Defining the Digital Preservation and Curation Research Agenda for the Next Decade

Distributed Architectures

The Planetary Data System (PDS)

Steven Hughes, Dan Crichton,
Chris A. Mattmann

November 7-8, 2005

Jet Propulsion Laboratory
California Institute of Technology
National Aeronautics and Space Administration
Pasadena, CA, USA

PDS Mission and Vision Statement

Jet Propulsion Laboratory
California Institute of Technology

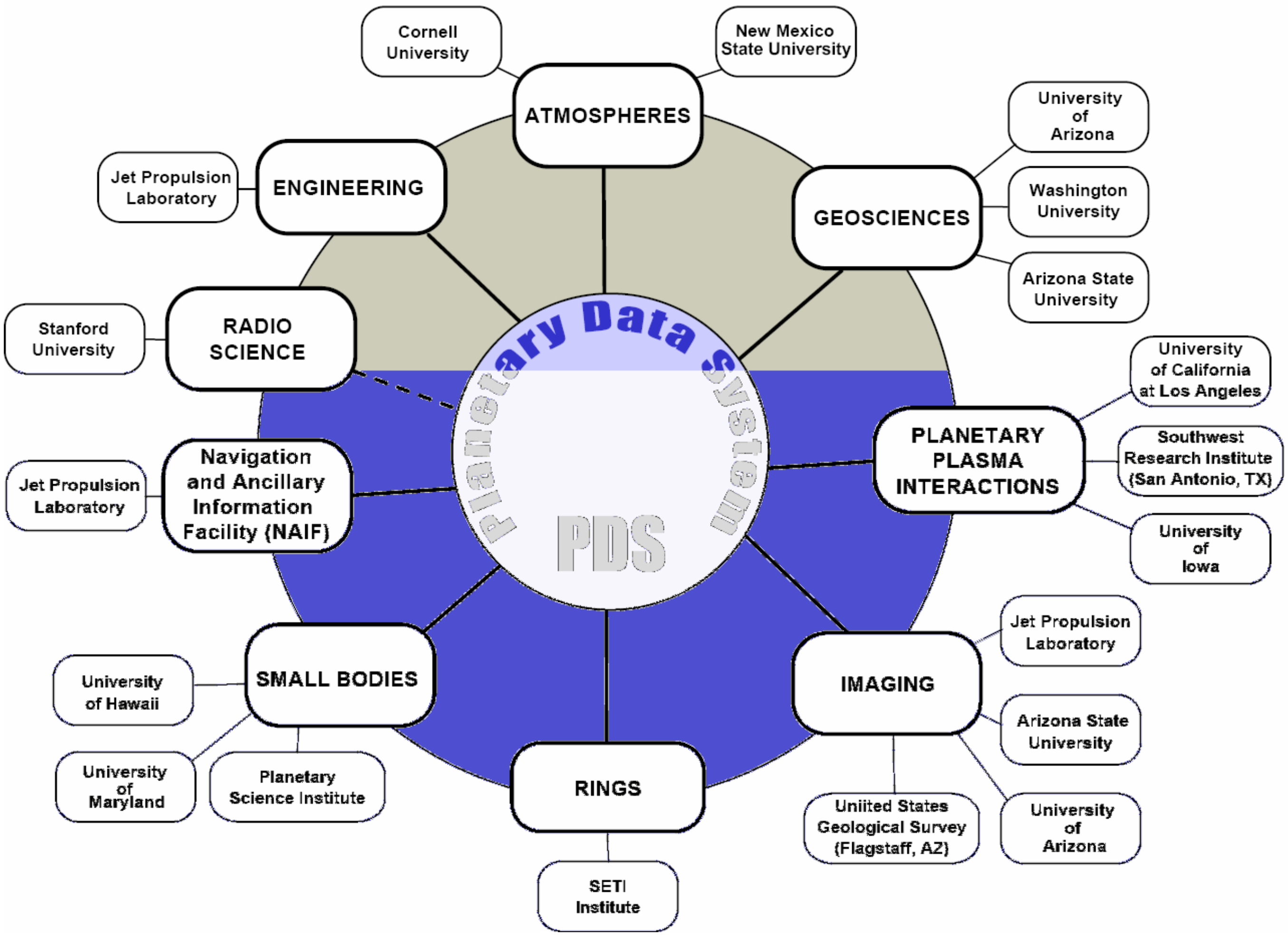
- Mission: The mission of the Planetary Data System is to facilitate achievement of NASA's planetary science goals by efficiently collecting, archiving, and making accessible digital data produced by or relevant to NASA's planetary missions, research programs, and data analysis programs

- Vision:
  - To gather and preserve the data obtained from exploration of the Solar System by the U.S. and other nations
  - To facilitate new and exciting discoveries by providing access to and ensuring usability of those data to the worldwide community
  - To inspire the public through availability and distribution of the body of knowledge reflected in the PDS data collection

- PDS is a federation of heterogeneous nodes including science and support nodes

Planetary Data System
NODES/SUBNODES/DATA NODES

Propulsion Laboratory
California Institute
of Technology

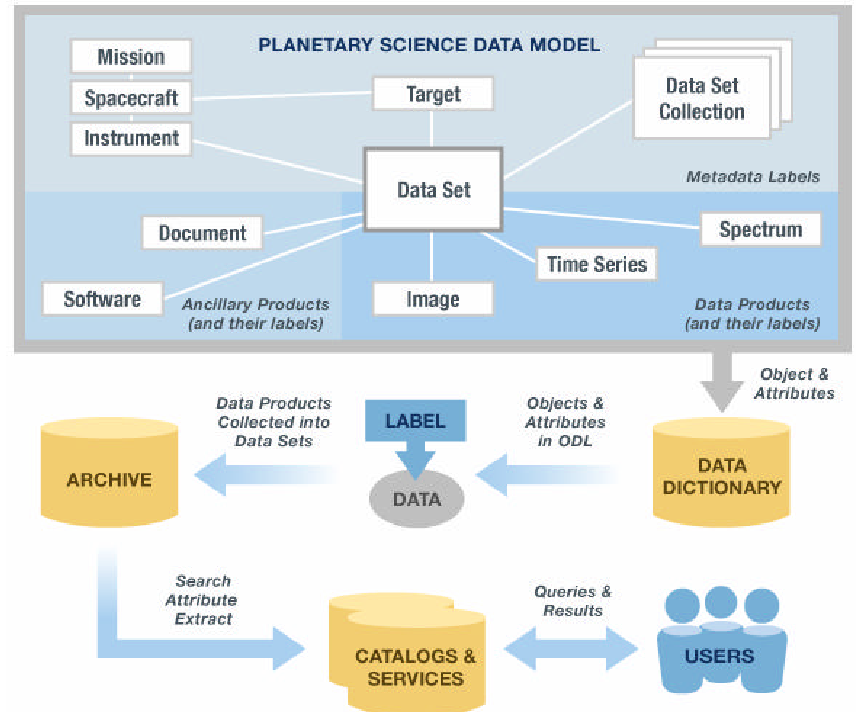


PDS Data Model

Jet Propulsion Laboratory
California Institute
of Technology

- Cross-disciplinary model applied to all missions using the PDS
  - Planetary Science Data Dictionary
    - Data Model
    - Vocabulary
  - Community developed
  - Standard structural specifications for products, data sets, volumes, etc
  - Specification for metadata captured to describe data

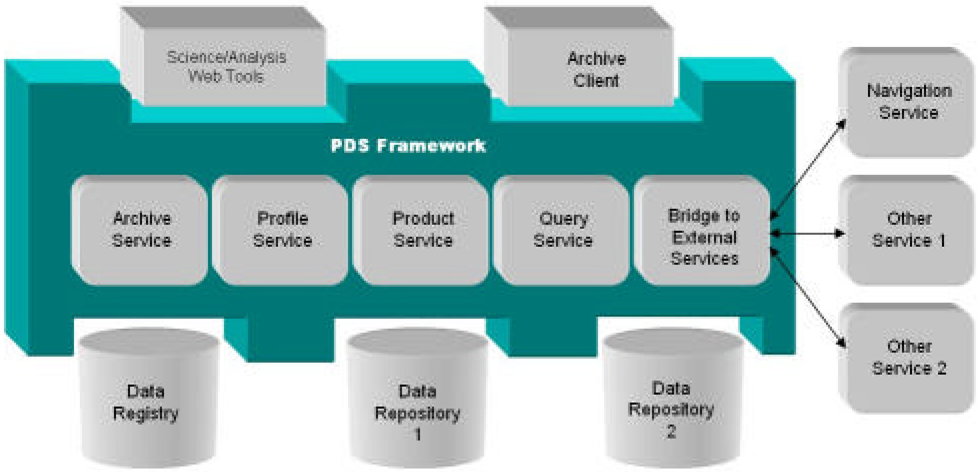
- Technology driven by the model
  - Automated generation and ingestion
  - Index and correlated search
  - International interoperability



Architectural Principles

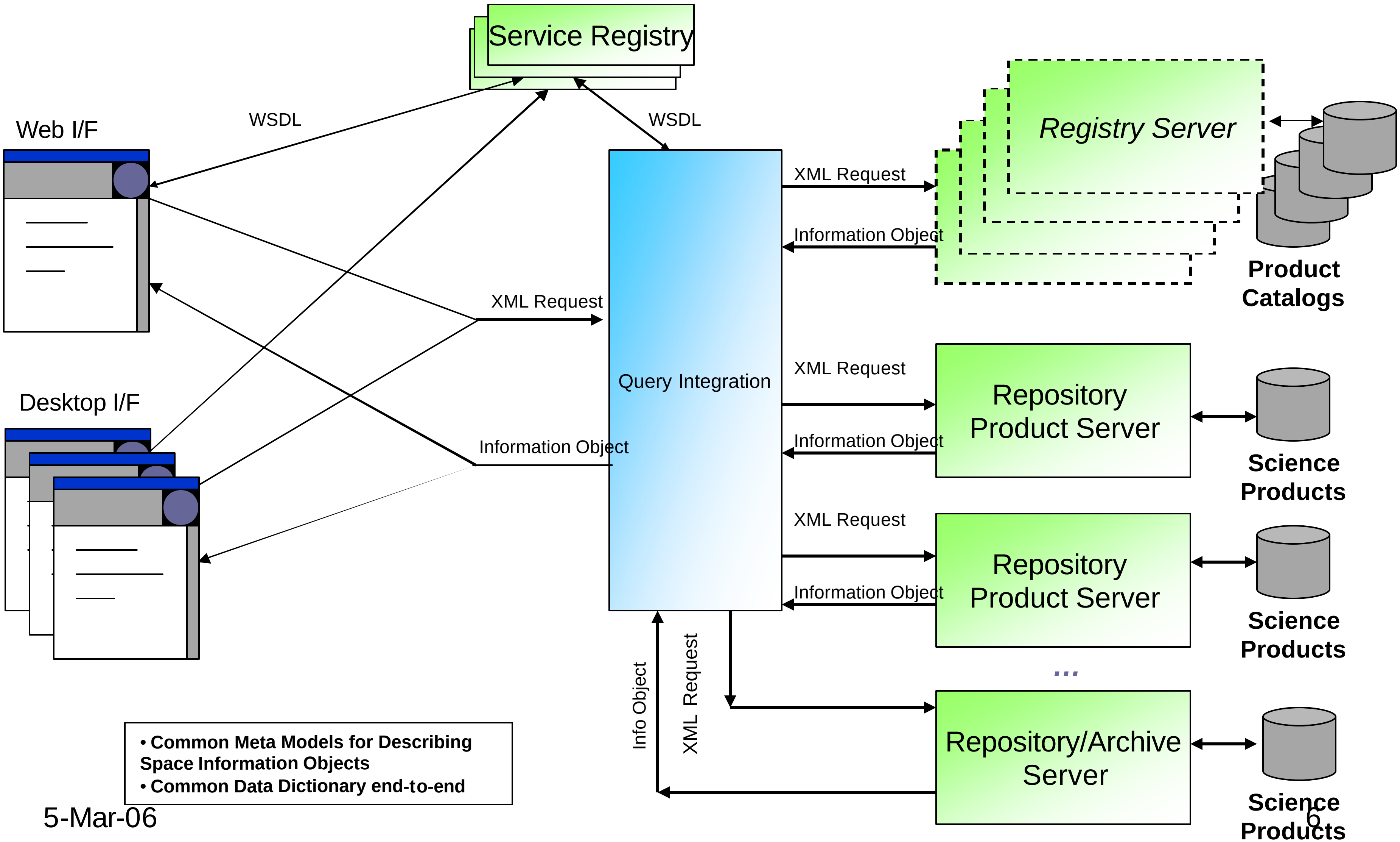
Jet Propulsion Laboratory
California Institute
of Technology

- Define a model for describing systems and their resources
- Separate the technology and the information model
- Require that communication between distributed systems use metadata
- Allow systems using different data dictionaries and metadata implementations to be integrated
- Provide data system location independence
- Encapsulate the messaging layer to support different messaging implementations
- Encapsulate individual data systems to hide uniqueness
- Provide scalability in linking both number of nodes and size of data sets
- Leverage open source



Conceptual Architecture

Jet Propulsion Laboratory
California Institute
of Technology

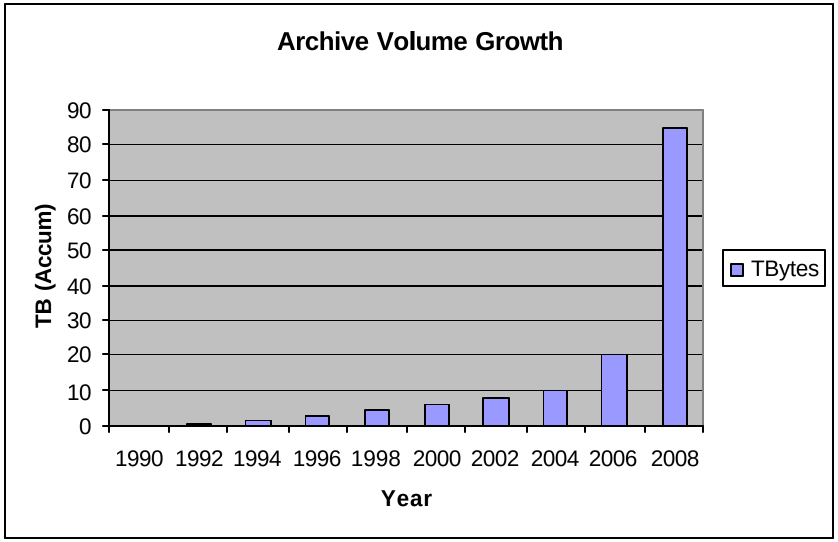


- Common Meta Models for Describing Space Information Objects
- Common Data Dictionary end-to-end

PDS Engineering Challenges
5 Year Timeframe

Jet Propulsion Laboratory
California Institute
of Technology

- Location Independence
- Heterogeneity
  - Science (Data Model)
  - Technology
- Data search across the PDS "Federation"
  - Product Discovery
  - Correlative Search
- Scalability of the PDS Functions
  - Data Volumes
  - Complexity
  - Number of missions
  - Number of sites (E.g., data nodes)
- Electronic Access and Distribution
  - Point-to-Point (Data Movement)
  - Point-to-Many (Distribution to user)
- Coordination of systems and data across the Planetary Science Enterprise
- Aging Technology
- International Interoperability
- Preservation
  - Long term usability of the archived product



Ontology and Data Modeling

Jet Propulsion Laboratory
California Institute
of Technology

- Benefits of an Ontology
  - Capture Conceptual Model
    - Object classes, attributes, and relationships
    - Rich Semantics (Cardinality, inverse relationships, …)
  - Representation of the Conceptual Model (Documentation)
    - Standard and novel notations
      - UML Class Diagrams
      - HTML Documents
      - Dynamic drill down GUIs
    - New plugins under development
  - Export to Notation/Language of Choice
    - RDFS/RDF, OWL, Frames, XML Schema/XML, XMI (UML), Relational Schema
  - Implementation Independence
  - Promotes Interoperability
    - Common tool semantics
    - Common domain semantics easily exploited
    - Inference within model is supported
  - Classification of new elements

Ontology and Data Modeling

Jet Propulsion Laboratory
California Institute
of Technology

- Ontology + Data = Knowledge Base
  - Ingest and validate the data
  - Validate the ontology
  - Equally accessible metadata and data
- Semantic Web Browsers for Metadata/Data Navigation
  - Generated semi-automatically
  - Support Text-, Facet-, and Forms-Based Search
  - Graphical Notations for GUIs under development
  - Allow co-location for searching of differing domain ontologies/knowledge bases

Jet Propulsion Laboratory
California Institute
of Technology

Year Ten Challenges

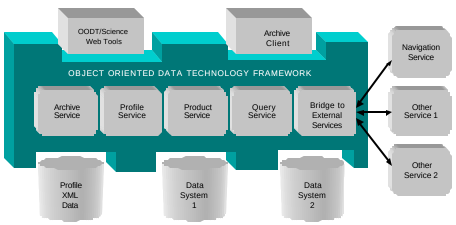
- Data Model Evolution
- Natural Language Queries
- International Interoperability
- Science analysis and science data processing as plugin services
- Intelligent software agents for mundane tasks (conversion/decompression, gathering resources, normalization, geometry calculations)
- Science model integration

Ontologies and the Semantic Web

- "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation." -- Tim Berners-Lee, James Hendler, Ora Lassila, "The Semantic Web", Scientific American, May 2001
  – Web technologies such as HTML, hyperlinks, and http protocol have focused on providing information for human consumption
  – New technologies such as XML, RDF, OWL and software agents will provide information sufficient for computer processing and reasoning
  – The semantic web is dependent on the existence of domain models/ontologies that provide the necessary semantics

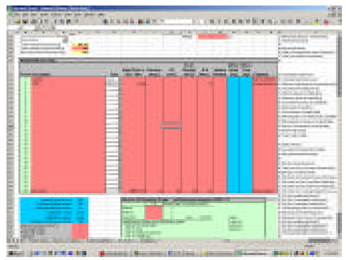
Roadmap
Ten Year Time Frame

Jet Propulsion Laboratory
California Institute of Technology



- Common distributed service infrastructure for sharing science product discovery and distribution
- Single point of entry

Middleware for linking PDS distributed data repositories

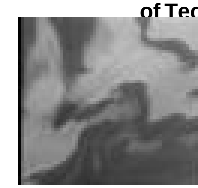


- Integrated tool suite ported to all major operating systems
- Intelligent "Google-like Searching" at product level
- Electronic data movement for large data sets

Tool and Data Sharing Across Planetary Community

- Data analysis and data processing on demand
- Science model integration
- Shared enterprise-quality services (Geometry Engines)

Interoperable and Enterprise Level Services



- International Interoperability
- Value added services and capabilities layered on top of grid infrastructure
- Standard information model used all the way up the pipeline
- Software agents and intelligent processing

Cross-Agency Interoperability

2005 **2009** **2012** **2015**