PDS Data Services Background

PDS Data Services Workshop

Phoenix, Nevada

Nov 5, 2019

Outline

- Background
- Current PDS Architecture
- Future PDS Architecture
- Data Services Requirements Discussion
- Other Considerations
- Next steps

Action Items

- March 2019 Face-to-Face: White Paper
  - "The decision was made by the group to proceed with writing a white paper that states the detailed path forward for PDS regarding our online presence. This will be of use for designing the CMU (architecture) study or will replace the CMU study as our (architectural) "roadmap" if the CMU study cannot be funded."
  - Dan C. will take leadership role in writing the white paper.
  - Tim M. will lead Project Office management side with white paper.

- August 2019 Face-to-Face
  - Action to hold a detailed workshop – goals already discussed.

White Paper Status and Forward Path

- White paper has been updated to v4 and posted to the EN site under data services.
- The forward path is to develop an implementation plan and architecture for the vision

PDS Federated Architecture Today

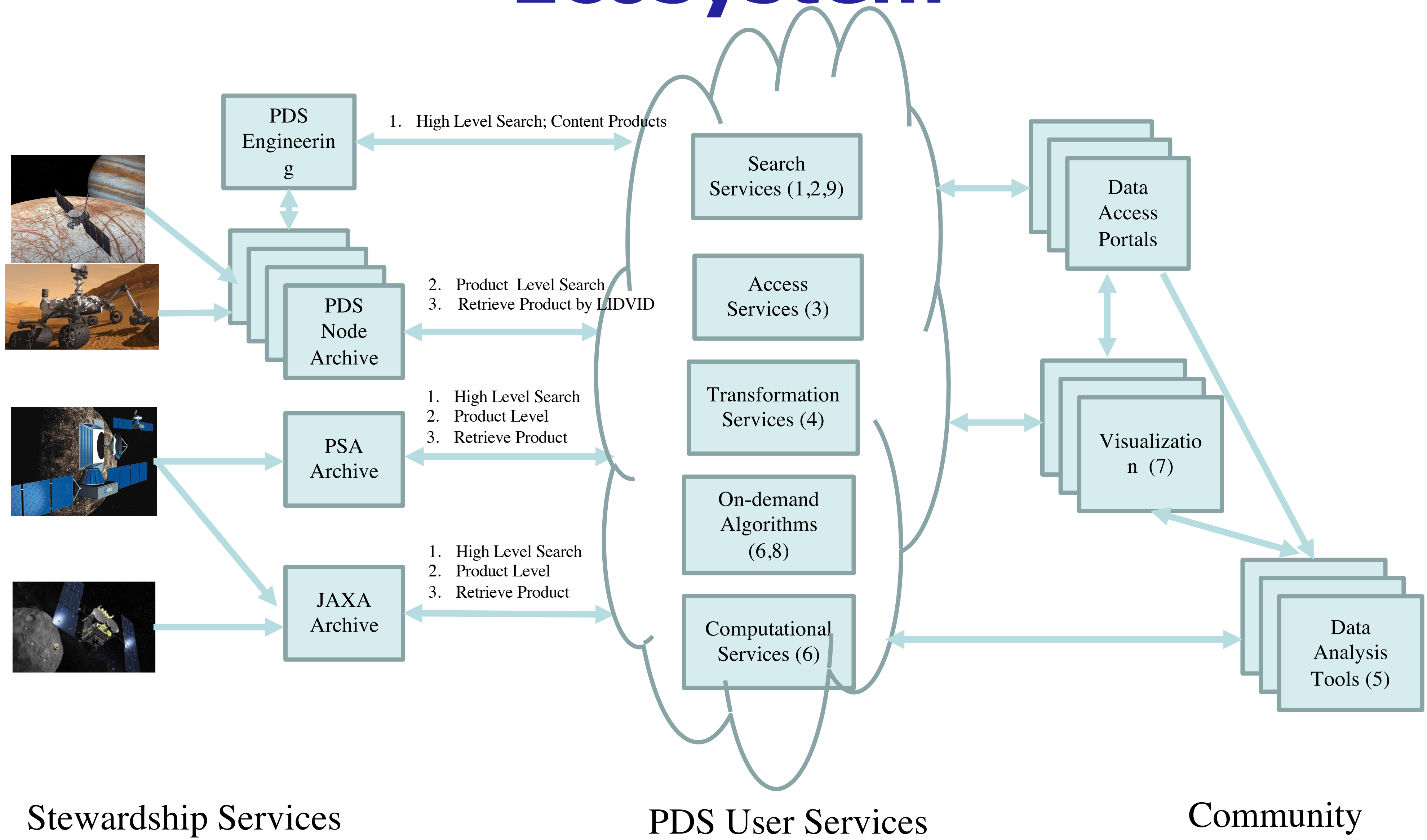
- A robust information model and process for archiving and stewarding data.
  - *This is a critical foundation for international discovery and use!*
  - *It is the de facto standard world-wide which is a huge opportunity*

- A multi-level search implementation following a system-of-systems approach
  - Users can get to the data
  - We have some great node services
  - We can link internationally

PDS Architecture Today: Key Challenges

- Inconsistency in common protocols/standards around data discovery across PDS and the archive community systems: APIs; linking between search engines
- Inconsistent metadata (PDS3); differing use of parameters (PDS3, PDS4, other) for search/query
- Adoption of differing search methodologies and technologies (*this is a fast moving target*)
- End-to-end "search chain" and its impact on user interface/user experience (UI/UX)
- Lack of explicit, consistent, and sharable web services for sharing PDS data and common functions ("data services") across the federation
- Differing approaches to navigating mission support and data access pages

Towards a Planetary Data Ecosystem



Enabling Data Services Functional Capabilities

1. High Level Search across the federation and IPDA
2. Product-level Search
   a. Within an archive (comprehensive)
   b. Across federated archives for different scenarios*
3. Retrieve products from an archive
4. Transform a PDS4 product
5. Integrate modern data analysis tools
6. Compute on PDS4 products
7. Interactive visualization for PDS4 data products
8. Metadata extraction using ML
9. Indexing on dynamic metadata

It's understood that little common metadata exists for every PDS product (e.g, LIDVID, LID, Time, Instrument, Target, etc)

State of Data Services Mapping Table

Capability	IPDA an PDS Support
1a. High Level Search within an Archive (PSA, PDS, etc)	PDAP; EPN-TAP; PDS Search Service API; PDS and PSA supporting services interfaces.
1b. High Level Search across archives (PSA, PDS, etc.)	PDS Search Service API; Search integration (e.g., passing search parameters) inconsistently implemented.
2a. Product Level Search within a Site	Differing support for REST-based API access
2b. Product Level Search across Sites	No PDS or IPDA-wide product-level search;
3. Retrieve Product	Most archives/node provide HTTP access for download. Inconsistent services for download (label, data).
4. Transform	PDS library exists but no service
5. Tool Integration	Integration with tools such as ISIS; no support for Juptyr notebooks; no support for languages such as R
6. Computational Support	Limited support for on-the-fly processing and running analytical results.
7. Visualization Support	Solar System Treks; VESPA
8. Metadata Extraction	Local node activities
9. Indexing on dynamic metadata	Local node activities

Overall Federated Architectural Concept

- The architecture must handle *variety* and *distributed services* as a core architectural principles.

- With PDS4 now in place, increase focus on integration of the federation and improving open access to data and services so they can be shared across PDS, IPDA and the community.
  - Improve integration and sharing of distributed search services (both EN and DN)
  - Drive consistent protocols for major functions (access, search, transform) as a basis for PDS and the international community to adopt
  - Improve web presence (UI/UX, mission support pages, etc) and search chain by leveraging of shared services and APIs
  - Explicitly publish common services for use by the community

Architecture Strategy: Open Data, Standards Models, Shared APIs

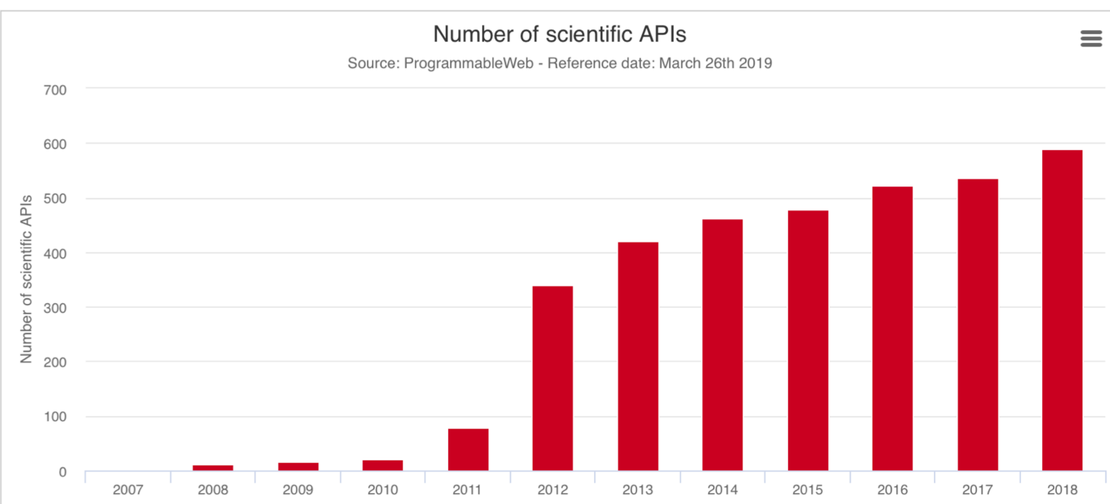
- Open Data – Planetary data online and open for use (*in the DNA of PDS*)

- Standard Data Models – Planetary data described by a consistent model (*PDS4)*

- Shared APIs – Well documented interfaces to get to data and services for use across PDS and by the community *(we need to do more of this!)*
  - *APIs directly driven by the PDS4 information model (e.g., discipline query models)*

Why are APIs important?

An API approach is an architectural approach that revolves around providing a program interface to a set of services to different applications serving different types of consumers. (Wikipedia)



NETWORKWORLD

SOFTWARE QUALITY

By Ole Lensmar, Network World
MAY 28, 2013 03:40 PM PT

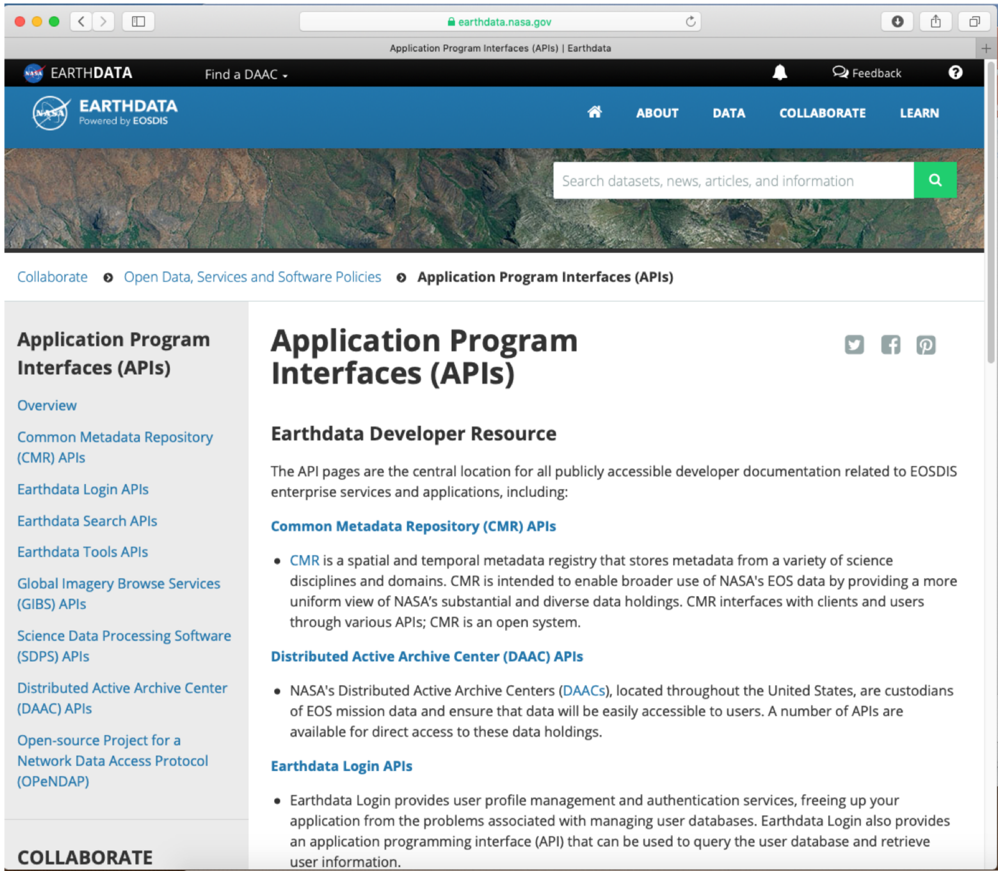
How open data and APIs fuel innovation

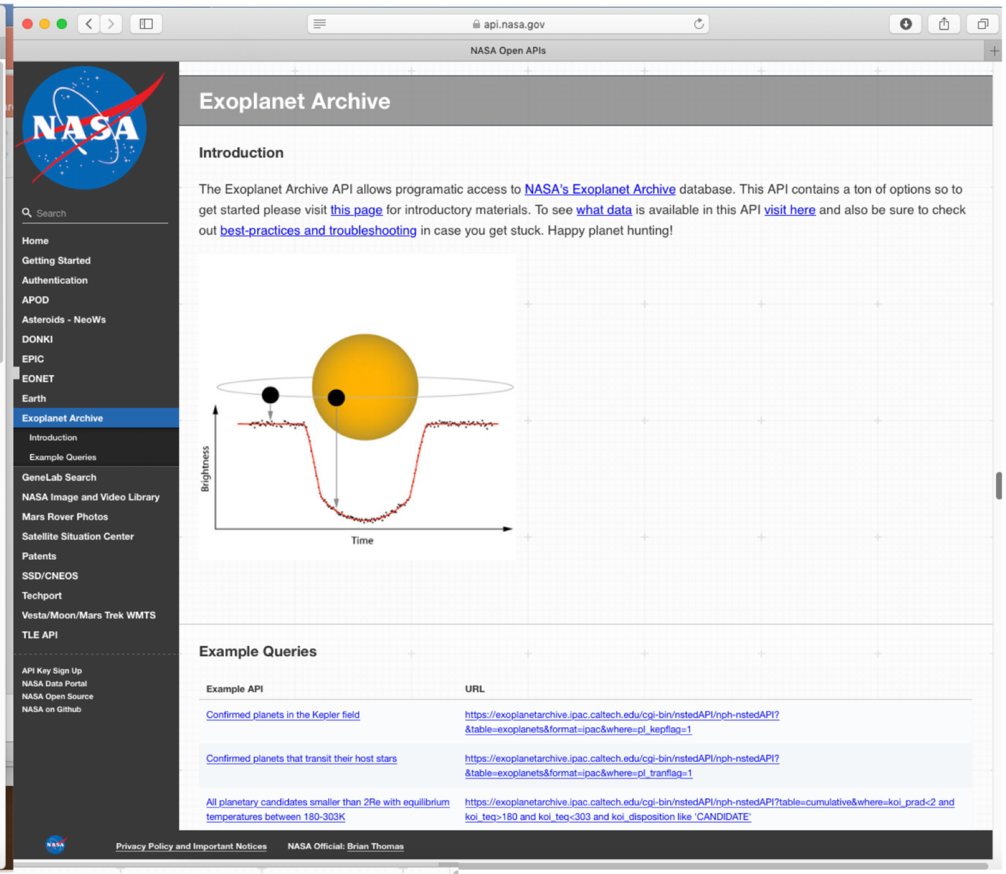
Relating Legos to open data, APIs, and quality software.

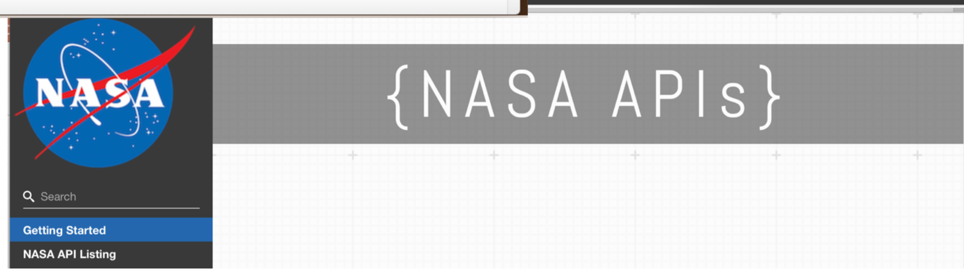
Just like many fellow developers and technology geeks, I was an avid Lego-builder during my youth. Those small plastic pieces gave me the possibility to create anything my imagination had in stock for me – and the cool thing was (and still is), the more basic the blocks were, the more freedom they gave me. Getting older I changed, and so did Lego - I can still sit with my friends complaining about the "dark years" of Lego, when those basic building blocks turned into pre-molded pieces of wings, buildings, or animals, putting a definitive stop to the creative outlet that provided so much joy in our youth. As stated by

European Commission (2019):Data on open collaboration

API Examples @ NASA

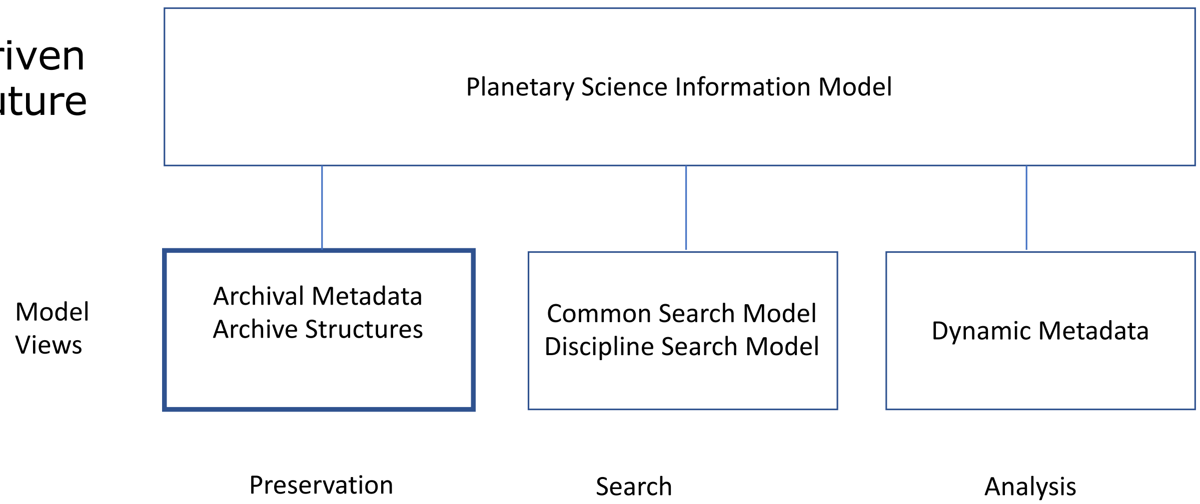






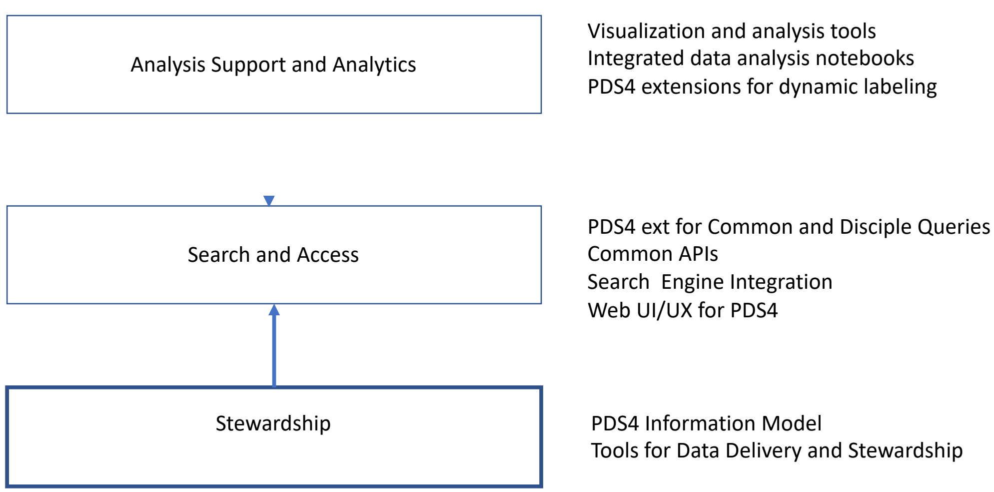
PDS4 Information Model: Drive User Services

- Focuses on stewardship for planetary science data
  - Given data variety, a model-driven architecture is central to the future of a data-driven environment
- Three areas can be supported
  - Stewardship and Management
  - **Discovery and Search**
  - **Analysis**
- Future capability needs
  - Explicit discipline search models
  - Non-archive metadata to label data for machine learning and other discovery approaches



Steps for Unifying the Architecture of the PDS Federation

1. Expand the information model to define common and discipline specific queries.
2. Support dynamic metadata for enhanced product level searching.
3. Support cross-code, cross-product federated searching using modern search engines.
4. Develop and use consistent APIs for searching, access, and retrieving data across systems.
5. Build and register consistent archive support pages to help users boot strap into archives.
6. Develop a next generation web design leveraging PDS4 metadata, data services, and integration of search across PDS to improve user experience.



Data Services Requirements Discussion

- Search
- Access
- Transform
- Computation
- Tool Integration
- Dynamic Indexing

Search

- Goal: *Integrated* search of PSD4-compliant federated archives
  - There is no single search that will ever provide *comprehensive* discovery across all planetary types using a single search string

- Web services PDS can offer:
  1. A consistent API search protocol at both the high level and product level
  2. A set of parameters for forwarding search strings
  3. A centralized index for product level search of common attributes (e.g., LIDVID)

Access

- Goal: Return every product, including both labels and data, from any PDS4 compliant archive

- Web services PDS can offer:
  1. Retrieve base product product based on LIDVID
  2. Retrieve label for a product based on LIDVID
  3. Retrieve related ancillary information

Transform

- Goal: Provide standard library of transformations

- Web services PDS can offer:
  1. Common transformations across all archives for PDS4 data offered as web service

Compute

- Goal: Provide standard processing services (e.g., subsetting, coordinate translation, etc)

- Web services PDS can offer
  1. Geometry services
  2. Subsetting services
  3. Other routine processing as data increases

Tool Integration

- Goal: Support integration with common tools and frameworks

- Web services PDS can offer
  - Jupyter Notebooks for planetary science data
  - Python, R, and other language bindings to search, retrieve, transform, and compute on data
  - Shared visualization tools

Dynamic Indexing

- Goal: Further enhance discovery and use of data through auto labeling using machine learning feature detection and classification methods based on PDS4 model
- Web services PDS can offer
  - Shared ML libraries for labeling including image and other data
- Note: this implies the ability to use dynamic metadata for search as well as user supplied metadata for analysis

Other Planning Considerations

- Community Collaborations
- Driving UI/UX capabilities forward
- The role of cloud computing to enable data services
- PDS role in integrating data analytic capabilities

Community Collaborations

- IPDA – Providing consistent APIs and data services will increase interoperability opportunities with international archives.

- Community – Providing consistent APIs and data services will enable the community to build analysis tools on top of PDS.

UI/UX Considerations

- Improving User Interface/User Experience (UI/UX) and leveraging PDS4 is critical to the future of improving the PDS.
  - This needs to be part of the plan!

- It is important to have a *consistent architectural foundation* on which to build an improved UI/UX web presence

Cloud Computing to Enable Data Services

- Part of the forward plan needs to determine the role of cloud computing for
  - Shared data services (*this is where cloud could provide a big advantage*)
  - Storage (primary, secondary/backup)

- Cloud services provide
  - Standards APIs for access, search tool support, data management, computation, machine learning, etc.

Cost and other trades need to be made as part of the implementation plan

PDS Role in Evolving Data Analytics

- Increasing technical capabilities are opening new opportunities for AI, ML, and visualization.
  - Increasing support in many different science disciplines to apply these techniques to improve how data is used and analyzed.

- Part of enabling an international planetary data ecosystem
  - PDS is leading the way in exploring its role through the Planetary Science Data Analytics and Informatics Conference and Planetary Data Workshop.
  - PDS can enable this but it needs to define how far it should go with its resources and its role.

Next Steps

- Tie off the "As-is" State of PDS Data Services
- Shift to architecture, developing, and deploying PDS data services
  - Following an agile release process (architect, develop, test, release, iterate)
  - Be use case driven
- Ensure Data Services architecture and framework definition is in place for the PPBE and senior review
- More to come in the afternoon…

Discussion

Backup

Definitions

- *Architecture (systems and software)* is "the fundamental organization of a system embodied in its components, their **relationships** to each other, and to the environment, and the principles guiding its design and evolution." (ISO/IEC/IEEE 42010:2011, IEEE-1471)

- We have often used three critical views to describe an architecture: data lifecycle, data architecture (models, dictionaries, etc), technology (services, tools, etc).

- What is also important are the protocols and standards that support the integration of software services and tools for PDS, IPDA, and users.