

ARCHIVING DERIVED DATA WITH THE PDS ATMOSPHERES NODE: THE EDUCATIONAL LABELING SYSTEM FOR ATMOSPHERES (ELSA)

Lynn D.V. Neakrase, Danae Hornung, Nancy Chanover, Lyle Huber, Reta Beebe, Joni Johnson, Kathrine Sweebe, Zena Stevenson. NASA Planetary Data System, Atmospheres Node, New Mexico State University, Las Cruces, NM 88003. lneakras@nmsu.edu

Introduction. New requirements put forth for data management from both research and development data analysis programs and peer-reviewed journals have increased the amount of derived data being submitted to the NASA Planetary Data System (PDS). PDS instituted a new archiving standard, PDS4, which requires users to become familiar with a new archiving process based in eXtensible Markup Language (XML). This new archiving standard was designated for use for all missions launching after November 2011, and this extended to new derived products submitted to PDS after this time. The Atmospheres Node (ATMOS) has endeavored to streamline the data submission process and promote familiarization of the new archiving standard by developing a new online PDS4 tool, *ELSA*. *ELSA* (*Educational Labeling System for Atmospheres*) is designed to enhance the interaction between ATMOS and data providers by streamlining the creation and editing of PDS4 label templates for building a valid PDS4 bundle for eventual submission to the archive.

PDS4 and Derived Data. The Atmospheres Node utilizes pre-made basic PDS4 XML label templates as starting points for common atmospheres products. These can be divided into two main groups: System Labels and Product Labels. Contrary to PDS3, individual labels for data products have little meaning without the connections to the overarching PDS4 Bundle structure. The PDS4 Bundle structure can be roughly described as a dataset, although under the PDS4 Standard the Bundle is often more than a dataset as has been described in previous iterations of the archive standards [1,2].

At the core of this new archiving system is the internal referencing mechanism referred to as the “logical_identifier”. Logical identifiers (LIDs) utilize Uniform Resource Names (URNs) to designate the location of the uniquely labeled product within the PDS Central Registry. Under

the PDS4 Standard, all PDS-based products consist of a 6-segment URN that uniquely identifies the products. PDS4 requires this URN in the format:

```
urn:nasa:pds:<bundle>:<collection>:<product>
```

ELSA. The primary focus of ELSA is the easy manipulation of product label templates for custom tailoring for data providers. It is intended as a starting point for their Bundle creation and ultimate submission of PDS4 compliant data. In this way, ELSA is an educational tool for learning the architecture of PDS4 archives while the user works with the node to create their data archive. ELSA is not an automated tool or “black-box” for creating bundles for submission to PDS, but rather, an aid, that automates parts of the label template tailoring process.

ELSA’s ‘Build-A-Bundle’ tool begins by querying the user for basic information about the bundle they intend to submit. The goal of this system is to create and identify all logical identifiers that will be used throughout the incoming bundle and begin to populate the appropriate base templates for the data provider. Logical identifiers (LIDs) come in two main types across the bundle: (1) internal product references referring to the items within this bundle or other closely related bundles; (2) context references for missions/instruments, targets, or facilities/laboratories. ELSA helps users by providing mechanisms for constructing internal references and selecting appropriate context references.

Once these references are constructed, they are used to populate bundle, collection and data product labels. The user is prompted to work on their Document Collection first to create LIDs for those products necessary for referencing in subsequent collections and data products. Currently ATMOS provides basic templates for documents and data that need to be edited by taking information from the bundle and

collection products. ELSA can handle any pre-made templates by appropriate referencing, but it currently only accommodates document file labels and basic data product (ASCII and binary tables, simple images/arrays, FITS) labels.

The top-down approach within the Build-A-Bundle tool stems from extensive work with data providers over the past 4 years, incorporating early PDS4 mission experience from LADEE and MAVEN, PDS3 migration efforts, and derived data products submitted to the node over the last several years. In this way ATMOS has been beta testing the process through which the node interacts with providers of derived data before the ELSA coding has been completed, ensuring a viable methodology for this streamlined approach.

Current Status. ELSA presently allows users to input and create internal references. These internal references are handled via a PHP/JavaScript framework on an ATMOS internal website for development. Prototyping of the interface to access the PDS Engineering Node context reference catalog is underway. The system successfully produces all the selected collection files and populates the constructed LID URNs for each. Currently the interface allows text editing of the labels within a browser window, and can be saved or downloaded as partially completed templates. These templates can be used offline for pipeline development, in the case where many products must be produced or the provider can opt to further work with the templates outside of the ELSA environment in a text editor of choice. ATMOS is internally beta testing ELSA with several derived data providers currently preparing archives by a combination of ELSA-generated products and hand-filled work.

Future Work and Conclusions. ATMOS hopes that the data providers who will be starting their submission process in 2018 will make extensive use of ELSA for the development of their archive products. ELSA is and will be an interactive environment to facilitate

communication between the node and the data providers, by providing a mechanism for data providers to understand the full scope of the archiving responsibility required by their funding programs.

Although ELSA is currently using a PHP/JavaScript framework to handle web form creation and manipulation, ATMOS is investigating moving some or all of this development to Django (a high-level Python web framework) to make use of previous Python scripts from our migration code development. ATMOS will continue to develop and expand ELSA to handle more seamless integration of the PDS4 Bundle-Collection-Product architecture. Security upgrades at ATMOS will allow ELSA to make use of a username/password log-in system to enable data providers to save progress on their various projects and work on them over multiple online sessions. Further integration of the product labels needs to be implemented to complete the bundle-building experience and allow the seamless population of URN references across the bundle. ATMOS also intends to tie this Build-A-Bundle into a submission system that links into our online review forms for further streamlining of the peer review process.

Current testing and application of ELSA to our incoming projects is already providing a useful tool for the node to interact with our data providers. Continued development will provide increased efficiency in how ATMOS serves both its data providers and end users by allowing manageable, repeatable archiving of derived data both through R&A programs and unfunded projects.

References. [1] *Planetary Data System Standard Reference, V.1.7.0*, (2016) https://pds.nasa.gov/pds4/doc/sr/current/StdRef_1.7.0_160929_clean.pdf; [2] Neakrase et al., (2016), Planetary Data System: Supporting archiving of derived data, 47th LPSC Abstract #2640.