PDS Data Services Workshop

D. Crichton
T. McClanahan
J. Padams

November 6, 2019

Goals of the workshop

- Develop an overall vision and plan for the PDS moving forward with planning and implementation of an integrated data services architecture and improved user services.
- Provide technical background on the architecture, technologies, and terminology surrounding data services
- Discuss data service capabilities and plans including
  - Evolution and leveraging of the PDS4 Information Model for improving search, access, and user services
  - Development on common software standards including APIs
  - Core services
  - Community engagement
- Discuss existing pilot and node activities in data services to identify requirements including
  - REST and APIs supported by different discipline nodes
  - Linking for search
  - Support for data-driven analytical techniques including machine learning and other methods
  - Leveraging of the PDS4 information model

2

Agenda – November 5, 2019

- Workshop overview and framing
- Technical overview
- Case Study: PO.DAAC
- Node pilots and data services activities
- Technical plans
- Project planning and wrap up

https://docs.google.com/spreadsheets/d/1PeJ5EqU6D69yv_g3Ubj-DRbxexS0y3h8bPJWvsTYEYs/edit#gid=0
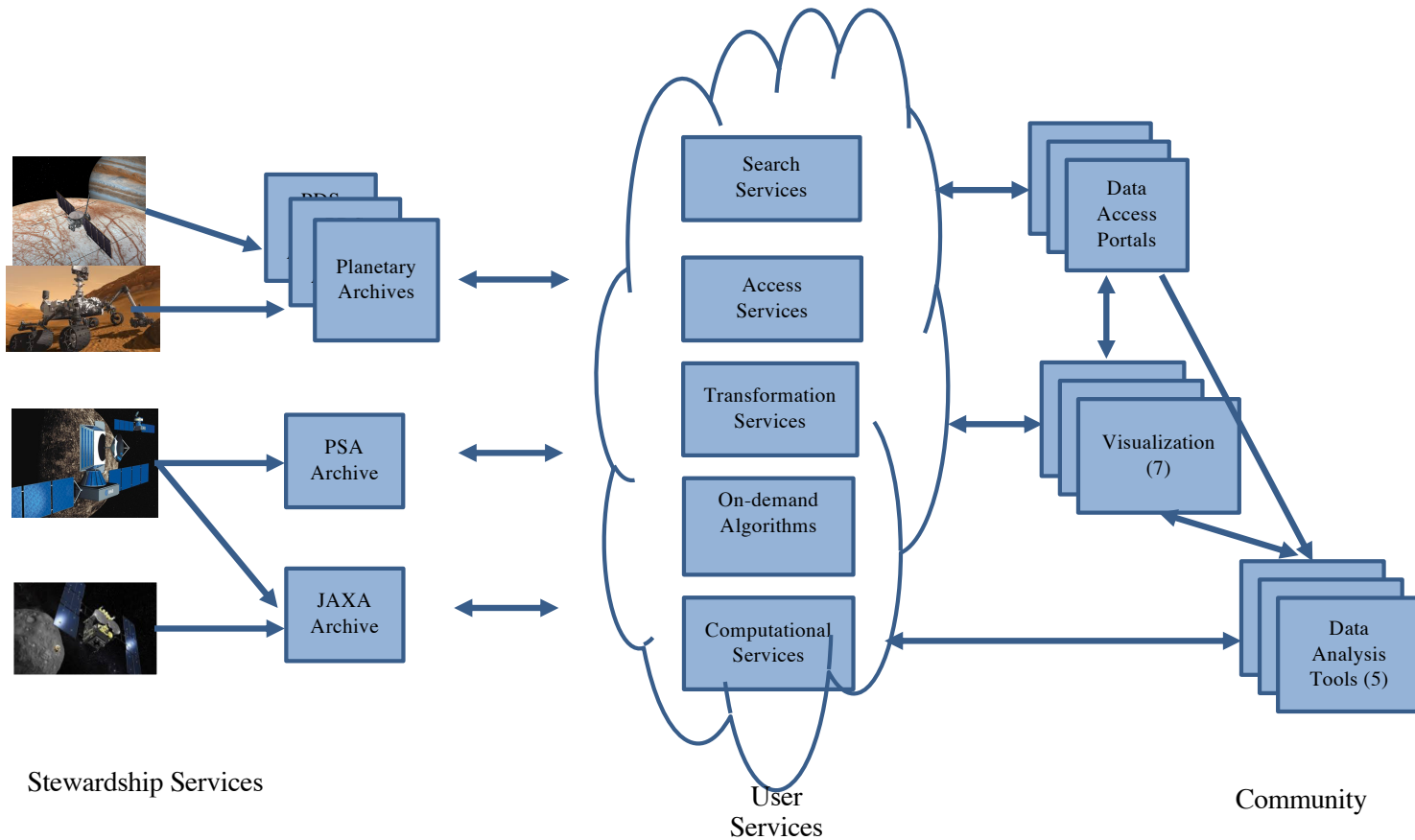
PDS Data Services Vision

Vision: Provide an integrated world-wide data services platform that enables the efficient discovery, dissemination, use and analysis of internationally sponsored planetary science archives

We accomplish this by:
- Capturing and curating archival data and metadata to enable both discovery and use by the community
- Providing a worldwide planetary science data portal as a gateway to archival data and services across PDS, IPDA and the broad planetary science community
- Providing consistent APIs for sharing archival data and services across PDS, among planetary archives, and within the planetary science community
- Supporting a federated cross-node, cross-agency search that enables users to get as close to the archived data and services as possible
- Enabling the on-going exploration and integration of modern tools and access methods to enable data discovery and analysis from visualization to mining of archival data

A Planetary Data Services Ecosystem

Search Services

Access Services

Transformation Services

On-demand Algorithms

Computational Services

PDS

Planetary Archives

PSA Archive

JAXA Archive

Data Access Portals

Visualization (7)

Data Analysis Tools (5)

Stewardship Services

User Services

Community

International * Community Engaged * Sustainable * Cross-Disciplinary

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.
  - Extend PDS4 information model and process to capture search data models
  - 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

Moving Forward

- Propose that Management Council vote on vision statement in executive session

- Begin early efforts to launch a data services project
  - Make progress to support 2020 PPBE Process and Senior Review

- Use next MC F2F as a driver (May 2020)

May 2020 Target

- DDWG begin planning for expansion of information model
  - Capture common and discipline query (search) models
  - Support separation of archive metadata from search metadata
- Proposed changes to PDS level 1/2/3 requirements
- Initial architecture document (v0.1)
- Initial search API developed (v0.1)
- Application of 3 use cases to demonstrate and validate API
- Initial Project plan and structure including schedule and resource estimates