

**PDS ANNEX AND THE PLANETARY DATA ECOSYSTEM.** T.M. Hare<sup>1</sup>, L. Huber<sup>2</sup>, J. Mafi<sup>3</sup>, L. Neakrase<sup>2</sup>, M. Hartke<sup>1</sup>, J. H. Padams<sup>4</sup>, <sup>1</sup>U.S. Geological Survey, Astrogeology Science Center, Flagstaff, AZ 86001 ([thare@usgs.gov](mailto:thare@usgs.gov)), <sup>2</sup>New Mexico State University, <sup>3</sup>The University of California, Los Angeles, <sup>4</sup>Jet Propulsion Laboratory, Pasadena, CA.

**Introduction:** The Planetary Data System (PDS) was established to archive and host data generated by NASA's planetary missions. While most archived products are delivered by mission instrument teams, smaller research products from individual NASA-funded investigators are also now required to be archived at the PDS or a PDS-equivalent repository. This new requirement has resulted in a significant increase in the variety and complexity of data holdings for the PDS.

**Annex:** To help address this influx of disparate data types, several NASA PDS nodes and the European Space Agency's Planetary Science Archive (PSA) have created the concept of an *Annex* data repository. An "Annex" repository typically holds data which does not meet all PDS rules for standard archives. This can simply happen if the data type does not fit well into the allowable PDS formats or if the data provider does not have the time or funding to meet all PDS archive requirements. For example, over the last ten years, the Cartography and Imaging Sciences Discipline Node Annex (or just Imaging Annex) has focused on archiving derived geospatial products (e.g., GIS-ready global image mosaics) which were difficult to support using typical PDS formats and often generated after a mission archive was completed [1, 2]. Similarly, both the Atmospheres (ATM) Node and the Planetary Plasma Interactions (PPI) Node have initiated their own annex repositories specializing in modeling research for various atmospheric and plasma studies. Generally, these studies also require domain-specific formats.

Currently, the PDS version 4 information model (PDS4) has 3 classes of product types called *observational data*, *supplementary data*, and *documentation* [3]. Generally, most derived data hosted in an Annex would be considered observational. However, for observational data, the PDS4 has very strict formatting rules allowing only four data structures: array, table, parsable byte stream, and encoded byte stream. As described in [3], "[PDS] data structures are tightly

*constrained, particularly for observational data... If data cannot be described in terms of these PDS4 'base' structures, they cannot be archived under PDS4 and must be revised."*

Consequently, much of these data have been accommodated in an *annex*, a concept of significant functional value to the planetary community, but one that presents a challenge to define and constrain over time as planetary data formats, directives, and technologies evolve.

**Annex Evolving:** Since 2016, the Imaging Annex has been considered a PDS-equivalent site (essentially an approved long-term data repository), but now finds itself needing to quickly evolve to meet newer NASA and community requirements defined by the Scientific Information Policy for the Science Mission Directorate SMD Policy Document SPD-41a [4] and recommendations from the Final Report of the Planetary Data Ecosystem (PDE) Independent Review Board [5].

SPD-41a states, "[The] *results of federally funded research and development need to be shared openly in order to maximize the benefit and reach of the information. Data need not only to be archived but also to be curated – that is, the data are assured to have continued accessibility and usability for multiple decades.*" Another stated requirement for SPD-41a includes encouraging FAIR practices [6] which requires a globally unique and persistent identifier (or Digital Object Identifier, DOI). FAIR stands for Findable, Accessible, Interoperable, and Reusable. Fortunately, the PDS was created to fulfill these defined requirements and now mandates the creation of DOIs for its archives. Unfortunately, for individual Annex releases, there is no official mechanism in place to issue PDS-created DOIs. Also, while data held in the Imaging Annex is "findable" using common search tools, it is not registered alongside other PDS archives.

NASA's Planetary Data Ecosystem (PDE) is an umbrella concept for our planetary user

community and supported by various facilities involving the planning, obtaining, analyzing, preserving, and sharing of data from planetary science missions [6]. The PDS plays a significant role in this ecosystem, but the concept goes much further than what can be supported by the PDS alone. From the PDE final report, **Recommendation 41** states, “NASA should establish guidelines for preserving high-level data sets of interest that are **not** appropriate to PDS archiving.” Examples for data which are not appropriate for PDS archives may include specialized data formats honed for a particular task, cloud-based formats and services, or ephemeral data results as the result of specific modeling research.

Thus restated, the idea of an annex data repository is warranted, but due to newer NASA requirements and planetary data ecosystem recommendations, the current implementations will need to evolve.

**Introducing Product\_External Data Type:** The PDS has already initiated plans to meet both SPD-41a requirements and the PDE report recommendations. The goals are to not only provide a method to register these annex holdings alongside other PDS archives but also provide PDS-created DOIs. This will provide a bridge for official PDS archives and PDS/PDE Annex data products to better enable search and support a persistent identifier. The goal is to extend this effort outside the PDS and make this available for other PDE-supported repositories.

As previously stated, most Annex data products fit naturally under the PDS4 *observational data* classification. But due to the limited formatting options allowed under observational data, we will introduce a new PDS4 classification called *product external*. This new structure, Product\_External, is most like the existing PDS4 Product\_Browse structure with some minor metadata additions for which most are already found in observational products. Product\_External will likely not be allowed in official PDS4-compliant archives and thus found only within external data repositories including PDS-supported annex repositories. Like Product\_Browse, Product\_External will not be locked to the strict PDS4 observational data structures (formats), nor will it try to describe every aspect of the data format. For example, a

Product\_Browse image just states that it is using the JPEG format and assumes the user can use existing software to view the image. For Product\_External, there will be format constraints, but with the goal for allowing widely used but domain-specific formats. Example formats likely to be allowed under Product\_External include the Hierarchical Data Format (HDF), Network Common Data Form (NetCDF), Cloud Optimized GeoTIFF (COG), or the open geospatial GeoPackage format. Note that the PDS4 information model allows for limited versions of (CDF) and GeoTIFF [7], but the optional format benefits of compression, internal tiles, and cloud-based streaming optimizations are not allowed in PDS4 archives.

**Summary:** The *annex* concept, to support the broader planetary data ecosystem including domain-specific and optimized cloud-streamable formats, is broadly accepted. The PDS4 Information Model and system of tools and services enables findability and discoverability of data. Thus, by describing annex data sets using the PDS4 Information Model, they can be seamlessly integrated into the existing PDS tools and services. This bridge will allow for annex data products, often not currently allowed in PDS4 archives, to be registered by the PDS. The updates to the PDS4 information model will not hamper the current secure archiving methods for mission data and will allow for more analysis-ready and future-forward research approaches expected by users within a robust planetary data ecosystem.

**References:** [1] Hare et al. (2019) 4th Planetary Data Workshop, abs #2151. [2] Hare, et al. (2015) 2nd Planetary Data Workshop, abs. #7060. [3] Data Design Working Group, PDS4 Concepts, DOI: 10.17189/6h0fng87 [4] <https://science.nasa.gov/researchers/science-data/science-information-policy> [5] [https://science.nasa.gov/science-red/s3fs-public/atoms/files/PDE\\_IRB\\_Final\\_Report.pdf](https://science.nasa.gov/science-red/s3fs-public/atoms/files/PDE_IRB_Final_Report.pdf) [6] <https://www.go-fair.org/fair-principles/> [7] <https://pds.nasa.gov/datastandards/documents/archiving/>.