SUMBITTING DATA TO THE PDS: DATA MANAGEMENT PLANS, DERIVED DATA, AND MORE.

L.D.V. Neakrase, L. Huber, N. Chanover, R. Beebe, and J. Johnson. NASA Planetary Data System, Atmospheres Node, Department of Astronomy, New Mexico State University, P.O.Box 30001 MSC 4500, Las Cruces, NM 88003-3001. lneakras@nmsu.edu

Introduction: Traditionally NASA's Planetary Data System (PDS) has been responsible for interfacing with missions and instrument teams for production of data archives of planetary data. In response to the NASA Plan for Increasing Access to the Results of Scientific Research, which involves ensuring long-term accessibility of digital scientific data and peer-reviewed publications, many more investigators are turning to the PDS for their archiving needs for derived data. This includes data resulting from further analysis of mission observations, modeling results used to interpret NASA mission data, groundbased telescopic observations, and laboratory and field measurements. The PDS has modernized its archiving approach by moving to a new archiving standard, PDS4 [1]. This new standard at its core is an XML-based, model-driven system for storing metadata in the new labels.

PDS4 and XML: The PDS is now using the PDS4 archive standard for controlling the metadata labels for all new data coming into the archive. This new system is driven by a core information model that maintains the standard for archiving. All possible label outcomes are derived from this information model and the labels are expressed in XML. XML (Extensible Markup Language) is a well-known standard designed for easy management of data over the internet. It is readily editable with most modern text editors, though specific XML editors exist (e.g., Oxygen (subscription), Eclipse (free)). The XML label format is controlled by the Information Model and guarantees consistency across all of the PDS. All products that have labels can be registered in the PDS4 Central Registry allowing ease of access through any of the PDS search engines.

Data submitted to the PDS is organized into Bundles, Collections, and Products. Products represent the file-level organization and include not only the data files themselves but also document sets and browse images, etc. with corresponding labels. These products can be organized into logical Collections, usually by processing level or purpose (e.g., data_raw, data_calibrated, browse, document, geometry, etc.). All collections are then organized by project as a Bundle. Bundles contain all data and documentation useful for describing, explaining, and using the data.

Data Management Plans: Most data analysis programs through ROSES [2] now require a Data Management Plan (DMP). The requirements of data management may vary by program, but in general all require data providers (PIs) to demonstrate how the data produced in their project will be accessible beyond the terms of their grant. In most cases, programs require the use of the PDS or some other PDS-equivalent archive. PDS is mandated to peer review all data coming into the archive and to preserve and maintain data for more than 50 years. The resulting restrictions on data formats are for usage of simple, proven formats (e.g., ASCII/CSV tables, FITS, IMG, PDF/A, etc.).

A good DMP should include basic information about the planned data submission, including what format and how much data there will be, and how it will impact the planetary community. A letter of support from the relevant PDS node is now required by many programs and should demonstrate that the node and the PI are in agreement of what is expected in the submitted data. PIs should demonstrate within the DMP that they understand the archiving process and what their responsibilities will be with respect to submitting data to the node. For PDS-equivalent archiving, the PI should demonstrate that the proposed archive location will maintain the data for a significant period of time (letter of support) and that the data have been reviewed or vetted with some sort of agreement from the facility or in peer-reviewed publications to prove equivalence.

Derived Data: In response to these new requirements for derived data produced by grant programs and published in journals, PDS has widened the scope of the data that can be archived. In the past PDS predominantly archived mission and mission-related data from spacecraft sent anywhere in the Solar System. Currently laboratory experiments/simulations, analog field studies, astronomical observatory observations, as well as new calibrations or combinations of mission data are fully acceptable under PDS4.

For aeolian research, typical data, outside of mission data that have been submitted to PDS include wind tunnel experiments/observations, analog field data for particle entrainment and flux studies and laboratory data pertaining to atmospheric compositional constraints. Many of these data are currently in the preparations stage but should be available later this year.

Conclusions: The key to creating a good archive is a good relationship with the PDS. For many PIs this begins with the DMP and the letter of support from the appropriate node. This process should begin as early as possible with contacting the PDS node and beginning to discuss the data to be archived. The node will pro-

vide you with a basic timeline of responsibilities for archiving. This will include helping tailor XML labels to your needs, including agreeing upon accepted archive formats for data and documentation. The final peer-review process should also be discussed and budgeted into the DMP. Typically peer review and lien resolution can take up to a month to complete, so submission of the data to PDS must be scheduled well in advance of the end of funding.

For projects that are not funded, PDS is still available to help archive your data. Derived data are often overlooked because of production late in a project's timeline. Projects often plan to create derived products but run out of time and/or money during their nominal funding period. PDS is still available for help after funding runs out, but the key to success is communicating with the node to schedule your project. Figure 1 illustrates an example of the division of labor on any project coming into the Atmospheres Node. Early and good communication with the relavant node is essential for producing a good DMP and hopefully in turn, a good archive.

References: [1] PDS Archiving Documentation, https://pds.nasa.gov/pds4/doc/index.shtml [2] Research Opportunities in Space and Earth Sciences 2017 (ROSES-2017), https://nspires.nasaprs.com/external/solicitations/summary.do?method=init&solId=%7BE757EF32-60E6-76AE-A276-21A1F8BA96BB%7D&path=open [3] PDS Atmospheres Node website, http://atmos.nmsu.edu

Figure 1. A step-wise list of the archiving responsibilities for planning an archive submission, including what the supporting node will provide. This example is specific to the Atmospheres Node, although it will be similar for