

PDS4 Data Dictionary

Tutorial

October 19, 2010

Edits by Simpson (2010-11-30), Hughes (2010-12-10), Simpson (2010-12-12)

Introduction and Purpose

The Planetary Data System (PDS) PDS4 Data Dictionary (DD) is an adjunct to the PDS4 Information Model; it defines 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.

Audience

This tutorial is addressed to people learning about PDS4 and its components.

Scope

The tutorial provides an introduction to the PDS4 Data Dictionary.

Related Documents

- Controlling Documents
 - PDS4 Information Model Specification - 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
 - 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 from the current system.
 - PDS4 Data Dictionary – Unabridged – V.0.1.1.1.c

Terminology

Following are definitions of some important terms used in the data dictionary.

1. An **attribute** is a property or characteristic that provides a unit of information about a class. For example, 'color=blue' defines an attribute of paper, fabric, or sky.
2. A **class** is the set of attributes which identifies a family. A class is generic -- a template from which individual members of the family may be constructed. In the example above paper, fabric, and sky would be classes.
3. An **association** is an attribute that establishes a relationship between two classes. For example the association 'has_Array_Axis' is used to relate the PDS4 parent class 'image_grayscale' to the child class, Array_Axis.
4. An **object** is constructed from a class. It is a specific instance of a class. In PDS4, a data object can be one of three types, digital, conceptual, or physical.
5. A **conceptual object** is an object that is intangible. For example, 'mission' is a conceptual object.
6. A **digital object** is an object consisting of digital data. For example, a binary image is a digital object.
7. A **physical object** is an object that is tangible. For example, the Lunar Orbiting Laser Altimeter is a physical object.
8. A **resource** is the referent (target) of a Uniform Resource Identifier; the thing to which a URI points.
9. The prefix **local** indicates a scope no larger than the enclosing context in which the value or expression has been defined.
10. The prefix **logical**, as used in the definition of logical identifier, indicates that the identifier logically groups a set of objects.

Note that the term **data element** is often used as a synonym for **attribute** (or **class** or both); here it is used only as a synonym for **attribute**. The term **keyword** should be 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 'title' and 'description'. 'Title' provides a common name for the attribute ("axes") and 'description' provides a statement that describes the attribute ("a count of the axes").

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

The PDS4 data dictionary focuses on data elements. Each is defined as illustrated in Figures 1 and 2. Following the figures, each ISO/IEC 11179 *meta-attribute* is defined. Note that, several terms are

used as both attributes and meta-attributes; their definitions may differ slightly depending on the context. The definitions here are only for the meta-attribute usage. Also, even though the ISO/IEC 11179 specification allows the inclusion of class definitions, this aspect of the model is not currently used and the PDS4 class definitions are simply included verbatim from the PDS4 Information Model.

- axes

steward: pds
name space id: pds:
version: 0.2.0.0.d

description: The axes attribute provides a count of the axes.
data_type: ASCII_Integer
minimum_value: 1
maximum_value: 16

Figure 1 - Definition of the attribute 'axes'

- sample_display_direction

steward: pds
name space id: pds:
version: 0.2.0.0.d

description: 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.

data_type: ASCII_Short_String_Collapsed - Enumerated
minimum_characters: 1
maximum_characters: 6
permissible values
DOWN
LEFT
RIGHT
UP

• Figure 2 - Definition of the attribute 'sample_display_direction'

- exposure_duration

steward: img
 name space id: img:
 version: 0.2.0.0.d

description: This element provides the value of the time interval between the opening and closing of an instrument aperture (such as a camera shutter).

data_type: ASCII_Real
 minimum_value: 0.0
 maximum_value: INF
 unit_of_measure_name: UnitOfMeasure_Time
 default_unit_id: s

• Figure 3 - Definition of the attribute 'exposure_duration'

1. The meta-attribute **class** provides the context within which the attribute is to be defined.
2. The meta-attribute **data_type** provides the hardware representation used to store a value in a physical file. The data type for `axes` is `ASCII_Integer`, meaning that values like -255, 0, and 7 are acceptable so far as storage is concerned (the fact that -255 is not a meaningful count of axes is addressed below). The data type for `sample_display_direction` is `ASCII_Short_String_Collapsed`, meaning that special characters such as 'line-feed' and 'tab' are replaced by spaces and then all multi-space sub-strings are reduced to a single space . The suffix "enumerated" is appended to the data_type as a reminder that only a fixed set of values for `sample_display_direction` is allowed. The current list of permissible values for data type is shown below; for definitions, see the PDS4 Data Dictionary.

ASCII_AnyURI	ASCII_Integer
ASCII_Boolean_TF	ASCII_Integer_Binary
ASCII_DOI	ASCII_Integer_Hex
ASCII_Date_DOY	ASCII_LID
ASCII+Date_Time	ASCII_LIDVID
ASCII_Date_Time_DOY	ASCII_MD5_Checksum
ASCII_Date_Time_UTC	ASCII_NonNegative_Integer
ASCII_Date_Time_YMD	ASCII_Real
ASCII_Date_YMD	ASCII_Short_String_Collapsed
ASCII_Directory_Path_Name	ASCII_Short_String_Preserved
ASCII_File_Name	ASCII_Text_Preserved
ASCII_File_Specification_Name	ASCII_Time
ASCII_Identifier	ASCII_VID

Bit	SignedMSB8
ComplexB16	UTF8_Short_String_Collapsed
ComplexB8	UTF8_Short_String_Preserved
IEEE754Double	UTF8_Text_Preserved
IEEE754Single	UnsignedByte
SignedLSB2	UnsignedLSB2
SignedLSB4	UnsignedLSB4
SignedLSB8	UnsignedMSB2
SignedMSB2	UnsignedMSB4
SignedMSB4	

3. The meta-attribute **description** provides a statement, picture in words, or account that describes the meta-attribute. For example, the description for axes explains that it is the count of axes in an array.
4. The meta-attribute **default_unit_id** is a unit identifier chosen as a default from among the possible unit identifiers specified by unit_of_measure_name.
5. The meta-attribute **maximum_characters** provides the upper inclusive bound on the number of characters.
6. The meta-attribute **maximum_value** provides the upper inclusive bound of a value.
7. The meta-attribute **minimum_characters** provides the lower inclusive bound on the number of characters.
8. The meta-attribute **minimum_value** provides the lower inclusive bound of a value.
9. The meta-attribute **steward** indicates the person or organization who manages the set of registered attributes and classes to which axes or sample_display_direction has been assigned. Axes and sample_display_direction have both been assigned to 'pds'.
10. The meta-attribute **name_space_id** identifies the XML Schema namespace container (see below) for a logical grouping of data elements. The name_space_id is assigned by the steward, who has set it to 'pds' in both Figure 1 and Figure 2.
11. The meta-attribute **permissible_value** is an expression which gives the allowed value(s). The permissible values for axes are set by data_type, maximum_value, and minimum_value; so no explicit entry for permissible_value is shown (Figure 1). Four values are allowed for sample_display_direction: "DOWN", "UP", "LEFT", and "RIGHT" (Figure 2).
12. The meta-attribute **title** is a name by which the data element is formally known. **Title** is implicit on the first line of each figure.
13. The meta-attribute **unit_of_measurement_name** indicates the division of quantity accepted as a standard of measurement.
14. The meta-attribute **value meaning** is the meaning or semantic content of a permissible value. Currently there is no example of a value meaning.
15. The meta-attribute **version** identifies the specific version of an attribute's definition.

Management of Attributes and Classes

Management of PDS4 includes assigning 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 register. 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 'PDS'. 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 and be assigned one or more namespaces. Namespaces are assigned via the Namespace Registry Service, which functions principally to ensure that namespaces are unique across all registration authorities intending on archiving with the PDS.

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.