Data Provider Validation Policy Discussion

Dan Crichton
PDS Management Council
April 2014

Action Item

- 2013-11-18/A (Crichton, 2013-11-19): Draft one or more policies on validation (and other 'entrance') requirements to be levied on future data providers. Propose new language on this topic to be added to the APG and to SIS and ICD templates.

Goals

- Improve the quality of PDS4 bundles from data providers by ensuring that PDS supplied tools coupled with manual verification is applied prior to delivery of data to a PDS node.
- Provide policy and guidance for consistency across the PDS.
- Provide clear, practical steps to be included in the data management plan.
- Encourage the use of PDS supplied and COTS tools (oXygen)

What this does not cover

- Whether something is “acceptable” as an archive product
  - What structures and format in being used
  - Science quality

Types of Data Provider Validation

Validation Type	Definition
Syntactic Validation	Validate the grammatical structure / syntax of the XML label
Semantic Validation	Validate that the interpretation of the values in the XML labels are consistent with the circumstances and contexts being described
Content Validation	Validate that the syntactic and semantic content of the XML label accurately describes the digital bits in the data objects being described (i.e., the physical structure and order of the bits are accurately described by the data objects in the xml label)
Referential Integrity	Validate that the relationships described, in and between digital objects described in the XML label, are consistent and represented

PDS Data Provider Validation Support

Validation Type	Support
Syntactic Validation	Validate tool - ensures accuracy of the grammar / syntax of the XML label against the appropriate schemas as well as any mission schema or schematron.
Semantic Validation	Validate tool - ensures accuracy of the semantic "meaning" of the XML label when run against all relevant schema and Schematron files.
Content Validation	Manual inspection which can be done either through visual inspection or adherence to a set of rules (e.g., # of column headers, checksum).
Referential Integrity	Validate tool (coming feature) - validate that the relationships described, in and between digital objects described in the XML label, are consistent and represented. NOTE: this can only be done with the local bundle. Full RI won't occur until ingestion/registration.

Policy for Data Provider Validation of PDS4 Bundles

Data providers delivering bundles shall ensure adherence to PDS4 standards by ensuring that the following criteria are met:

1) Syntactic validation: a) the XML label is validated against the appropriate schema rules; b) a mission schematron is syntatically correct; and c) a mission schema is syntactically correct.

2) Semantic validation: the XML label is validated against the appropriate schematron rules.

3) Content validation: the XML label accurately describes the data product.

4) Referential integrity: the relationships described, in and between digital objects described in the XML label, are consistent and represented

PDS supplied software tools will be made available to data providers in order to support syntactic, semantic, specific content rules and referential integrity validation. Data Providers should use visual inspection to validate content that can not be done programmatically.

Examples

(1) syntactic validation:

(a) xml tags match (accurate): <mission_name>Voyager</mission_name>

(b) xml tags match (error): <mission_name>Voyager</name>

(2) semantic validation:

(a) value is meaningful (accurate): <mission_name>Voyager</mission_name>

(b) value is meaningful (error): <mission_name>Mars</mission_name>

(3) content validation:

(a) data object matches type of data (accurate):

character data file described using <Table_Character>

(b) data object matches type of data (error):

binary data file described using <Table_Character>

Next steps

- Determine whether this is the right set of functions in accordance with the goals
- Review and discuss policy
- Identify appropriate language for the DM Plan