SEARCH AND RETRIEVAL, WEB DESIGN AND REGISTRY, ROAD TO IMPROVEMENT

L.D.V. Neakrase, R.F. Beebe, L.F. Huber, N. Chanover, J.J. Johnson, I. Trejo
Planetary Data System, Atmospheres Node, Dept. of Astronomy, P.O. Box 30001, MSC 4500, New Mexico State University, Las Cruces, NM 88003-8001

Introduction. Over the past few years PDS has been moving to releasing data exclusively under the new PDS4 archive standard. Part of the draw for moving to PDS4 is increased flexibility, consistency, and interconnectivity of data products and documentation. The result of this upgrade is to improve PDS inter-node connectivity through better referencing and eventually better search and retrieval of desired data by end-users. While PDS continues to move forward with migration of legacy products, testing of various mechanisms for improving search and retrieval is necessary to leverage PDS4 design decisions with usability for end-users.

Web Wizardry. Part of the development of better access to PDS data begins with how it is presented to the end-users. Reviews of users' complaints revealed that a basic problem is that inexperienced users are discipline-centered while the structure of the PDS archive is mission-centered. Facing marketing of the first missions to archive under the new PDS4 standard, LADEE and MAVEN, the Atmospheres node (ATM) began working on web page layouts (together with Small Bodies (SBN) and Planetary Plasma Interactions (PPI)) to better accommodate users looking for data. The pages focused on providing more base information about the purpose and goals of the mission with stepping-off points for more detail as needed. The goal of this effort was to provide first-time or novice users access to information vital for understanding the data while maintaining clean access for experienced users to get to the data as efficiently as possible. The resulting pages provided mission information and images at the top of the pages, with links to individual instrument pages that are organized in a similar way. The instrument pages provide information about the capabilities of the instrument, with clear links to the instrument bundles, document and data collections.

The desire to provide tools to access individual data sets at the file level led to the creation of indices containing information above and beyond what is found in individual label files. Dealing with PDS4 LADEE and MAVEN data and migrated Juno data quickly revealed that effective indices for each instrument are highly individualized and considerable effort and interaction with the instrument teams is required to optimize retrieval. Once generated, these indices provided metadata that allowed the creation of tailored search engines for an individual instrument. For the first time, users could organize the data by science discipline keywords, better facilitating search and retrieval at the product-level, rather than bulk data sets or collections.

Better Indices for Better Search. The key to this effort was the culling of specific information to form the indices. In order to better search a collection, more information is often necessary which can be merged with metadata from the data product labels. Information such as timing of observing sequences, targeting of various features, specific spectral components, or mission derived geometry and instrument pointing (SPICE geometry) can often be discovered within documentation and other ancillary sources from mission personnel.

The scope of necessary information and the difficulty of obtaining needed information vary greatly from mission to mission. For a mission like Juno the orbit is relatively known but the challenge is tracking and documenting variations in data acquisition and intent as the instrument teams learn and modify goals and approaches.

For long-lived complex missions like Cassini, the varying orbit and sequence of targets, combine with operational changes, introduction of new goals, changes in data processing, and shrinking mission budgets provide huge challenges in providing adequate metadata and search indices to allow users to effectively exploit the data.

Cassini and Science Facets. Website design for presenting mission data together with better indexing allows better searches, but science end-users typically want more science facet searches than bulk mission searches. Reorganization of the mission data to represent science facets is necessary once a mission is complete to better find complementary observations by different instruments. Furthermore, reorganization can provide easier comparison between missions and legacy data.

Based on PDS3 data organization, the Cassini team was tasked with reorganizing the data into common science facets to make sense of a very long and complicated mission to the Saturn system. The resulting facets were woven into a Cassini mission system of web pages, built largely on providing multiple access points to better serve the community.

The benefit of the mission team being heavily involved in the process was to incorporate much corporate knowledge from the teams collecting the data during the mission. Detailed files were made to incorporate that knowledge including event timing across all instruments and targets during the lifetime of the mission. These files allowed better comparison of coordinated observations, providing a better sense of the science goals of the mission as a whole. User-guides were commissioned from the instrument teams and were written by team scientists (grads/post-docs) that were using the data, to better represent the usability and processing techniques from a practical standpoint.

Preliminary versions of the Cassini pages are online at:https://atmos.nmsu.edu/data_and_services/atmospheres_data/Cassini/Cassini.html. The team has made a major effort to provide mission and discipline data to the user via extensive web pages. PDS4 migration of the data and integration of data retrieval is yet to come.

Future Work, Leveraging PDS4. The efforts of the Cassini Team represent an attempt to bridge the gap between the PDS responsibility to archive mission data and be a usable repository for future scientists to be able to access use the data. As the PDS4 archiving standard continues to mature more of its capabilities will become apparent. For example, the next logical step would be to leverage the PDS4 Central Registry for better search and retrieval options.

Science facet organization can be accomplished using the concept of '*secondary bundles*' or rather a simple reorganization of the registered data products. Secondary bundles allow new bundles to be registered through the Central Registry that could be fundamentally different groupings of data already present in the archive. These new bundles could preserve the specific indices as new documentation in a document collection and data collections could be based on science facets directly. The new data collections would contain no new data, but rather an inventory file comprised of listings of related products pointing to the previously registered mission data.

The use of the Logical Identifier (LID), also known as a Uniform Resource Name (URN), allows every registered product to be uniquely identified like a serial number. Better leveraging of the registry could allow direct linking to the latest documentation on web pages based on these LIDs, meaning that multiple copies of document files would not have to exist in multiple places. Leveraging PDS4 in this way should enable better file management and reduce the possibility of finding multiple copies of outdated documents on multiple PDS sites.

The real strengths of the PDS4 system are just now beginning to become apparent. As more legacy data are migrated and further stabilization of the PDS4 Information Model continues, better web designs making use of the new architecture should be possible. The end product should be a better-integrated, more-consistent archive, that better serves the planetary science community. The evolution of web designs like the Cassini model, will pave to road to improving the ability to search and retrieve data at the PDS.