EPSC Abstracts
Vol. 13, EPSC-DPS2019-856-1, 2019
EPSC-DPS Joint Meeting 2019

EPSC
EPSC-DPS Joint Meeting 2019
15–20 September 2019 | Geneva, Switzerland

Mega-Metadata: Rosetta for PDS4

Anne C. Raugh (1), David J. Heather (2), and James M. Bauer (3)
(1) University of Maryland, Astronomy, USA (araugh@umd.edu), (2) ESA European Space Astronomy Centre, Madrid, Spain (dheather@cosmos.esa.int), (3) University of Maryland, Astronomy, USA (gerbsb@umd.edu)

Abstract

The data returned by the European Space Agency's (ESA's) Rosetta mission over its lifetime are unique, varied, and of extremely high interest. 27 instruments have, to date, produced over 3000 individual data sets for the mission archive. The Rosetta archives, however, were designed at the turn of the century – before the new information model-based PDS4 Standards, and before the Planetary Science Archive (PSA) was established. The Rosetta mission archive is a last-generation product of the PDS3 standards, and the task of migrating these data and the rich metadata incorporated into the PDS3 labels via the PDS3 mission dictionary is formidable. We report on steps taken to convert the monolithic PDS3 mission metadata dictionary into a set of PDS4 information model-based "local dictionaries" following the best practices established for the PSA PDS4 mission archives.

1. Introduction

The initial design of the Rosetta mission archives was undertaken in the early 2000s. When the option of defining mission keywords for use in PDS3 labels became available, Rosetta made quick use of it. PDS3 mission dictionaries, however, were inherently unstructured because of limitations in the PDS3 standards. For a mission like Rosetta, with over two dozen instruments generating data, this is far from an ideal situation.

The organizational structure presented by PDS4 namespaces and metadata modeling tools enables logical, functional, and systematic organization of metadata across the entire mission, thus supporting interoperability across the Rosetta instruments. The PSA has also established best practices for the mission archives being developed under their supervision. Following these guidelines for the Rosetta PDS3-to-PDS4 migration will support interoperability across PSA missions as well. The expected return on the investment in migrating these data, then, is high.

2. Migration from PDS3 to PDS4

The first aspect of the migration to be designed is the metadata. The process is not a straight-forward mapping of PDS3 keyword to PDS4 concept, unfortunately. One of the primary reasons for the development of PDS4 was that the PDS3 metadata definitions had been so frequently interpreted and re-interpreted from context to context that attempting to trace precise keyword meanings in any arbitrary label was a nearly intractable problem. Migration, then, involves determining the correct local interpretation of PDS3 common dictionary keywords so they can be properly mapped to the more precise PDS4 equivalents, where they exist, or assigned to a local dictionary, where they can be defined in context.

For keywords defined in the PDS3 mission dictionary, the process is more straight-forward. The definitions collected for the Rosetta dictionary provide the primary content needed to produce the PDS4 dictionaries, and they have but a single context for interpretation.

3. PSA Best Practices

The PSA, which oversees the archiving efforts for all ESA planetary missions, has established best practices for its missions aimed at ensuring new mission data can be integrated smoothly into their PDS4-based archive, while supporting the autonomy of pipeline development for instrument teams. To this end, PSA has established guidelines for creating mission namespace hierarchies wherein the main mission namespace can be subdivided by instrument. Each instrument team, then, develops the instrument-specific metadata dictionary. In an active mission, this gives the instrument team direct control over a very valuable validation tool [1].

4. General Approach

Our general approach to developing PDS4 dictionaries for Rosetta is to start by parsing the PDS3 mission dictionary into the new mission and

instrument namespaces. These baseline sets of "attributes" (in PDS4 parlance) are then organized into functional groups called "classes". PDS4 dictionary classes define the structural and existential metadata requirements for the namespace and are key to validation.

Once the baseline dictionary content is defined for an instrument, the pipeline products for that instrument will be analysed to identify other PDS3 keywords that need to be incorporated into the instrument dictionary. This involves tracking down details of how the keywords were used in the instrument label context. Details of that nature are typically summarized in interface control documents, all of which are available in the Rosetta PDS3 archive.

Keywords flagged for inclusion in the instrument dictionary must then be incorporated into the dictionary classes or used to define new classes. Unlike the keywords in the Rosetta mission dictionary, the definitions and constraints for these keywords cannot be simply cut and pasted into the new PDS4 dictionaries. These keywords are precisely those in PDS3 that were often subject to re-interpretation. Precise definitions and constraints will be established for the Rosetta instrument context for these keywords.

Once the input data are collected and validation constraints defined, the actual preparation of the dictionary files becomes a largely mechanical task. The goal of this effort is to produce the dictionaries for each level of the Rosetta PDS4 archive, and mappings from the PDS3 label content to the PDS4 design, so that migration of the data to PDS4 format can begin.

5. Status

We will report on the current status of the effort, including any unexpected obstacles encountered, any lessons already learned, and projected schedules for completion.

References

[1] Raugh, A. and Hughes, J.: PDS4 Dictionaries as Quality Assurance Tools, AAS/Division for Planetary Sciences Meeting Abstracts, 114.12, 2018.