MAPPING THE STARBASE: CONTEXT PRODUCT DATABASE, ELSA BRIDGES GAP

L.D.V. Neakrase, K. Sweebe, T.D. Pagán, J. Karl, L. Huber, N. Chanover, V.I. Martinez, A. Hyder
Planetary Data System, Atmospheres Node, Dept. of Astronomy, P.O.Box 30001, MSC 4500, New Mexico State University, Las Cruces, NM 88003-8001

Introduction. In an effort to aid derived data providers, the PDS Atmospheres Node has developed a new online tool, the Educational Labeling System at Atmospheres (ELSA). ELSA provides an online web environment to build PDS4 bundles for simple data while teaching users about PDS4 structures.

Under the PDS4 archiving standard, datasets are organized by an Information-Model-driven hierarchy:

Bundle → Collection(s) → Basic Products

Every registered product consists of a metadata label and the product itself (data, document, inventory files, context products, etc.). The labels are stand-alone XML files that contain all the necessary information about the companion files, including linkages to related files. All products must have unique identifiers (logical identifiers – LIDs) that are used to register and reference every product under this system. Logical Identifiers are implemented through the use of Uniform Resource Names (URNs) and for PDS will have the form:

urn:nasa:pds:bundle_id:collection_id:product_id

There are formal and informal rules for constructing unique, valid URNs for PDS4. Many PDS documents for data providers outline the rules for these identifiers. New providers or providers with limited experience with PDS4 may need to create these unique references for funded research projects, but may be intimidated by the complexity of the PDS4 archiving standard. ELSA is designed to aid in this task.

Within the PDS system, context references, or products that refer to observing system components like investigations, instrument hosts (spacecraft, facilities), instruments, and targets, are managed by the PDS Engineering Node (EN). This 'Context Bundle' contains all the necessary bits for referencing missions, laboratories, instrument packages, and designates what targets are used in every product.

ELSA aims to allow easy access to the engineering context bundle to provide consistent, reliable referencing within label templates across the bundle. This allows data providers to easily add references to their labels and get the correct LIDs automatically.

PDS4 Context Products and Databases. Context products work as a system of PDS-curated references to aid in the eventual search and retrieval of specific, user-desired products. The system of references form a hierarchy providing the context for archive products tagging them to a specific mission or facility with information about how the data were collected with instruments. The context product system builds the observing system necessary for collecting data and has a hierarchy that looks like:

Investigation → Host(s) → Instrument(s) → Target(s)

Investigation products provide the framework for a set of data and fall into multiple categories: *missions*, *observing* and *field campaigns*, *individual* and *other investigations*. These describe the top-level for the type of science being done, information including references to the instrument host or facility, instruments, and targets.

Host products form the top level providing information about the spacecraft or facilities where instruments are located. *Instrument Hosts* typically refer to spacecraft but could also include a range of mobile platforms (rovers, landers, aircraft, balloons, etc.). *Facilities* tend to include laboratories and observatories. Host products provide a '*Where*' to the science being done.

Instruments are the devices that are used to collect the data (spectrometers, accelerometers, cameras, meteorology packages, etc.) and can also include telescopes (which can be both instrument and instrument host where telescopes belong to an observatory and can have instruments mounted to them). Portable devices can also be referenced as instruments (portable telescopes, wind sensors, seismic arrays, etc.). Instrument products provide the '*How*' the data are collected.

Targets provide the objects of the study and can range from stars to planets to comets/asteroids to collected samples and synthetic samples and even laboratory analog conditions. Target products provide the '*For What*' purpose of the science.

Together this system of products provides the ***Context*** of the files associated with a particular PDS4 archive bundle.

ELSA, Python, and Django. ELSA initially began as a student project developing with Java and PHP. Continued student activity required an easier way to maintain the code. The decision was made to leverage our previous coding experience for PDS3-to-PDS4 migration efforts with Python to expand our capabilities in aiding data providers with label support. A natural extension of this was to incorporate Python's Django libraries to create a web-interface for creating and editing label templates. The goal for ELSA is therefore to provide a way for novice and expert data providers to quickly create viable label templates that can be populated with the provider's own code (or by hand).

ELSA and Starbase. ELSA implements a population script to crawl the PDS EN Context Bundle located in the Starbase Backup Repository to find and store information about context products. ELSA performs a linear sweep of each of the following Observing System Components: 1) investigations, 2) instrument hosts, 3) instruments, 4) targets, and 5) facilities. For each product, ELSA uses the current naming scheme of the hyperlinks listed in the Starbase Backup Repository of each product to find pertinent information for database population. The current naming scheme of the links/filenames to each product follows the following scheme:

observing_system_component_type.observing_system_component_name_versionid.xml

ELSA checks if the product exists by name. If the product does exist, the version id of the latest product in the repository is checked against the version id of the pre-existing product in the database. If the current product has a higher version id, then it is newer than the pre-existing product, so the pre-existing product object is updated to represent the most recent version of the product. If the product does not exist, a new object is created in ELSA's database that is representative of the most recent product in the Context Bundle. ELSA then crawls the individual context product's label for its LID and its internal references. The internal references are used to relate other objects in ELSA's database for faster search of products during the bundle building process.

As a user selects context references for their bundle, the user can filter their search by host type (mission/instrument host vs. facility/laboratory-observatory). If a user selects a mission, the user is prompted to select from a list of member instrument hosts. Selections are added to the correct locations in the user's bundle. Once a user selects the instrument host, the user is given a list of instruments and targets that are filtered by relation to the instrument host, while populating instrument host information in the user's bundle. A similar filtering process is also followed when a user selects facility rather than mission. This method of filtration provides a shorter list of observing system components at each step providing easier readability for the user and faster search time for ELSA.

ELSA's population script has been able to find inconsistencies in the PDS EN Context Bundle. ELSA uses hierarchal relations like mission to instrument host and instrument host to instrument to relate objects. Problematic objects have been found when a lower order product is related to a higher order product that does not exist within the PDS EN Context Bundle. Two examples found within the PDS EN Context Bundle occur in the Giotto spacecraft instrument host and in the Canon EOS Digital Rebel XT instrument products. The Giotto spacecraft instrument host points to a mission called 'Support Archives' and the Canon EOS Digital Rebel XT points to an instrument host called 'Unk', both of which do not exist in the PDS EN Context Bundle. To get around these inconsistencies, ELSA currently manually adds the missing product objects to provide a path for a user to reach these lower order products.

Conclusions and Future Work. ELSA, in trying to provide a clean approach for data providers to auto-populate label templates with consistent correct context references, has also provided a tool that is good at identifying inconsistencies within the PDS EN Context Bundle. Incomplete, incorrect, sparsely populated context products do not satisfy the primary purpose of their existence – uniquely, consistently identifying the context of every product for search and retrieval. Current efforts with ELSA have already led to changes in the context library at EN. Continued work will not only improve the quality and usability of the products within the Starbase repository, but also ensure more consistent auto-population tools like ELSA and any others in development.