Standards Change Request

PSDD Enhancement --- add "Case" as new attribute SCR3-1124.v6

Provenance:

Date: 2008-02-28
Author(s): R. Joyner (EN)
Working Group: R.Joyner (lead), L.Huber, T.King, E.Rye, D.Simpson,
S.Hardman (technical advisor),
Date: 2008-03-05: edited by Simpson
Date: 2008-03-10: edited by Simpson after WG telecon

Problem Summary:

Validation and Label Template design warrant more rigorous PSDD constructs to warn if keyword values are not PDS compliant. Most keyword values must consist of uppercase characters (e.g., DATA_SET_NAME, UNIT, etc); but there exists a set of keywords for which the value may consist of mixed-case characters (e.g., DESCRIPTION, NOTE, TEXT, etc), and there is a small set of keywords that must be lowercase (e.g., BL_NAME).

This SCR adds an additional attribute to the PSDD enabling:

- automated validation of case.

- data producers, through examination of the PSDD, to determine case requirements for each keyword.

The CASE_TYPE attribute is applicable to all data types (e.g., CHARACTER, ALPHANUMERIC, INTEGER, DATE) as the keyword-values may have character component parts.

Examples

```
DATE must be upper case: START_DATE           = 2008-02-24T23:59:59
REAL may be mixed case: A_AXIS_RADIUS         = 1.234E2 or 1.234e2
INT  may be mixed case: ACCUMULATION_COUNT    = 16#4B# or 16#4b#
FILE_NAME must be upper case:    FILE_NAME    = "DATAFILE.DAT"
DESCRIPTION may be mixed case:   DESCRIPTION  = "Blue and small"
```

Current Urgency:

None

Proposed Solution:

Add a new attribute to the PSDD: CASE_TYPE

Impact Assessment:

(1a) PDS Standards Reference; Sections requiring modifications:

- Section 5.4 (Integers)

Change: Integer values must be presented as a string of digits, optionally preceded by a sign. Specifically, no comma or point should be used to group digits. Values that are to be interpreted as integers must not be enclosed in quotation marks of any kind.

to: Base-10 integer values must be presented as a string of digits, optionally preceded by a sign. Specifically, no comma or point should be used to group digits. Values that are to be interpreted as integers must not be enclosed in quotation marks of any kind. Non-decimal integer values must be enclosed by "#" and preceded by the radix. Upper case letters A, B, C, ... (or their lower case equivalents) must be used for digits larger than 9.

- Section 11.4 Software Packaging Under Previous Versions of the Standards

Change: 2. SUN Software

- The problem in this case is preserving the SUN file names, since case is significant in file names on that platform. Since the ISO standard requires all file and directory names to be uppercase, a disk premastered as an ISO CD may encounter problems in the casesensitive SUN environment. Specifically, some CD readers mounted on SUN systems show file names as uppercase regardless of the format prior to mastering. If build routines ("make" files, for example) refer to lowercase file names, the corresponding files will not be found.

to: 2. SUN Software

- The problem in this case is preserving the SUN file names, since case is significant in file names on that platform. Since the ISO standard requires all file and directory names to be uppercase, a disk premastered as an ISO CD may encounter problems in the case-sensitive SUN environment. Specifically, some CD readers mounted on SUN systems show file names as lower case regardless of the format prior to mastering. If build routines ("make" files, for example) refer to upper case file names, the corresponding files will not be found.

(1b) PDS Standards Reference; Sections that address "case" and that are not to be modified:

(1) Section 5.1.2 Label format
- All values in a PDS label should be in upper case, except values for descriptive elements.

(2) Section 5.3.1 Label Standards Identifiers
- This keyword will use the upper case representation of the catalog identifier, e.g., PDSCAT1R47, PDSCAT1R48, etc.

(3) Section 5.3.3.2 Use of Pointers in Detached and Combined Detached Labels
- File names must be in uppercase characters.

(4) Section 6.3.1 Restrictions on DATA_SET_ID and DATA_SET_COLLECTION_ID
- Within the DATA_SET_ID and DATA_SET_COLLECTION_ID, acronyms are separated by hyphens. The only characters allowed are:
  - Uppercase characters, A-Z

(5) Section 8.2 Formation of Directory Names
- A directory name must consist of only uppercase alphanumeric characters and the underscore character (i.e., A-Z, 0-9, or "_"). No lowercase letters (i.e., a-z) or special characters (e.g., "#", "&", "*") are allowed.

(6) Section 10.1.1 ISO 9660 Level 1 Specification
- The base name and extension may only contain characters from the following set: the upper case alphanumeric characters (A-Z, 0-9) and the underscore ("_").

(7) Section 11.3 Packaging Software Files on a CD or DVD
- The ISO 9660 Level 1 standard requires all pathnames and directory names to be in uppercase, and to be limited to eight characters with a three-character file extension for file names.

(8) Section 12.1.2 Notation
- 1. Lower case words, some containing underscores, are used to denote syntactic categories. For example: units_expression

(9) Section 12.2.1 ODL Character Set - Letters
- The letters are the uppercase letters A - Z and the lowercase letters a - z. ODL language elements are not case sensitive. Thus the following identifiers are equivalent:
  - IMAGE_NUMBER
  - Image_Number
  - image_number

  Case is significant inside of literal text strings, i.e., string "abc" is not the same as the string "ABC".

(10) Section 12.2.3 ODL Character Set - Special Characters
- Case is significant within the quotes of a text string.

- Case is not significant within delimiting apostrophes (a.k.a. "single quotes").

(11) Section 12.3.1.2 Integer Numbers In Based Notation
- If X is greater than 10, then the letters A, B, C, D, E, F (or their lower case counterparts) are used as needed for the additional digits.

(12) Section 12.3.1.3 Real Numbers
- Note that the letter ÎEÌ in the exponent of a real number may appear in either upper or lower case.

(13) Section 12.3.2.5.1 Combining Date and Time
- The letter T separating the date from the time may be specified in either upper or lower case.

(14) Section 12.3.4 Identifiers
- Because ODL is not case sensitive, lower case characters in an identifier can be converted to their upper case equivalent upon input to simplify comparisons and parsing.

(15) Section 12.5.4.1 Implementation of Symbolic Literal Values
- Symbolic values are converted to upper case on input. This means that a lowercase string is converted to the equivalent uppercase string; as in the following example:

```
Original string:  SPACECRAFT_NAME = 'Voyager_2'
Converted string: SPACECRAFT_NAME = 'VOYAGER_2'
```

(16) Section 12.7.3 ODL/PVL Usage
(5) - Keywords and standard values must be in upper case. Literals and strings may be in upper case, lower case, or mixed case.

(17) Section 18.1 SI Units
- SI units are all written in mixed case; symbols are also mixed case except for those derived from proper names.

(18) Section 19.4 Volume Naming
- Volume names must be no more than 60 characters in length and in upper case.

(19) Section 19.4.1 Volume ID
- The volume ID must be in upper case.

(20) Section 19.5 Volume Set Naming
- Volume set names may be at most 60 characters in length and must be in upper case.

(21) Section 19.5.1 Volume Set ID
- VOLUME_SET_IDs must be in upper case, and are composed by concatenating the following fields, separated by underscores, using abbreviations if necessary:

(22) Section A.16.5 Example

- DESCRIPTION = "The geographical feature name with all diacritical marks stripped off. This name is stored in upper case only so that it can be used for sorting and search purposes.

(23) Section D.1.2 Example
- filename is the upper-case file name
  ext is the upper-case file extension

- In the table, directory names are enclosed in square brackets ([]), upper-case letters indicate an actual directory or file name, and lower-case letters indicate the general form of a set of directory or file names.

(2) Archive Preparation Guide: None.

(3) Proposer's Archive Guide: None.

(4) Planetary Science Data Dictionary: New keyword definition.

(5) PDS Tools:
- EN Tools: VTool and LTDTool will require minor code modifications
- DN Tools: impact TBD

Additional Information:

Each element in the PSDD will be modified to set CASE_TYPE to:
- "LOWERCASE",
- "UPPERCASE",
- "MIXEDCASE",
- "NOMIXEDCASE"

All elements will have CASE_TYPE set to UPPERCASE; with the following exceptions:

(1) Keywords ending in either DESC or DESCRIPTION will have CASE_TYPE = MIXEDCASE:

- ABSTRACT_DESC
- ALGORITHM_DESC
- CITATION_DESC
- COLUMN_DESCRIPTION
- COMMAND_DESC
- CONTAMINATION_DESC
- COORDINATE_SYSTEM_DESC
- DATA_QUALITY_DESC
- DATA_SET_COLLECTION_DESC
- DATA_SET_COLLECTION_USAGE_DESC
- DATA_SET_DESC
- DATA_SET_OR_INST_PARM_DESC
- DATA_SET_TERSE_DESC
- DATA_SOURCE_DESC
- DESCRIPTION
- DETECTOR_DESC
- DISCIPLINE_DESC
- EARTH_BASE_DESC
- ELECTRONICS_DESC
- EVENT_TYPE_DESC
- EXPERTISE_AREA_DESC

- FEATURE_TYPE_DESC
- FORMAT_DESC
- FORMATION_RULE_DESC
- FRAME_PARAMETER_DESC
- INST_CMPRS_DESC
- INSTRUMENT_CALIBRATION_DESC
- INSTRUMENT_DESC
- INSTRUMENT_FORMATTED_DESC
- INSTRUMENT_HOST_DESC
- INSTRUMENT_MODE_DESC
- INSTRUMENT_MOUNTING_DESC
- MAP_DESC
- MAP_PROJECTION_DESC
- MEASUREMENT_ATMOSPHERE_DESC
- MEASUREMENT_SOURCE_DESC
- MEASUREMENT_STANDARD_DESC
- MEASUREMENT_WAVE_CALBRT_DESC
- MEDIUM_DESC
- METHOD_DESC
- MISSION_DESC
- MISSION_PHASE_DESC
- MODEL_DESC
- MOSAIC_DESC
- NODE_DESC
- OPERATIONAL_CONSID_DESC
- OPTICS_DESC
- ORDER_STATUS_DESC
- PARAMETER_DESC
- PEER_REVIEW_RESULTS_DESC
- PLATFORM_OR_MOUNTING_DESC
- PROCESSING_LEVEL_DESC
- QUATERNION_DESC
- RATIONALE_DESC
- RECEIVER_DESCRIPTION
- REFERENCE_DESC
- REFERENCE_POINT_DESC
- REGION_DESC
- REQUEST_DESC
- RESEARCH_TOPIC_DESC
- RESOLUTION_DESC
- ROLE_DESC
- ROTATIONAL_ELEMENT_DESC
- SAMPLING_DESC
- SCAN_PARAMETER_DESC
- SELECTION_QUERY_DESC
- SENSITIVITY_DESC
- SOFTWARE_ACCESSIBILITY_DESC
- SOFTWARE_DESC
- SPACECRAFT_DESC
- SPACECRAFT_ORIENTATION_DESC
- SPACECRAFT_POINTING_MODE_DESC
- SPECIALTY_DESC
- SPECTRAL_ORDER_DESC
- STANDARD_VALUE_SET_DESC
- STAR_DESCRIPTION
- SUPPORT_REQUEST_DESC
- SYSTEM_BULLETIN_DESC
- TABLE_DESC
- TARGET_DESC
- TEMPERATURE_TRANSLATION_DESC
- VECTOR_COMPONENT_TYPE_DESC
- VOLUME_DESC

(2) Keywords ending in NOTE will have CASE_TYPE = MIXEDCASE:

- ARCHIVE_STATUS_NOTE
- CONFIDENCE_LEVEL_NOTE
- INVENTORY_SPECIAL_ORDER_NOTE
- LABEL_REVISION_NOTE
- NOTE
- STATUS_NOTE
- SYSTEM_EVENT_USER_NOTE
- TEMPLATE_NOTE
- USAGE_NOTE

(3) Keywords ending in TEXT will have CASE_TYPE = MIXEDCASE:

- ABSTRACT_TEXT
- ADDRESS_TEXT
- COMMENT_TEXT
- HELP_TEXT
- KEYWORD_VALUE_HELP_TEXT
- PROCESSING_HISTORY_TEXT
- RELEASE_PARAMETER_TEXT
- TRANSFER_COMMAND_TEXT
- VOLUME_INSERT_TEXT

(4) Keywords ending in CHECKSUM will have CASE_TYPE = MIXEDCASE:

- CHECKSUM
- MD5_CHECKSUM

(5) The following keywords will have CASE_TYPE = LOWERCASE

- BL_NAME
- BL_SQL_FORMAT
- TERSE_NAME

(6) Any keyword which can take on a numerical value in scientific notation (e.g., 1.234e5) or base notation (e.g., 16#4b#) will have CASE_TYPE = MIXEDCASE. This includes all keywords where GENERAL_DATA_TYPE is specified as REAL, INTEGER, NON DECIMAL, or NON_DECIMAL.

Requested Changes:
---

(1) Add the following element definition to the PSDD:

```
OBJECT = ELEMENT_DEFINITION
  NAME = CASE_TYPE
  STATUS_TYPE = PROPOSED
  GENERAL_DATA_TYPE = IDENTIFIER
  UNIT_ID = NONE
  STANDARD_VALUE_TYPE = STATIC
  MAXIMUM_LENGTH = 11
  DESCRIPTION = "
      The CASE_TYPE keyword defines whether a keyword value must be
      comprised of upper case characters (i.e., CASE_TYPE = UPPERCASE)
      or lower case characters (i.e., CASE_TYPE = LOWERCASE).  If there
      are no restrictions the value is set to MIXEDCASE (i.e.,
      CASE_TYPE = MIXEDCASE.  If upper or lower case (but not both) is
```

```
        allowed then CASE_TYPE = NOMIXEDCASE.  Upper case characters
        consist of 'A' through 'Z'.  Lower case characters consist of
        'a' through 'z'.  The mixed case character set includes both."
    STANDARD_VALUE_SET = {
       "LOWERCASE",
       "UPPERCASE",
       "MIXEDCASE",
       "NOMIXEDCASE"
       }
END_OBJECT = ELEMENT_DEFINITION
END
```

(2) Add CASE_TYPE to all elements in the PSDD in accordance with the rules specified above.