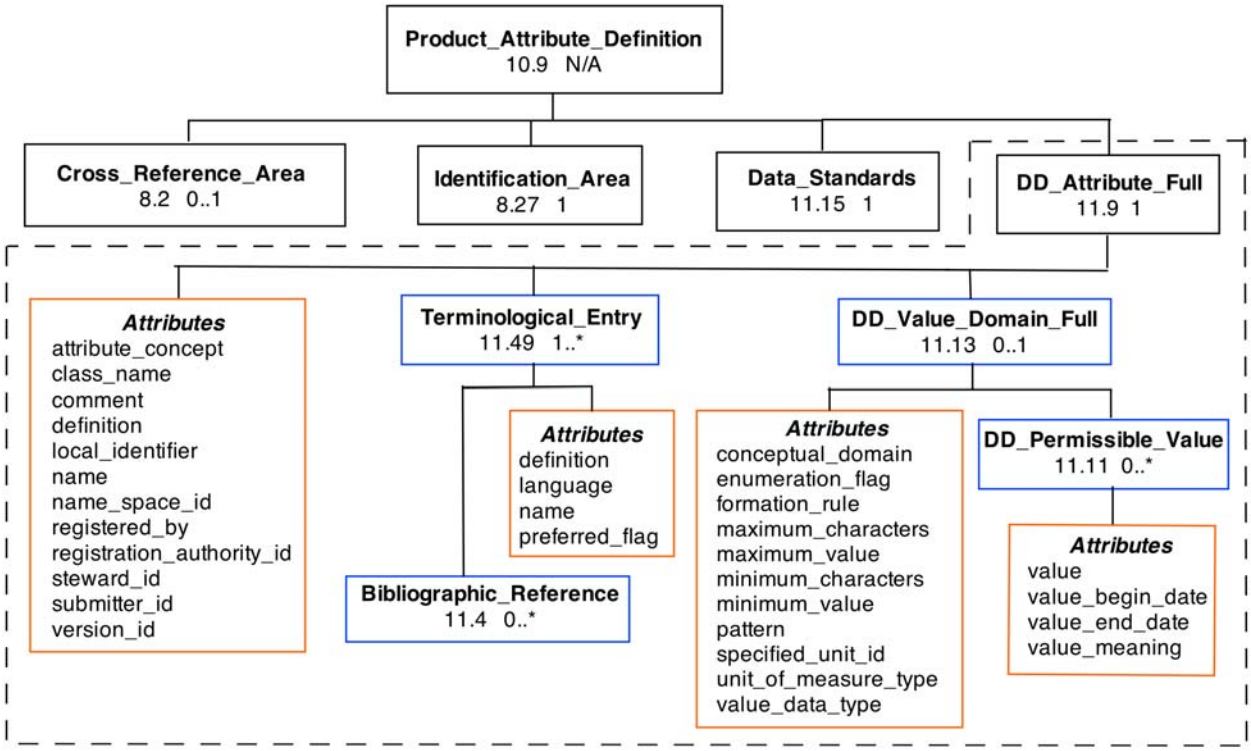

Figure 1. Model of a PDS4 Data Dictionary attribute definition. Boxes with decimal numbers represent classes; decimal numbers are the relevant section numbers in the PDS4 Information Model (version 0.5.0.0.g), and accompanying numbers show cardinality. Structure outside the dashed line is used for DD management.

Figures 2, 3, and 4 show examples of how attributes and their meta-attributes are presented through the PDS4 Data Dictionaries. Note that only some of the meta-attributes are displayed in the DD. Where a meta-attribute is shown as "required" but it does not appear in an example (Figures 2, 3, or 4), that is because the Data Dictionary does not display the attribute — often because the attribute has the same value throughout the Dictionary (for example 17 and 18 below). Note also that several terms are used as both meta-attributes here and attributes elsewhere (*e.g.*, **name** and **comment**); their definitions are allowed to differ slightly depending on the context. The definitions here apply only to meta-attribute usage; see the Data Dictionary itself for definitions in the attribute context. Also, even though the ISO/IEC 11179 specification allows the inclusion of class definitions, this aspect of the model is not currently used; the PDS4 class definitions in the Data Dictionary are simply copied verbatim from the PDS4 Information Model.

- axes
  - steward_id: pds
  - name_space_id: pds
  - class_name: Array_2D
  - version_id: 0.5.0.0.g
  - attribute_concept: COUNT
  - definition: The axes attribute provides a count of the axes.
  - conceptual_domain: INTEGER
  - value_data_type: ASCII_Integer
  - minimum_value: 1
  - maximum_value: 16
  - permissible_value: 2

Figure 2 - PDS4 Unabridged Data Dictionary entry for the attribute 'axes' as used with class Array_2D

• sample_display_direction
steward_id: pds
name_space_id: pds
class_name: Image_2D_Display
version_id: 0.5.0.0.g
definition: The sample_display_direction attribute is the preferred orientation of samples within a line for viewing on a display device. The default is right, meaning samples are viewed from left to right on the display. "sample_display_direction" must be used with "line_display_direction". Image rotation attributes such as TWIST_ANGLE, CELESTIAL_NORTH_CLOCK_ANGLE, and BODY_POLE_CLOCK_ANGLE are defined under the assumption that the image is displayed in its preferred orientation.
value_data_type: ASCII_Short_String_Collapsed
minimum_characters: 1
maximum_characters: 255
permissible_values:
RIGHT
DOW
LEFT
UP

Figure 3 - PDS4 Abridged Data Dictionary entry for the attribute `'sample_display_direction'`

• exposure_duration
steward_id: img
name_space_id: img
version_id: 0.3.1.1.Beta
definition: This element provides the value of the time interval between the opening and closing of an instrument aperture (such as a camera shutter).
value_data_type: ASCII_Real
minimum_value: 0.0
maximum_value: INF
unit_of_measure_type: UnitOfMeasure_Time
specified_unit_id: s

Figure 4 - PDS4 Abridged Data Dictionary entry for the attribute `'exposure_duration'`

Meta-attributes are defined below.

1. The meta-attribute **attribute_concept** (required; supplied by the steward) provides the type of information (a classification) conveyed by the attribute. For example, `axes` has **attribute_concept** set to COUNT (Figure 2).

2. The meta-attribute **class_name** (required) provides the common name by which the class is identified; it is the class within which the attribute is used. Specification of the class-attribute pair is a unique entry within a data dictionary. **class_name** is set to Array_2D in Figure 2 and Image_2D_Display in Figure 3. Since the Abridged Data Dictionary collapses many class-attribute definitions into one, no value for **class_name** is shown in Figure 4.

3. The meta-attribute **comment** (optional) is a character string expressing one or more remarks or thoughts relevant to the attribute. No **comment** is shown in Figures 2, 3, or 4.

4. The meta-attribute **conceptual_domain** (optional; supplied by the steward) provides the domain to which the value has been assigned. In Figure 2 **conceptual_domain** is set to INTEGER; no value is

shown for the Abridged DD entries in Figures 3 and 4.

5. The meta-attribute **definition** (required) provides a statement, picture in words, or account that defines the attribute. The **definition** used in the Data Dictionary is selected from one or more **definition** choices in **Terminological_Entry** using **preferred_flag.**

6. The meta-attribute **enumeration_flag** (optional) indicates whether there is an enumerated set of permissible values. If not specified, the default value 'F' (false) may be assumed. **enumeration_flag** is set to 'T' (true) for the examples in Figures 2 and 3; but it is not shown explicitly in the DD listings because presence of **DD_Permissible_Value** implies the value 'T' (see #16 below).

7. The meta-attribute **formation_rule** (optional) provides a user-friendly instruction for forming values of the attribute; **pattern** is a symbolic instruction, which is not intended to be user-friendly. **formation_rule** is not used in Figures 2, 3, and 4.

8. The meta-attribute **local_identifier** (required) is a character string which uniquely identifies DD_Attribute_Full within its label. In a complex label, **local_identifier** provides a shorthand term that can be used when making internal references within the label. No **local_identifier** is shown in Figures 2, 3, or 4.

9. The meta-attribute **maximum_characters** (optional) provides the upper, inclusive bound on the number of characters. **maximum_characters** is set to 255 in Figure 3, although 5 would have been sufficient for the **DD_Permissible_Values** listed.

10. The meta-attribute **maximum_value** (optional) provides the upper, inclusive bound on the value. **maximum_value** is set to 'INF' (infinity) in Figure 4.

11. The meta-attribute **minimum_characters** (optional) provides the lower, inclusive bound on the number of characters. **minimum_characters** is set to 1 in Figure 3.

12. The meta-attribute **minimum_value** (optional) provides the lower, inclusive bound on the value. **minimum_value** is set to 0.0 in Figure 4.

13. The meta-attribute **name** (required) provides a word or combination of words by which the attribute is known. The **name** used in the Data Dictionary is selected from one or more **name** choices in **Terminological_Entry** using **preferred_flag. name** is implicit in the first line of Figures 2, 3, and 4.

14. The meta-attribute **name_space_id** (required) is the abbreviation of the XML schema namespace container (see below) for this logical grouping of classes and attributes. The **name_space_id** is assigned by the steward (see below); the PDS steward has set **name_space_id** to 'pds' in Figures 2 and 3, and the IMG steward has set it to 'img' in Figure 4.

15. The meta-attribute **pattern** (optional) provides a symbolic instruction for forming values. **formation_rule** is a user-friendly instruction, which is intended to convey the same information in words that can be readily understood by a reader/user. **pattern** is not used in Figures 2, 3, and 4.

16. DD_**Permissible_Value** is a class, optionally used within attribute definitions; so it is not a meta-attribute *per se*, but its own attributes perform the same function. If **DD_Permissible_Value** appears, it is accompanied by the attribute **enumeration_flag** set to 'T'. If multiple values are allowed, a separate **DD_Permissible_Value** definition is required for each. The meta-attributes under

DD_Permissible_Value are:

16A. **value** (required for each **DD_Permissible_Value**) is a single, allowed numerical or character string value. The only permissible value for `axes` when used with Array_2D is 2 (Figure 2). In Figure 3 there are four possible values for `sample_display_direction`: "RIGHT", "DOWN", "LEFT", and "UP". Note that **enumeration_flag** is 'T' in Figures 2 and 3. In Figure 4 there is no enumeration; an infinite number of values is possible.

16B. **value_begin_date** (optional) provides the first date on which the permissible **value** is in effect. **value_begin_date** is not used in Figures 2, 3, and 4.

16C. **value_end_date** (optional) provides the last date on which the permissible **value** is in effect. **value_end_date** is not used in Figures 2, 3, and 4.

16D. **value_meaning** (required for each **DD_Permissible_Value**) is the meaning or semantic content of the associated permissible **value**. Currently there are no examples of **value_meaning** in the Data Dictionary; however, the "PDS4 Data Type Definitions" chapter in the DD achieves the same result (but only for **data_type** values). **value_meaning** is not used in Figures 2, 3, and 4.

17. The meta-attribute **registered_by** (required) provides the name of the person or organization that registered the attribute. **registered_by** is not used in Figures 2, 3, and 4.

18. The meta-attribute **registration_authority_id** (required) is the name of the organization that registered the attribute definition. For all PDS attributes, **registration_authority_id** will be set to '0001_NASA_PDS_1'. **registration_authority_id** is not shown in Figures 2, 3, and 4.

19. The meta-attribute **specified_unit_id** (optional) provides the units in which other meta-attributes such as **maximum_value** and **DD_Permissible_Value** are given. **specified_unit_id** is 's' in Figure 4.

20. The meta-attribute **steward_id** (required) is the person or organization who manages the set of registered attributes and classes. `Sample_display_direction` and `axes` have both been assigned to 'pds' (Figures 2 and 3); `exposure_duration` has been assigned to 'img' (Figure 4).

21. The meta-attribute **submitter_id** (required) is the name of the author who submits the attribute definition to the steward. **submitter_id** is not used in Figures 2, 3, or 4.

22. **Terminological_Entry** is a required class; there may be multiple instances of **Terminological_Entry** within a single **DD_Attribute_Full**. Meta-attributes under **Terminological_Entry** are:

22A. **definition** (required) provides a statement, picture in words, or account that defines the term. **definition** of `axes` is "a count of the axes" (Figure 2).

22B. **language** (required) indicates the language used for the **name** and **definition** of the term. **language** is not shown in Figures 2, 3, and 4.

22C. **name** (required) provides a name for the term in the **language** selected.

22D. **preferred_flag** (required) indicates whether this entry is preferred over all other entries — for example, because of a **language** choice.

Terminological_Entry is not shown explicitly in Figures 2, 3 ,or 4.

23. The meta-attribute **unit_of_measure_type** (optional) provides the named grouping of units to be used for this attribute. **unit_of_measure_type** is set to UnitOfMeasure_Time in Figure 4.

24. The meta-attribute **value_data_type** (required) provides the data type used to represent the value. The **value_data_type** for `axes` is ASCII_Integer (Figure 2), meaning that values like -255, 0, and 7 are acceptable so far as storage is concerned (that -255 is not a meaningful count of axes is addressed using **DD_Permissible_Value**). The **value_data_type** for `sample_display_direction` is ASCII_Short_String_Collapsed (Figure 3), meaning that special characters such as 'line-feed' and 'tab' are replaced by spaces and then each multi-space sub-string is reduced to a single space. **value_data_type** is an enumerated attribute; acceptable values can be found in the PDS4 Data Dictionary.

25. The meta-attribute **version_id** (required; supplied by steward) identifies the version of the attribute's definition. **version_id** is set to 0.5.0.0.g in Figures 2 and 3 and is set to 0.3.1.1.Beta in Figure 4.

Management of Attributes and Classes

Management of PDS4 includes assignment of responsibility for maintenance of each attribute and class in the Data Dictionary. The ISO/IEC 11179 reference model provides two attributes for this purpose: 'registration authority' and 'steward'. 'Namespace' also plays a role but only for the XML implementation.

A **registration authority** is an organization responsible for maintaining a registry. The ISO/IEC 11179 reference model allows many registration authorities, each of which is uniquely identified. Each registration authority has, by definition, its own model and therefore, implicitly, its own local dictionary. Each registration authority can design, develop, and manage its own model and dictionary using any data modeling methodology and independently of the other Registration Authorities.

The registration authority for the Planetary Data System is '0001_NASA_PDS_1'. PDS has designed, developed, and managed its model using the 'object oriented' methodology. An important constraint levied by the model is that each attribute and class must have a unique name within the model. But PDS attributes and classes may duplicate those maintained by other registration authorities.

A **steward** is a person or organization who manages a set of registered attributes and classes, typically as an agent of another or others. Each attribute and class in the PDS4 Data Dictionary is assigned to a single steward. Stewards for PDS4 include PDS, the discipline nodes, and any mission wishing to conform to the PDS4 Information Model. A registration authority must have at least one steward; but it may have many. Stewards are uniquely identified within a registration authority. A single steward may operate across many registration authorities.

Namespace is an abstract container or environment created to hold a logical grouping of unique identifiers or symbols (*i.e.*, names). An identifier defined in a namespace is associated with that namespace. The same identifier may be independently defined in multiple namespaces. Namespaces are not a functional component of the PDS4 Information Model or Data Dictionary; rather, they are assigned and used for implementation into XML Schema. A steward may ask for a namespace from the PDS Namespace Registry Service, which will make the assignment if the namespace has not previously been assigned.

Acknowledgements

The PDS4 Data Dictionary and the PDS4 Information Model are joint efforts involving representatives from each of the PDS nodes functioning as the PDS4 Data Design Working Group.

J. S. Hughes
R. A. Simpson