



A Data Model for a Data Element Dictionary

IPDA Data Modeling Task Group

IPDA Steering Committee Meeting
11-12 July 2008
Montreal, Canada

steve.hughes@jpl.nasa.gov





Topics

- **Introduction**
- **Terminology**
- **Problem Statement**
 - Problem 1 - Limited Definitions
 - Problem 2 - Enabling Interoperability
- **Case Study**
 - Questions to ask of a Data Dictionary
- **ISO/IEC 11179**
- **Status**
- **Conclusions**



Introduction



- The Data Dictionary Modeling task group was constituted in the summer of 2007, with mandate to “develop a data model for the data dictionary”.
- This effort is related to the standard data modeling task because a data dictionary is needed to support the data model.
 - The IPDA draft standard data model is currently under assessment.
 - The IPDA standard model, a version of the PDS data model will require a similar data dictionary.



Terminology

- A Data Model defines the entities to be processed, their **attributes**, and the relationships that add meaning.
- An **attribute** has alternate names.
 - Data Element
 - Vocabulary Term
- The set of all **attributes** in a data model is also called its vocabulary and is collected into a **data dictionary**.
- When defining an **attribute**, a set of **meta-attributes** or “attributes about attributes” are used.
 - The *name* of an attribute is a meta-attribute. For example when defining the data element *sample_type*, the meta-attribute, *attribute_name*, has the value “sample_type”.



Problem Statement



- The data model for the existing Planetary Science Data Dictionary (PSDD) is limited in its capabilities and needs an upgrade.
- This task does not address the content of the data dictionary.
 - The content of the data dictionary is addressed when the data model is addressed.



Problem 1 – Limited Definitions



- The definition of an attribute is limited in most data modeling tools.
 - Meta-attributes that are typically captured.
 - Name
 - Description
 - Data type
 - Valid values
 - Cardinality of values
 - Additional meta-attributes that are often needed.
 - Alternate names (aliases, acronyms, etc)
 - Registration authority
 - Representation information
 - Classification Schemes
 - Namespace
 - Valid Value definition and formation rule
 - ...

* See IPDA Implementation Requirements in Backup



Problem 2 – Enabling Interoperability

- Option 1 - The adoption of a common vocabulary by two or more repositories enables inter-operability between those entities.
- Option 2 - Where a single vocabulary is not in common between two repositories, inter-operability is dependant on the **identification** of commonality between vocabulary terms.
 - The use of a common mechanism for defining vocabulary terms makes it easier to identify commonalities between vocabulary terms.
 - A common mechanism for defining vocabulary terms provides interoperability at a deeper level.



Planetary Science Data Dictionary Example



- Data Element_Name - Sample_Type
- Data_Element_Description - The sample_type element indicates the data storage representation of sample value (*within an image*).
- Value_Type: Identifier (*Enumerated*)
- Valid_Values:
 - ieee_real
 - lsb_integer
 - lsb_unsigned_integer
 - msb_integer
 - msb_unsigned_integer
 - pc_real
 - unsigned_integer
 - vax_real



Some Questions a Data Dictionary Should Answer



- What is the organization that is responsible for managing the data element Sample_Type? Who can be contacted about its management? What is the source of the definition?
- What changes have been made to the definition and by whom? What is its version?
- What is the rationale for the data element? Is this a common concept shared by other data elements?
- What alternate names does the data element have?
- What natural languages are used in the definition?
- How is the data element classified? E.g. science vs operations, science discipline, etc.



More Questions

- Who controls the valid value “ieee_real? Who can be contacted about its management? What is the source of the valid value?
- What is the definition of “ieee_real”?
- What changes have been made to the definition of “ieee_real” and by whom? What is its version? What is the window within which this value was valid?
- What is the rationale for the set of valid values? Is this a common valid value concept that other data elements might use?
- What character set is used for the valid values? How many characters can be used for a valid value?
- How is the valid value represented and what is the unit of measurement?
- What are the minimum and maximum values allowed?
- What values signify Missing, Not Applicable, and Unknown for this data element.



ISO/IEC 11179



- ISO/IEC 11179 - Metadata registries (MDR)
 - Addresses the semantics of data, the representation of data, and the registration of the descriptions of that data.
 - It is through these descriptions that an accurate understanding of the semantics and a useful depiction of the data are found.
- The purposes of ISO/IEC 11179 are to promote the following:
 - Standard description of data
 - Common understanding of data across organizational elements and between organizations
 - Re-use and standardization of data over time, space, and applications
 - Harmonization and standardization of data within an organization and across organizations
 - Management of the components of data
 - Re-use of the components of data
- In ISO/IEC 11179 the basic container for data is called a **data element**. It may exist purely as an abstraction or exist in some application system.



- **Data Element**
 - Data Element, Data Element Concept
 - Concept Domain, Value Domain
- **Administration** – Status, Creation_Date, Effective_Date, Last_Change_Date, Registration_Authority, Steward, Submitter, Language
- **Classification Scheme** - User defined
 - Scope – e.g. Common, Imaging, Atmospheres
 - Mission – e.g. MEX, MER
- **Other**
 - Description, Data Type, Enumeration, Unit of Measurement, Value Meaning, Alternate Names, Language, Reference Documents, Relationships, Terminology
 - Effective begin and end date



Status



- Use cases have been defined.
- Requirements have been written.
- A data element registry has been prototyped.
 - Several data dictionary models were reviewed.
 - CCSDS Data Entity Dictionary Specification Language (DEDSL)
 - SPASE Data Dictionary
 - Three data dictionary models have been captured in an ontology.
 - Planetary Science Data Dictionary (PSDD) – PDS Data Dictionary
 - ISO/IEC 11179-3:1994 – Specification and standardization of data elements
 - ISO/IEC 11179-3:2002 – Metadata registries (MDR)
 - ISO/IEC 11179-3:2002 is being prototyped
 - Ontology was used to develop a prototype data element registry.
 - Planetary Science Data Dictionary content is being loaded.
 - Work is not yet complete.



Conclusions

- A data dictionary is required to support a data model since the data modeling tools typically do not capture all the information needed to completely defined the attributes used in the data model.
- A common data model for data dictionaries provides inter-operability at the most basic level.
 - Common terms should be used for defining data elements.
- The ISO/IEC 11179 Metadata Registry standard was designed to provide a common data model.
- The results are positive so far.
 - An ISO/IEC 11179 Data Model has been developed.
 - A ISO/IEC 11179 compliant database has been configured.
 - The content of the Planetary Science Data Dictionary is being test loaded.



Backup



Requirements

1. Manage Data Dictionary
2. Implement the Data Dictionary
3. Use the Data Dictionary to Create and Validate
4. Use Data Dictionary for Software Development



Implement the Data Dictionary (1 of 2)

1. The Data Dictionary shall define data elements in compliance with the data dictionary model .
2. The Data Dictionary shall define a “data-type” value set in compliance with the data dictionary model.
3. The Data Dictionary shall define a “units-of-measurement” value set in compliance with the data dictionary model.
4. The Data Dictionary shall define a “special-values” value set in compliance with the data dictionary model.
5. The Data Dictionary shall define a “data-element-value-alias” value set in compliance with the data dictionary model.
6. The Data Dictionary shall define a “data-element-value-formation-rule” value set in compliance with the data dictionary model.
7. The Data Dictionary shall define a “standard-value” value set in compliance with the data dictionary model.
8. The Data Dictionary shall define a “namespace” value set in compliance with the data dictionary model.
9. The Data Dictionary shall define a “general-classification” value set in compliance with the data dictionary model.
10. The Data Dictionary shall define a “system-classification” value set in compliance with the data dictionary model.
11. The Data Dictionary shall define linked-in-kind data element relationships in compliance with the data dictionary model.
12. The Data Dictionary shall define “pointer relationships” in compliance with the data dictionary model.
13. The Data Dictionary shall define “sfdu-relationships” in compliance with the data dictionary model.
14. The Data Dictionary shall have a naming standard for the data element title (common name).
15. The Data Dictionary shall provide a namespace attribute for indicating control authorities for groups of data elements.
16. The Data Dictionary shall provide a general data type attribute for classifying a data element according to a non-implementation-specific list of data types.
17. The Data Dictionary shall provide a general classification type attribute for classifying groups of data elements according to common characteristics
18. The Data Dictionary shall provide a system classification type attribute for classifying groups of data elements according to the data system that uses it



Implement the Data Dictionary (2 of 2)

1. The Data Dictionary shall provide a unit attribute for specifying the default unit of measure that is applicable to the referenced data element.
2. The Data Dictionary shall provide a linked-in-kind attribute for identifying relationships between data elements where the use / specification of a data element necessitates the specification of a complimentary data element.
3. The Data Dictionary shall provide an alias attribute for specifying one or more aliases that are applicable to the referenced data element.
4. The Data Dictionary shall provide a standard value type attribute for specifying the type of standard value that is appropriate for the referenced data element.
5. The Data Dictionary shall provide a minimum and maximum column value attribute for specifying the minimum and maximum numeric values that are applicable to the referenced data element.
6. The Data Dictionary shall provide a minimum and maximum length value attribute for specifying the minimum and maximum permissible length of the character values that are applicable to the referenced data element.
7. The Data Dictionary shall provide two IPDA-identifier attributes for specifying the unique instance of the data element and a locally defined instance of the referenced data element.
8. The Data Dictionary shall provide a textual-description attribute for defining the referenced data element.
9. The Data Dictionary shall provide a data-element-type attribute for specifying the type of data element.
10. The Data Dictionary shall provide a data-element-formation-rule attribute that supplies a rule that is to be applied during the creation of a value for the data element (e.g., the values supplied for reference_key_id must conform to the rules used by a specific professional journal for referencing citations).
11. The Data Dictionary shall provide a special-values attribute for specifying which of the special-values (e.g., "N/A", "UNK", "NULL") are permissible or not permissible as values for the referenced data element.
12. The Data Dictionary shall provide a series-set attribute for specifying if it is permissible or not permissible to specify values in a series or set.
13. The Data Dictionary shall provide a data-element-partial-label attribute that specifies if it is permissible or not permissible for a data element to exist within a partial-label (e.g., a FMT file).
14. The Data Dictionary shall provide a has-units attribute that specifies if it is permissible or not permissible to associate a unit with the referenced data element.
15. The Data Dictionary shall provide a can-be-locally-defined attribute that specifies if it is permissible or not permissible for a data element to be locally defined (i.e., overwrite the attributes of a data element in the PSDD).
16. The Data Dictionary shall provide a can-take-on-identifiers attribute that specifies if it is permissible or not permissible to pre-pend an identifier to a data element.
17. The Data Dictionary shall provide a can-be-a-pointer attribute that specifies if it is permissible or not permissible for the data element / object to be characterized as one of the three types of pointers (e.g., (1) data location pointer, (2) include pointer, and (3) related information pointer).
18. The Data Dictionary shall provide attributes for defining object classes.



ISO/IEC 11179 – High Level Meta-Model



data_element_concept_conceptual_domain_relationship

