The PDS Archives: Ensuring Quality and Usability

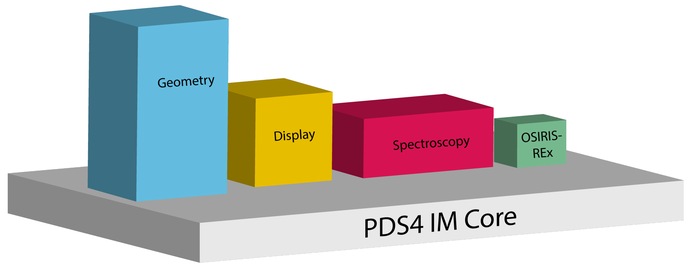
Anne Raugh (*araugh@umd.edu*) and James Bauer
University of Mayrland, College Park, MD

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.

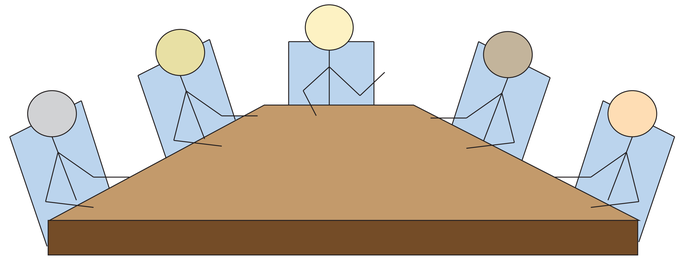
1. Data Standards

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

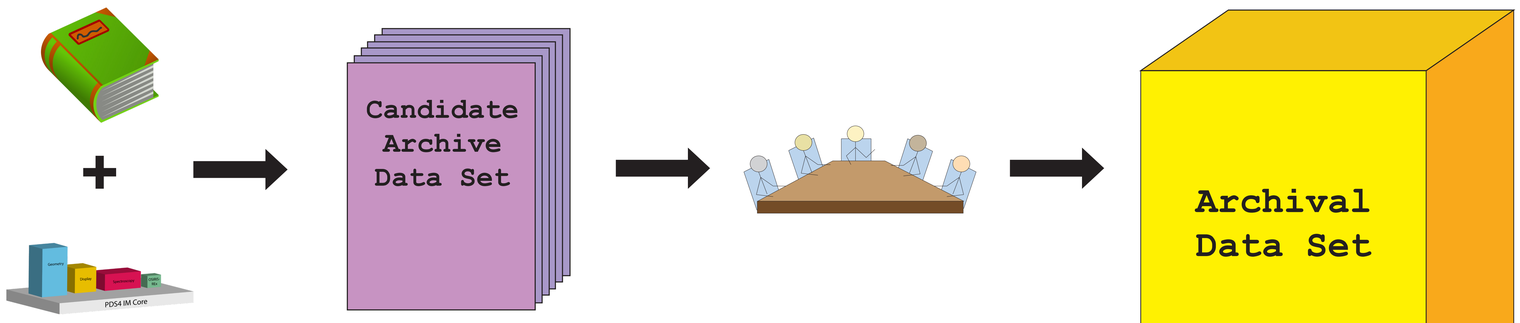
The next element, also essential to the data design process, is documentation via label metadata. The metadata describe the physical structure, of course, but the PDS Information Model (IM) contains extensive metadata structures to document – in a uniform and consistent way across the archive – the additional 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.