The Importance of Quality Metadata in PDS4 Product Labels

Thomas C. Stein, tstein@wustl.edu

Dept. of Earth, Environmental, and Planetary Sciences
McDonnell Center for the Space Sciences
Washington University in St. Louis

High-quality metadata are foundational to the usability, longevity, and scientific value of Planetary Data System (PDS) archives. The PDS4 Information Model (IM) enforces well-structured metadata that provide consistency, completeness, and clarity. These qualities are key to search relevance, API interoperability, automated analysis pipelines, and data reuse across missions and disciplines. Successful production of such metadata begins early in instrument and mission development phases.

PDS4 product labels are key to productive scientific workflows. By embedding identifiers, context, and provenance, they enable faceted search, semantic tagging, FAIR-aligned filtering, API-driven access, label-based retrieval, automated pipelines, provenance tracking, and cross-mission synthesis. This metadata-centric approach treats labels as a contract between data producers, repositories, and users, ensuring that products remain discoverable, interpretable, and reproducible over time.

Barriers to creating robust metadata include varying levels of optional input supplied by providers, incomplete product references, occasional misapplication of values, and inconsistent use among related data sets. The PDS4 IM addresses many long-standing issues from the older PDS3 approach (e.g., ambiguous field names, missing units, inconsistent formats). It defines the classes, attributes, relationships, and constraints that govern all PDS4 labels, ensuring semantic integrity and cross-node interoperability. Discipline-specific extensions evolve within a governed frame□work that preserves compatibility and shared vocabulary.