

WIND, DUNES, AND DATA: PDS IS EXPANDING, IMPORTANT NEW TOOLS. L.D.V. Neakrase, S.K. Rees, L. Huber, N. Chanover, T.D. Pagán, K. Sweebe, V.I. Martinez. PDS Atmospheres Node, Department of Astronomy, New Mexico State University, P.O. Box 30001 MSC 4500, Las Cruces, NM 88003-8001. lneakras@nmsu.edu

Introduction: With the advent of new requirements and recommendations within many NASA funding opportunities (e.g., ROSES [1]), data management throughout the project and making data publicly available at the end is more important than ever. Within the aeolian community, it is not always clear which NASA's Planetary Data System (PDS) nodes are appropriate for aeolian laboratory and field data, planetary analog studies, or reprocessing/recalibrating of planetary mission data [2]. Further development of the PDS4 Archiving Standards [3] is now making a wider range of data products appropriate for the archive.

Proposing with PDS: Planning for a PDS archive begins with the proposal. The initial phase of interacting with PDS is to contact the appropriate node during the proposal preparation. This allows interactions with the node to begin, checking archive format restrictions for the data and documentation, and give the node an initial estimate of the types of data and the total volume to be archived. This initial contact with PDS is extremely important for a successful archive allowing good communication with the node personnel and the node can begin to plan for the ingestion of new data into the archive. This allows the node to also plan for allotting disk storage space and double checking that all needed product references are ready for proposed archive products.

Once the proposal is selected, the first step should be to contact the PDS node immediately and let them know of the selection. At ATM, we will provide an instructional guide [4] to help educate and plan for the archiving process. This guide also provides a checklist timeline that will help organize data collection from the archive perspective. Remember, the node also has responsibilities to track all funded projects and report quarterly to the PDS Project Office on progress or on issues that may require help from NASA Headquarters program officers. Remaining in contact with the node is necessary to make sure archiving stays on target for deadlines. It is also important that the node (and subsequently NASA programs) stay informed of any issues, needs, or general problems. The node can provide assistance throughout the project and may be aware of other solutions relating to reporting delays or other issues with HQ.

Working with PDS4: Producing an archive with PDS will be a multi-step process. First and foremost, there will have to be data. Once finalized data products are produced, those products will need to have PDS4-compliant labels. A label contains all descriptive

metadata for the associated data (specifics about the product, pointers to documents and calibrations, etc.). PDS will likely provide base templates, sample products, and/or other instructions to begin the process. In order to do this the node should be provided detailed descriptions of the proposed scope of the data products. This may simply be a redacted version of your funded proposal, with iterative contact to plan out the structure of the archive bundle. All labels are dictated by the PDS4 Information Model (IM) and are expressed as XML files (eXtensible Markup Language). Label templates will likely consist of partially filled out XML forms that will need to be completed. Completion could simply be done by hand, or by script depending on the number of files needed. XML files can be edited with any basic text editor, but an XML-aware or XML-specific editor may be the best option. These more specialized editors can make use of XML functionality by allowing in-line syntactical validation based on control documents referenced in the PDS4 labels (namely the schema (.xsd) – IM organization, and schematron (.sch) – expected values sets). PDS ATM node personnel contacts will be available to advise and help throughout the process.

Types of Data for the Archive: Data format requirements for the archive are set for the long-term stability of the archive. Simply put, not all data are created equal, and sometimes what is used as a scientist in a particular field is not appropriate for an archive. Simple formats are usually best (e.g., ASCII and/or CSV tables, uncompressed images, simple arrays). (See PDS's guides on Archiving Formats [5].) The more modern the format, the less likely it is that the standard (if there is one) will meet the PDS archiving requirements. FITS files, for example, are not technically acceptable in the archive unless specific rules limiting the nesting of parts of the files are followed. If formats other than the simple ones are needed the node representative should be contacted to double check that those formats are acceptable, before doing all the work!

Upcoming Tools: Many of the nodes are working on tools for aiding in the creation of aspects of the PDS4 labels to help in creating your archive bundles. ATM is currently developing the Educational Labeling System at Atmospheres (ELSA) that will help teach the organization of PDS4 while allowing users to create tailored templates for their bundle needs. The system will allow users to explore different options, allowing users to auto-fill parts of the labels with referencing that

can otherwise be tricky to do by hand. (*If you are interested in helping to beta-test this software, please contact Lynn Neakras (lnearkas@nmsu.edu) for more information. We are anticipating an online working beta by Q4 of 2020.*)

Completing the Archive: Once all data are successfully labeled and validated by the receiving node, the completed bundle(s) will need to be peer-reviewed. At ATM, that review consists of an online review by PDS personnel and scientists appropriate to the data, followed by a telecon to discuss the results. When this is completed, there will likely be liens on the data archive, meaning there will be errors or edits that need to be completed before the dataset can be certified as part of the PDS archive. This process can be lengthy and many data providers find that they do not budget enough time for the entire review process. The review is important for multiple reasons. First, peer-reviewed data, like a peer-reviewed paper, catches errors and allows the community to help make the dataset more useful for a larger audience. Also like peer-reviewed papers, the PDS is moving to assigning *Digital Object Identifiers* (DOIs). The work and effort put into archiving data can then be referenced just like publications. DOIs will also make the certified data easier to find for future users to use in proposing new work. Peer-reviewed data also strengthens the confidence level of the validity of the data and separates the archive from simple data repositories.

Final Thoughts: Archiving is a responsibility to be taken seriously for all funded projects. Please remember to allot enough time for the nodes to help get through the process. This includes design and preparation of the labels for all products, directory layout of the bundle, all documentation and users' guide(s). Data providers should allow ~3-6 months at the end of the funding period to conduct the peer review and lien resolution steps, before an archive will be complete. Communication with the node is paramount to archiving success, particularly since node personnel are juggling many projects and working with dozens of data providers in various stages of archive completion. We the PDS are here to help archive and access data.

Atmospheres (ATM) is making updates to our web pages to better accommodate a wider range of data that are now becoming more common. This goes beyond our typical mission support. We will be redesigning pages for planetary analog lab and field data, which will greatly enhance the work that can be proposed with both mission and non-mission data. Some of the new data sets that pertain to aeolian processes across the Solar System that we are planning to archive include: historical wind tunnel threshold data, dust devil field observations as terrestrial analogs, wind tunnel

comparison data for multiple planetary analog environments, atmospheric composition laboratory data for hydrocarbons pertaining to gas giants and Titan, and hot-ion atmospheric escape modeling/laboratory work,. With advances in GIS support from the Cartography and Imaging Science Node we plan to convert the Mars Digital Dunes Database into archivable products using PDS4 internal referencing to preserve the current work and make future additions more efficient and usable.

Acknowledgments: This work is supported by NASA through Cooperative Agreement number NNX16AD93A.

References: [1] Research Opportunities in Space and Earth Sciences 2020 (ROSES-2020) website, <https://nspires.nasaprs.com/external/solicitations/summary.do?solId=%7bBCEE336B-D550-CCBA-1C8C-7A866DB06F45%7d&path=&method=init> ; [2] Information for Data Providers website at PDS, <https://pds.nasa.gov/home/proposers/> ; [3] Planetary Data System Standards Reference (V1.13.0), https://pds.nasa.gov/datastandards/documents/sr/current/StdRef_1.13.0.pdf ; [4] Welcome to Archiving at the Atmospheres Node, https://atmos.nmsu.edu/logs/Welcome_to_Archiving_at_ATM_V1.3.pdf ; [5] Archiving Formats at PDS website, policy and guides for CDF, FITS, and MP4, <https://pds.nasa.gov/datastandards/documents/archiving/> .