Archiving with PDS Atmospheres: Educational Labeling System for Atmospheres (ELSA)

Lynn D.V. Neakrase, Danae Hornung, Nancy Chanover, Lyle Huber, Reta Beebe, Joni Johnson, Kathrine Sweebe, Zena Stevenson

PDS ATMOSPHERES

ELSA

NM STATE UNIVERSITY

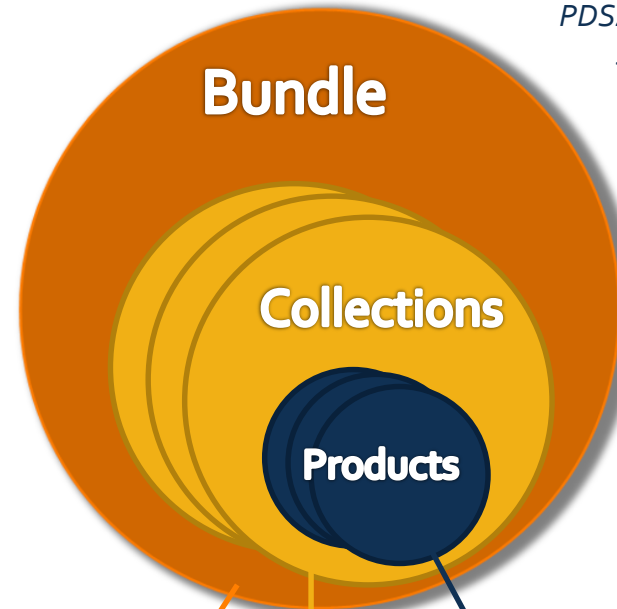# Introduction

- Archiving under PDS4 is a new requirement for data analysis programs and few have experience working with this new standard

- PDS4 moves away from the use of ODL for label metadata by employing XML, a more widely available language for metadata management

- PDS Atmospheres Node is developing an educational labeling tool to aid with the transition to PDS4 by providing an interface for data providers to learn PDS4 while building their archive products for submission to PDS

PDS4 Archiving Standard

- **+** Bundles, Collections, Basic Products

- **+** Everything is a product, all products have XML labels

- **+** Logical Identifiers (URNs) link everything together

- **+** All labels have the same organized layout

PDS4 Hierarchical Structure

Bundle

Collections

Products

urn:nasa:pds:*bundle_id*:*collection_id*:*product_id*

Logical Identifier (URN) Example

PDS4 Archiving Process

- **+** Communication between the data provider and the Discipline Node begins during the proposal process and continues through final validation and certification
  - ▪ We are not adversaries, we're here to help in the process!

- **+** Iteratively design product labels for all aspects of the bundle – work from both node and data provider
  - ▪ Submitting data samples for 'spot-checks' and syntax validation

- **+** Submit Bundle to node for review
  - ▪ Bundle, Collections, Products including documentation!

- **+** Peer Review & Lien Resolution

- **+** Final Validation and Certification

Data Provider Motivation

+ Communication with data providers in preparation for PDART and DAP programs has driven our philosophy regarding submitting derived data to ATMOS

  ▪ Most providers have little or no experience submitting data to the PDS (PDS3 or PDS4)

  ▪ Many may only have a limited number of projects submitting data to ATMOS

  ▪ Very few have much experience working with XML files

  ▪ Providers plan budgets to archive data but may underestimate the time needed to learn everything about XML/PDS4

  ▪ Most providers are looking for the quick clean solution providing maximum output for minimal effort

Provider Feedback & Experience

+ Labeling for data providers is based largely on templates with specific fields to be completed

+ Tailoring templates with the node can provide more detailed metadata and better labels with a mindset on searchability
  - Iterative label design work with node is needed for more complex projects

+ Most common issue is knowing how to fill in all the inter-connected fields (logical identifiers, relationships, etc.) and understanding the PDS4 hierarchical structure

+ *ATMOS is using this feedback and experience to create a tool for providers that provides easy access to PDS4 and how it works – providing a line of communication and support from the node*

Educational Labeling System for Atmospheres (ELSA)

+ ELSA is being designed as a multi-purpose tool environment at ATMOS to ease the burden of archiving

+ The **goal** of ELSA is to provide an online environment for providers to _learn more about PDS4_, _build PDS4 bundles using label templates_, _submit completed bundles to ATMOS_, and _manage the reviews/lien resolutions_

Educational Labeling System for Atmospheres

PDS4 Label Tutorial
The **Tutorial System** provides examples of how to build label templates and instructs users in PDS4 Bundle hierarchical structures and organization by allowing the users to walk through simple products using the Build-A-Bundle system.
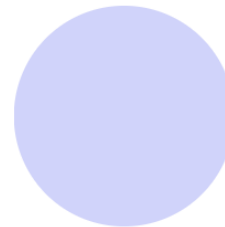
Build-A-Bundle
Build-A-Bundle System serves as a starting point for designing and setting up PDS4 label templates for all parts of the archive bundle.

Data Submission
The **Data Submission Tool** allows secure transfer of completed Bundles to the PDS. Submitted data can then be queued up for peer review.

Online Peer Review
The **Online Peer Review System** is an online review form for focusing reviewers' efforts and efficiently querying reviewers for pertinent information regarding the data under review.

Screen-capture of the front page of the ELSA environment. Currently ATMOS is testing the Build-A-Bundle tool with incoming PDART submissions

ELSA – Label Tutorial

+ Currently the Label Tutorial contains links to documents and other resources for learning about PDS4

    - Many PDS resources based on our work on LADEE, MAVEN, and PDS3-Migration work – and work with Data Analysis projects

+ ATMOS intends to have a walkthrough tutorial of the Build-A-Bundle tool linked to this page to demonstrate the basic functionality of how templates are produced and used

Label Tutorial Tool

- Resources for PDS4
- Walkthrough of Build-A-Bundle

ATMOS is integrating more and more of their systems to aid in the educational approach to learning about PDS4

ELSA – Build-A-Bundle

+ Currently the Build-A-Bundle queries users for the name of their bundle and allows many common collections to be selected

+ The system then builds the LID (URNs) and populates the templates

+ ATMOS is designing a way to browse context products (missions/instruments/targets) for easy template population

Build-A-Bundle

- Label Template design tool
- Auto-population of LIDs and basic set up of all label templates for entire bundle

Build-A-Bundle will allow providers to edit and populate templates with the goal of being able to generate usable, partially-filled templates.

9

ELSA – Build-A-Bundle

+ Build-A-Bundle will have user/password access (free account) so that projects can be saved, downloaded, managed by the provider/user

+ Eventually product labels will also be added to this system allowing for auto-population of all LIDs from bundle/collection labels

+ Basic products labels can use premade basic templates or tailored (with node help) templates that can be loaded into a user's workspace

+ Currently ELSA's Build-A-Bundle tool is written in PHP/JavaScript but is being converted to Django – to allow more seamless integration of Python code from other node efforts

+ Entire bundle template packages could then be integrated into a Submission tool with the possibility of some rudimentary pipelining tools

ELSA – Data Submission

+ Data Submission will be the area where ATMOS can link to completed portions of a provider's bundle

+ The password protected area will be a 'staging area' for node-driven peer reviews of submitted data

+ Currently this is handled manually within the node as the password security is being developed

Data Submission

- Holding area for completed bundles
- Staging area for starting peer review

Eventually ATMOS will have a password protected area for providers to collate completed pieces of their bundle and prepare for the peer review process.

ELSA – Peer Review

+ ATMOS has designed a short web-form for use during the peer review process
    + Reviewers access web-form and answer a series of ~4 questions, responses are converted to PDF and collated for formal review

+ Integration with Build-A-Bundle and Data Submission will allow auto-generation of a Bundle Review page through which the provider will be able to access and track the review process including lien resolutions

Peer Review

- On-line web-form review site
- Auto-generated provider tracking page

Currently this is only the web-form questionnaire. The intent is to add the tracking page once password system is in place.

Current Status

+ Currently ATMOS is beta testing the process by manually aiding incoming PDART14 projects

+ *ELSA* is being tested at various stages for in-house production of templates that have been delivered to data providers

+ ATMOS is planning on select open beta testing of label template creation by the end of the year (2017)

+ Current functionality includes

  - Bundle, Collection label creation – auto-generate, -populate LIDs

  - Prototyping Context Product selection and population

  - Beginning to look at query-driven editing of basic product templates (focus on documents and basic tables)

  - Online review forms used for reviews, manual collation of summary PDFs

Conclusions

+ ATMOS is producing an environment that eases the burden of archiving data with the PDS while enhancing communication with data providers, and providing a means to learn more about PDS4 and how it works

+ ELSA is an encompassing environment that provides tools for manipulation of templates, with multiple stepping-off points for users to become familiar with PDS4

+ Because of the interconnectivity of the XML-PDS4 references, providers tend to understand PDS4 better if the bundle is tackled in one effort, creating bundle, collection, and product labels in the beginning so that everything that needs to be completed can be understood earlier in the process

References

+ If you are planning a derived data set or have general questions or comments about PDS4, please don't hesitate to contact us:

https://atmos.nmsu.edu/ (Main Atmospheres Page)

Or see me after the talk or any time this week for more contact info.