**PDS4 Data Dictionary**

**Tutorial**

October 19, 2010

**Introduction**

The Planetary Data System (PDS) PDS4 Data Dictionary defines the organization and components of the PDS4 product labels. The components of a product label are objects created from classes and their attributes.

**Audience**

The PDS4 Data Dictionary is intended for use by programmers and data engineers who require formal definitions of the components and the organization of PDS4 products.

**Acknowledgements**

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

**Scope**

The PDS4 Data Dictionary defines the common and discipline level classes and attributes used to create PDS4 product labels. It also defines the meta-attributes (i.e. attributes about attributes) used to define attributes.

**Related Documents**

* Controlling Documents
	+ PDS4 Information Model Specification - The PDS4 Information Model Specification document is the source for class, attribute, and data type definitions.
	+ ISO/IEC 11179:3 Registry Metamodel and Basic Attributes Specification, 2003 - The ISO/IEC 11179 specification provides the reference schema for the PDS4 data dictionary.
* Reference Document
	+ Planetary Science Data Dictionary - The online version of the PDS3 data dictionary was used as the source for a few data elements carried over from the PDS3 data standards.

**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.
2. A **class** is the set of attributes which identifies a family. A class is generic -- a template from which individual members of each family may be constructed.
3. A **data 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.
4. A **conceptual object** is an object that is not tangible. For example, a mission is a conceptual object.
5. A **digital object** is an object consisting of digital information. For example, an image is a digital object.
6. A **physical object** is an object that is tangible. For example, a spacecraft instrument is a physical object.
7. A **descriptive class** provides information that is useful for the interpretation of a data object.
8. A **structural class** provides information that defines the components and organization of a digital object.
9. A **resource** is the referent of any Uniform Resource Identifier. The concept of resource is primitive in the Web architecture and is used in the definition of its fundamental elements.
10. The term **formal** as used in the definition of attributes that are names indicates that an established procedure was involved in creating the name.
11. The term **local** indicates a local scope where scope is an enclosing context where values and expressions are associated.
12. The term **logical** as used in the definition of logical identifier indicates that the identifier logically groups a set of objects.

**Attributes and Classes in the Context of the PDS4 Information Model**

The PDS4 Information Model consists of a formal definition of the attributes and classes required for the PDS archive and supporting systems. An attribute is defined using a standard set of meta-attributes. A class is defined using one or more attributes. An alternate way of saying this is that a class contains all objects that have the same set of attributes.

For example, in the PDS4 Information Model the attributes axis\_order and axes are used in the definition of the image\_grayscale class. The axes attribute provides the number of axes and the axis\_order attribute indicates how the axis elements are physically serialized with respect to the axes.

Attributes also add meaning to classes by defining relationships between classes. For example the attribute has\_Array\_Axis is used to relate the parent class image\_grayscale to the child class, Axis.

**Attributes and Classes in the Context of the PDS4 Data Dictionary**

Typically a data dictionary, also called a controlled vocabulary, is an index of terms used in a data system. In the PDS4 Data Dictionary the index consists of the attributes and their definitions that are defined in the PDS4 Information Model. The classes defined in the PDS4 Information Model are also included in the PDS4 Data Dictionary.

Note that the terms *data element* and *element* are often used as synonyms of *attribute*. However the term *keyword* should be avoided since *keyword* is more closely associated with search terms, as in publication keywords. The term *object* as used in the PDS3 data model for a synonym of *class* is not used. An *object*  is created from a class.

**PDS4 Data Dictionary Structure**

To define an attribute, meta-attributes are used. For example, the attribute axes is defined using the meta-attributes title and description. The title provides a common name for the attribute and description provides a statement that describes or gives an account of the attribute.

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 enables data system stability and interoperability and keeps data dictionary developers from having to continually “reinvent the wheel”.

The PDS4 data dictionary schema focuses on data elements. Each is defined as illustrated in Figure 1. Following Figure 1, each ISO/IEC 11179 meta-attribute is defined. Note that even though the ISO/IEC 11179 specification also allows the inclusion of class definitions, this aspect of the model is not currently used and the PDS4 classes definitions are simply included verbatim from the PDS4 Information Model.



Figure - Definition of the attribute Axes

1. The meta-attribute **begin date** provides the lower bound of the time interval within which the defined item is in effect.
2. The meta-attribute **conceptual domain** is a classification of the value domains. For example, the conceptual domain for axes is CD\_INTEGER. For the current list of permissible values for conceptual domain see the PDS4 Data Dictionary.
3. A **data element** is a unit of data for which the definition, identification, representation and permissible values are specified by means of a set of meta-attributes. For example, the concept of axes is used in the PDS archive to provide a count of the axes in an array. The data element aspect of this concept is the named attribute (or data element) axes.
4. The meta-attribute **data element concept** is a classification of the data elements. The data element concept for axes is DEC\_COUNT since all data elements ending in “s” imply count. For the current list of permissible values for data element concept see the PDS4 Data Dictionary.
5. The meta-attribute **data\_type** provides the hardware representation used to store a value in a physical file. For the current list of permissible values for data type see the PDS4 Data Dictionary.
6. The meta-attribute **description** provides a statement, picture in words, or account that describes.
7. The meta-attribute **end date** provides the upper bound of the time interval within which the defined item is in effect.
8. The meta-attribute **maximum\_value** provides the upper inclusive bound of a value.
9. The meta-attribute **maximum\_characters** provides the upper inclusive bound on the number of characters.
10. The meta-attribute **minimum\_value** provides the lower inclusive bound of a value.
11. The meta-attribute **minimum\_characters** provides the lower inclusive bound on the number of characters.
12. The meta-attribute **steward** indicates the person or organization who manages a set of registered attributes and classes.
13. The meta-attribute **name\_space\_id** identifies the XML Schema namespace container for a logical grouping of data elements. The name\_space\_id is assigned by the steward.
14. The meta-attribute **permissible\_value** is an expression of a value meaning allowed in a specific value domain. A permissible value for axes is "2".
15. The meta-attribute **title** is a name by which the data element is formally known.
16. The meta-attribute **value\_domain** is specified by a list of all its permissible values. These can be enumerated or non-enumerated. In PDS4, the value domain for axes is the value range from "1" to "6".
17. The meta-attribute **unit\_of\_measurement** indicates the division of quantity accepted as a standard of measurement.
18. The meta-attribute **version** identifies the specific variation of a data element.
19. The meta-attribute **value meaning** is the meaning or semantic content of a permissible value.

**Governance of Attributes and Classes**

Governance is a method or system of management. For the PDS4 data dictionary governance focuses on who governs each attribute and class in the data dictionary. The ISO/IEC 11179 reference model provides two attributes for this purpose, the registration authority and the steward. Namespaces also have a governance aspect but only for the XML implementation.

**Registration Authority**

A Registration Authority is an organization responsible for maintaining a register. The ISO/IEC 11179 reference model and the PDS4 data dictionary allow many Registration Authorities and each Registration Authority is uniquely identified within the set of Registration Authorities (e.g., PDS, ESA, IPDA). 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, however the results must ultimately be transformable so as to be compliant with the PDS4 Data Dictionary model. The PDS has designed, developed, and manages its model using the object\_oriented methodology. An important constraints levied by the model is that each attribute and class must have a unique name within the model.

The Registration Authority for the Planetary Data System is “PDS”. Again other Registration Authorities are allowed. The model and dictionary owned by each Registration Authority are independent, in both function and content, of those owned other Registration Authorities. This means that duplicate class and attribute names are allowed, as long as they are governed by different Registration Authorities.

**Steward**

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 registered into the PDS4 Data Dictionary is assigned a single steward. Propose stewards for the PDS4 Information Model include the PDS, the discipline nodes, and any mission wishing to conform to the PDS4 Information Model.

A Registration Authority must have at least one but may have many Stewards. Stewards are unique within a Registration Authority but may operate across many Registration Authorities.

**Namespace**

A 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 can be independently defined in multiple namespaces.

Namespaces are not a functional component of the PDS4 Information Model or Data Dictionary. Namespaces are assigned and used for implementation into XML Schema. A Steward may ask for and is 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.

**Types of Objects**

The PDS4 Information Model uses the Open Archive Information System (OAIS) Information Object as a key organizational concept. An Information Object is simply an object and its description. This simple definition is extended to include digital, conceptual, and physical objects and their description. Since little is known about an object apart from its description, the definitions of digital\_object, conceptual\_object, and physical object are trivial. In fact only the digital object has an attribute, a sequence of digital bits representing the binary object in a computer system. Therefore the information model consists almost entirely of descriptive classes. The relationship between descriptive and actual objects and their classes are presented in the following diagram.

 **PDS4 Information Objects – The Classes and their Objects**

Description Object (description) Data Object (thing being described)

Class Object Class Object

Image\_Grayscale Image\_Grayscale Digital\_Object Digital data (image) in file

 (model/schema) (XML Label) (attr:bit\_string)

Instrument Instrument Physical Object Instrument on Spacecraft (attr:none)

Mission Mission Conceptual Object Active Mission

 (attr:none)

**PDS Information Model Version Identifiers**

The version identifier used for the PDS Information Model consists of three digits separated by periods, for example, 1.2.3. The first digit is incremented when there has been a change to a required element such as making a required element optional or restricting an enumerated list. This class of change implies that the model is no longer backwards compatible . The second digit is incremented when there is a change to an optional element, such as augmenting a list of enumerated values or making an optional element required. This class of changes implies that the model continues to be backwards compatible. The final digit is incremented when there has been a change to text or a description.

For the beta release of the Information Model, the three digit version identifier has been prefixed with a zero and suffixed with a lower case letter, for example, 0.1.2.3.b. The zero will either be replaced with a 4 or simply eliminated with the first release of the PDS4 data standards. The lower case letter will be incremented for each release to simplify the identification of releases.