

## PLANETARY DATA SYSTEM: SUPPORTING ARCHIVING OF DERIVED DATA

Lynn D. V. Neakrase<sup>1</sup>, Reta F. Beebe<sup>1</sup>, Nancy J. Chanover<sup>1</sup>, Lyle F. Huber<sup>1</sup>, Daniel Crichton<sup>2</sup>, Sean Hardman<sup>2</sup>

<sup>1</sup>Planetary Data System, Atmospheres Node, Dept. of Astronomy, New Mexico State University, Las Cruces, NM 88003-8001.

<sup>2</sup>Planetary Data System, Engineering Node, Jet Propulsion Laboratory, Pasadena, CA 91109.

**Introduction.** Derived data comprise a wealth of information from scientific investigations that are useful to the community and are desired by NASA's Planetary Data System for archiving. Many planetary missions, field expeditions, and laboratory experiments produce data that should be preserved for posterity because of useful results, unique derivations, or accumulating progress in planetary exploration. NASA's Planetary Data System (PDS) [1] is charged with long-term preservation and archiving of digital data products pertaining to planetary exploration. Although the bulk of work done by the PDS is directly related to planetary mission data archiving, supporting data and derived data are also vital parts to the archive.

PDS is comprised of six science discipline nodes (Atmospheres, Cartography and Imaging Systems, Geosciences, Planetary Plasma Interactions, Ring-Moon Systems, and Small Bodies) located around the United States. These six separate discipline nodes use a common archiving standard for interoperability and long-term stability. The standard for archiving, known as PDS4, represents the fourth incarnation of PDS archive standards and has been recently implemented as of ~2012. New requirements through NASA ROSES-2015 and beyond include a Data Management Plan that will further increase derived data archiving across the PDS from programs such as Planetary Data Archiving, Restoration and Tools (PDART) and the other Data Analysis Programs (DAPs). Therefore, knowledge about the general practices of preparing and submitting data to the PDS under the new PDS4 standard will be useful to planning and costing proposals, preparation and submission of the data, and final review and lien resolution of the completed archive. For the work presented here, the Atmospheres Node at New Mexico State University [2] represents one approach

to archiving under PDS4, although the methodology employed by the other nodes mirrors the general practices under the new standard.

**PDS4: The New Standard.** PDS4 represents the latest archiving standard and a vast improvement over the past. PDS4 employs a centralized Information Model to unify the handling of the descriptive metadata and is expressed using eXtensible Markup Language (XML) labels. XML is a common modern markup language that allows easier access across multiple platforms and operating systems with a multitude of publicly available editors.

Beyond the increased accessibility of XML, the PDS4 system also includes a centralized registry necessary for ease of access of all registered data from any PDS search interface at any node. Key to this process are the enhanced product logical identifiers (URNs) that provide unique IDs that can also be used in Google-like searches directly similar to DOIs (once products are fully registered in the system).

At the Atmospheres Node, a further initiative has begun to evolve our website designs to aid users in finding and using scientific data, starting with mission-based and now including ground-based observations, field data, and lab data. "Resource products", or websites, also can have logical-identifier URNs and can be registered within the system allowing them to be prioritized in search results. Thus providing users with user guides and background information about the requested data up-front as they begin their research. Websites have been designed to answer science scenarios including novice users getting acquainted with PDS for the first time and ease of access to data for veteran users.

**Preparation of Derived Data.** In order to provide earlier, better input for proposers and general usefulness to data providers, the Atmospheres Node has begun to refine our submission process in response to the new PDS4 standards. The basic procedures are congruent with PDS as a whole, but we provide examples specific to our node. Regardless of your intent to submit to a funded research program opportunity, or as a self-motivated project in response to publishers' requests for data management, the process is the same.

**Step 1:** Contact the appropriate node with the request to submit data. Dialogue between PDS and the data provider should be done as early as possible. Information provided to PDS should include the scope of the project and the type of data to be submitted. Projected data volumes could also be useful for PDS storage planning.

**Step 2:** PDS and the provider iteratively work to design the archive to provide the best labels to the proposed data with the most complete metadata for long-term storage and future use. This step supplies the provider with label templates to use with the data products and ensures the resulting archive organization will be PDS4-compliant.

**Step 3:** Data providers are expected to provide adequate documentation for their submitted data. Documentation should be seen as an accompanying "Users' Guide" and should include information about data generation, calibration, assumptions, and usability, with background references. References to refereed journal articles can be accommodated in the documentation without the inclusion of PDFs of the article, but they could also be included in the document collection barring copyright issues.

**Step 4:** All data submitted to the PDS must be peer-reviewed for integrity before it can be released as "certified data". Internal PDS staff and selected external experts are

utilized as reviewers to ensure PDS usability and scientific viability. Results of the review may provide liens to be resolved by the providers. Once the liens are resolved the data can be released to the public.

**Step 5:** While the final reviews are concluding, PDS will design websites to act as distribution points for the data. Searchability through the PDS4 Registry is only one part of distributing the newly archived data. Web page resource products also serve as a one-stop-shop for learning about the data contained within and allow users easy access to background information and Users' Guide documentation.

PDS4 allows the Planetary Data System to move forward to better serve the planetary science community in the years to come. The key to a good planetary archive is good planetary data. Good planetary data under PDS4 contains more than just the data but also includes information about how to use the data effectively through detailed documentation and useful websites. In any case the most important step in the process is mutual communication between the node and data provider. Exact timing of each of the steps provided here are variable depending on the scope of the project and the complexity of the data.

[1] NASA Planetary Data System website: <http://pds.nasa.gov>.

[2] NASA PDS Atmospheres Node: <http://pds-atmospheres.nmsu.edu>