The PDS Archives: Ensuring Quality and Usability

Raugh Anne[1] and Bauer James[1]

[1]University of Maryland College Park

November 16, 2022

Abstract

When NASA established the Planetary Data System (PDS) in the late 1980s, its mandate to the PDS was not merely to preserve the bytes from NASA's planetary science missions, but to maintain the usability of the data for present and future generations. Two fundamental pillars support this ambitious goal: The external peer review required for acceptance of all archived data submissions; and the PDS Standards for data and metadata formatting and completeness. The PDS external peer review process is at least equivalent to, if not more rigorous than, the journal refereeing process(1). Data reviewers who are field experts but not affiliated with the data preparer, nor involved in the PDS consulting process, are brought in to review documentation and completeness. They are specifically charged to attempt to use the data to perform some scientific investigation (reproducing published results, comparison to correlated observations for consistency, etc.). If the reviewers are not successful, the impediments are documented and the data submission is amended by the preparer until the reviewers are satisfied. This process demonstrates immediate usability of the data. The PDS Standards, and in particular the recently-implemented version based on the PDS4 Information Model, require exhaustive metadata documenting data structure, observing circumstances, provenance, analytical metadata, and so on using the same templates across the entire archive. The associated schematic enforcement of at least minimal requirements for metadata completeness and quality provides a foundation for discoverability, interoperability, and usability of data from disparate sources throughout the archive. Together, the PDS external peer review and the Information Model-based PDS4 standards ensure both quality and usability for data accepted into the PDS archive, for this and future generations of planetary scientists. Reference: (1) Raugh, A. and Bauer, J., PDS Data Sets as Peer-Reviewed References, Poster presented at the 15th Annual Meeting of the Asia Oceania Geosciences Society Meeting, 03-08 June 2018, Honolulu, Hawai'i.

The PDS Archives: Ensuring Quality and Usability

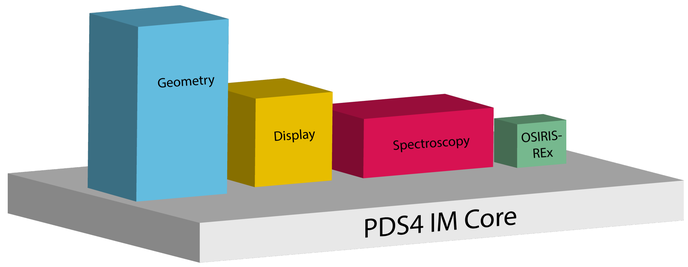
Anne Raugh (*araugh@umd.edu*) and James Bauer
University of Mayrland, College Park, MD

AGU 100 FALL MEETING Washington, D.C. | 10-14 Dec 2018
IN41E-0891

The Planetary Data System is charged with preserving and maintaining the usability of the data resulting from NASA's planetary science missions and programs, for future generations. Three main factors contribute to PDS and its data suppliers ensuring that data archived are complete, of high quality, and will be usable for the next 50-100 years.

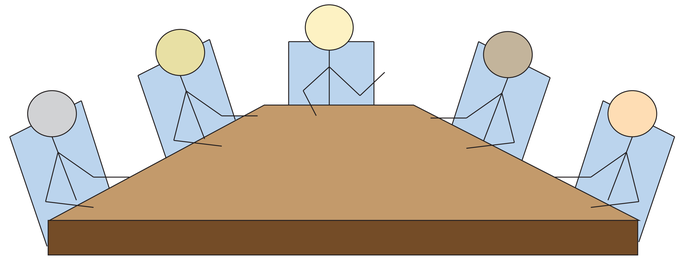
1. Data Standards

The first critical element is data file structure. PDS archival data must be independent of operating environment and software to have any hope of being "usable" past the decade in which the files were created. PDS standards constrain data files to very simple structures, very explicitly defined, and very hard to mis-read. The goal here is twofold: first, to ensure that writing I/O routines for PDS archive files will be easy to do, regardless of how programming languages and environments evolve; and second, to limit the archive to simple, well-documented data structures that are independent of the availability of software or the original data creators. Simple structures mean coding errors are likely to be easily recognized; limited structures mean large parts of the archive holdings become accessible through a single new I/O routine.



2. Information Model

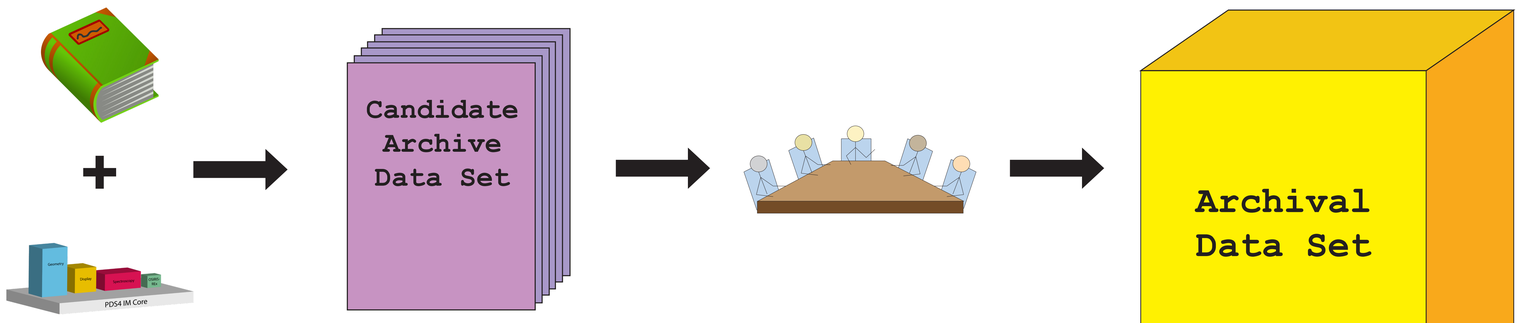
The next element, also essential to the data design process, is documentation via label metadata. The metadata describe the physical structure, of course, but the PDS Information Model (IM) contains extensive metadata structures to document – in a uniform and consistent way across the archive – the additional metadata that define the context and significance of the data. Observational parameters, instrument settings, pointing, display orientation, and so on are all included in the IM. At an even higher level, the IM metadata relate data products to other elements of the PDS archives (e.g., raw predecessors of calibrated data products) and to the published literature (e.g., papers directly cited by the data product, or publications describing the instrument or calibration procedures).



3. Peer Review

Planning and design are not, themselves, sufficient to guarantee a complete and maintainable data set. To build a stable, usable, generational archive it is vitally important to *demonstrate* quality and usability for all candidate submissions before they are accepted for archiving. PDS does validate submissions against its own standards and information model; but to truly demonstrate that the data are complete and usable on intake, PDS conducts an external peer review – a referee process – on each candidate dataset. Scientists unaffiliated with the data provider or PDS are asked to exercise the data. They attempt a real-world analysis (reproducing a published result; performing calibration; comparison with correlated observations; etc.) and report on results. Any omissions, corrections, or problems encountered are discussed with the data providers at an open meeting, and a plan is agreed for editing the data submission accordingly before it is accepted for archiving.

Small Bodies Node



The PDS4 Standards and Information Model guide data preparers to designing simple, well-documented candidate archive submissions.

The referee process exercises the data and demonstrates usability.

The end result is an archival data set: well-documented, well-formatted data that has demonstrable value and usability. PDS can be confident that that usability can be maintained for generations of planetary scientists.