

A horizontal banner image featuring a sequence of celestial bodies: a blue planet with white clouds, a brown planet, a reddish-brown planet, and a large white planet with a blue ring system. The text "Planetary Data System" is overlaid in white on the right side.

Planetary Data System

# **PDS 2010 System Design**

Management Council Face-to-Face  
August 18-19, 2009

Distributed Infrastructure Design Team

# Agenda for PDS 2010 Discussion

- Introduction and Design Overview
- Data Design

- System Design

- Example
- Mission Infusion Opportunity
- Standards Reference Plans
- Project Plans and Next Steps (Budget, Schedule, Transition and Migration, etc)

# Topics

- Purpose
- Design Team
- Design Decisions
- System Overview
- Ingestion and Distribution Flows
- Transition to Implementation

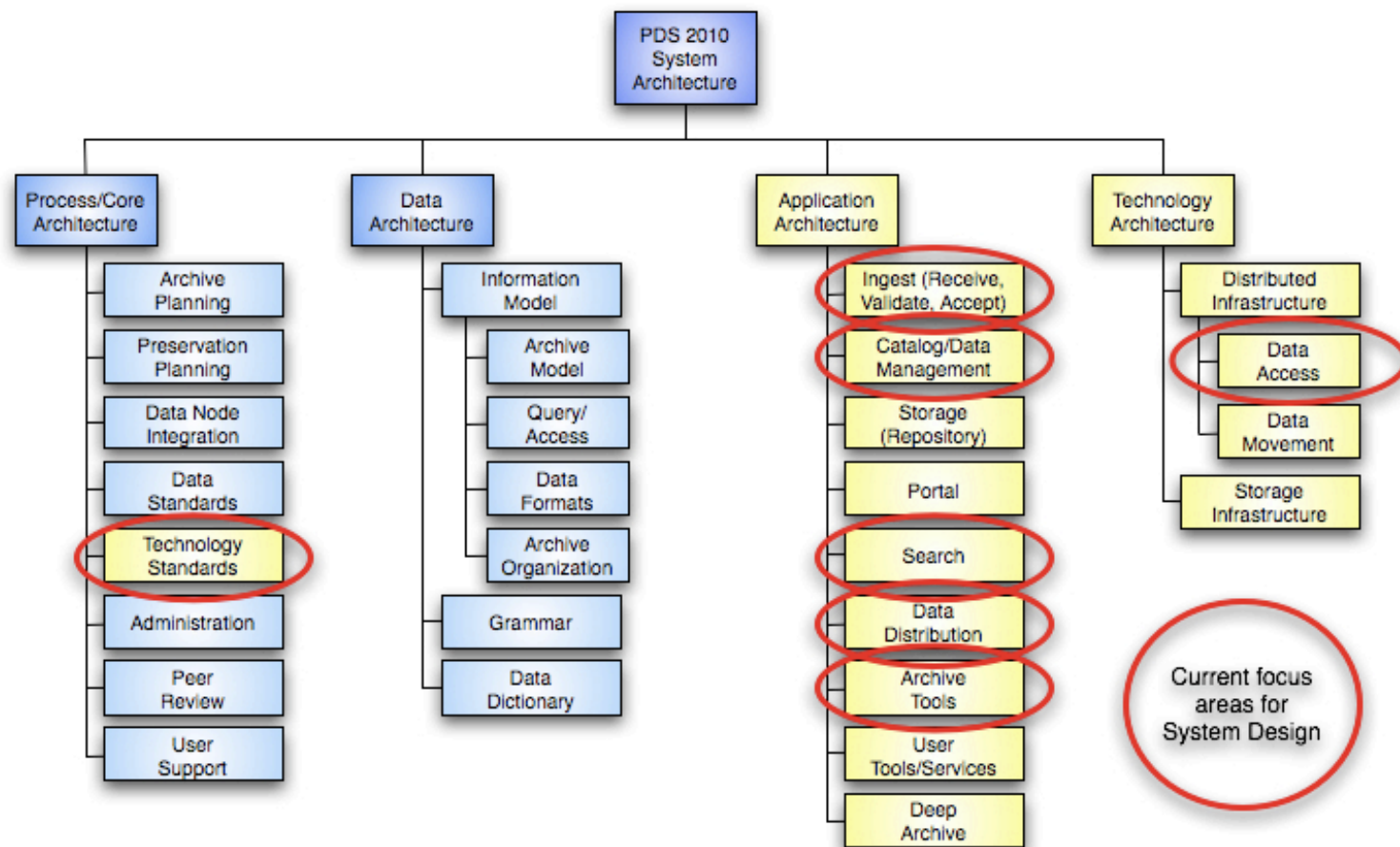
# Purpose

- Discuss major design decisions.
- Provide an overview of the system design.
- Describe the end-to-end flow of data focusing on ingestion and distribution.
- Describe the implementation plan for moving forward with system development.

# Design Team

- Formed the design team back in January, which consists of the following personnel:
  - Sean Hardman (Engineering)
  - Todd King (PPI)
  - Mike Martin (Management)
  - Paul Ramirez (Engineering)
  - Alice Stanboli (Imaging)
  - Tom Stein (Geosciences)
- Weekly teleconferences
- Current artifacts are captured on the PDS Wiki and Engineering Node web sites:
  - <http://oodt.jpl.nasa.gov/wiki/pages/viewpage.action?pageId=2600>
  - <http://pds-engineering.jpl.nasa.gov/index.cfm?pid=100&cid=134>

# System Elements



# **Design Decisions and System Overview**

# Design Decisions

- Service-Based Design
  - Support remote access to data and services to bring the federation together both for ingestion and distribution.
- System of Registries
  - Adopt a system of registries to support improved tracking and access.
- XML Representation
  - Adopt a modern data language/grammar where possible for all tool implementations.



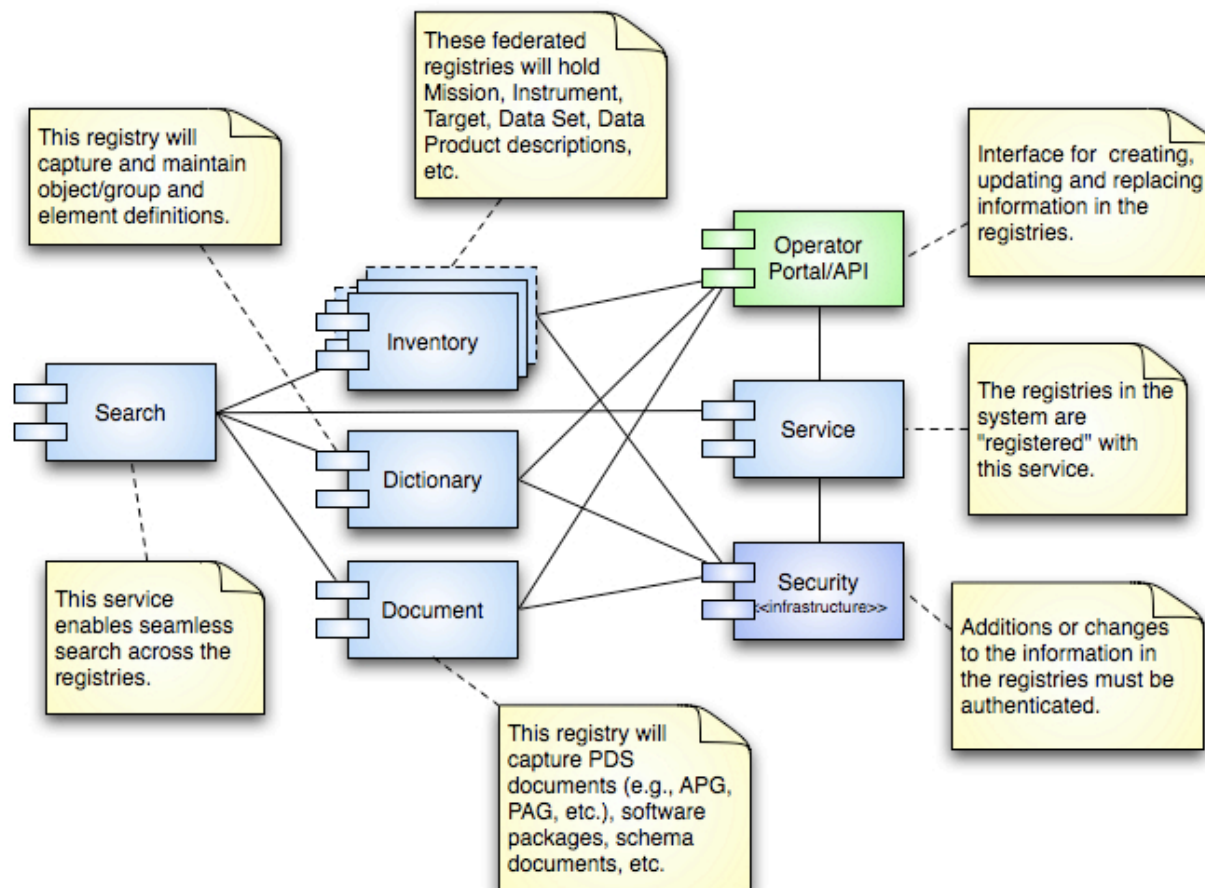
# Service-Based Design

- There are several advantages to adopting a Service-Oriented Architecture (SOA):
  - Captures many of the best practices of previous architectures.
  - Well suited for a distributed system.
  - A service-based architecture provides currency and timeliness for the system.
- Currently developing a SOA solution that suits PDS.
- Service-based functionality will focus on search and retrieval of data.
- A tool-based approach is still appropriate for certain functions.

# System of Registries

- A registry provides services for sharing content and metadata.
- A federated registry allows cooperating registries to appear and act as a single virtual registry.
  - Provides seamless information integration and sharing
  - Preserves local governance
- A query into the federation returns results from all cooperating registries.

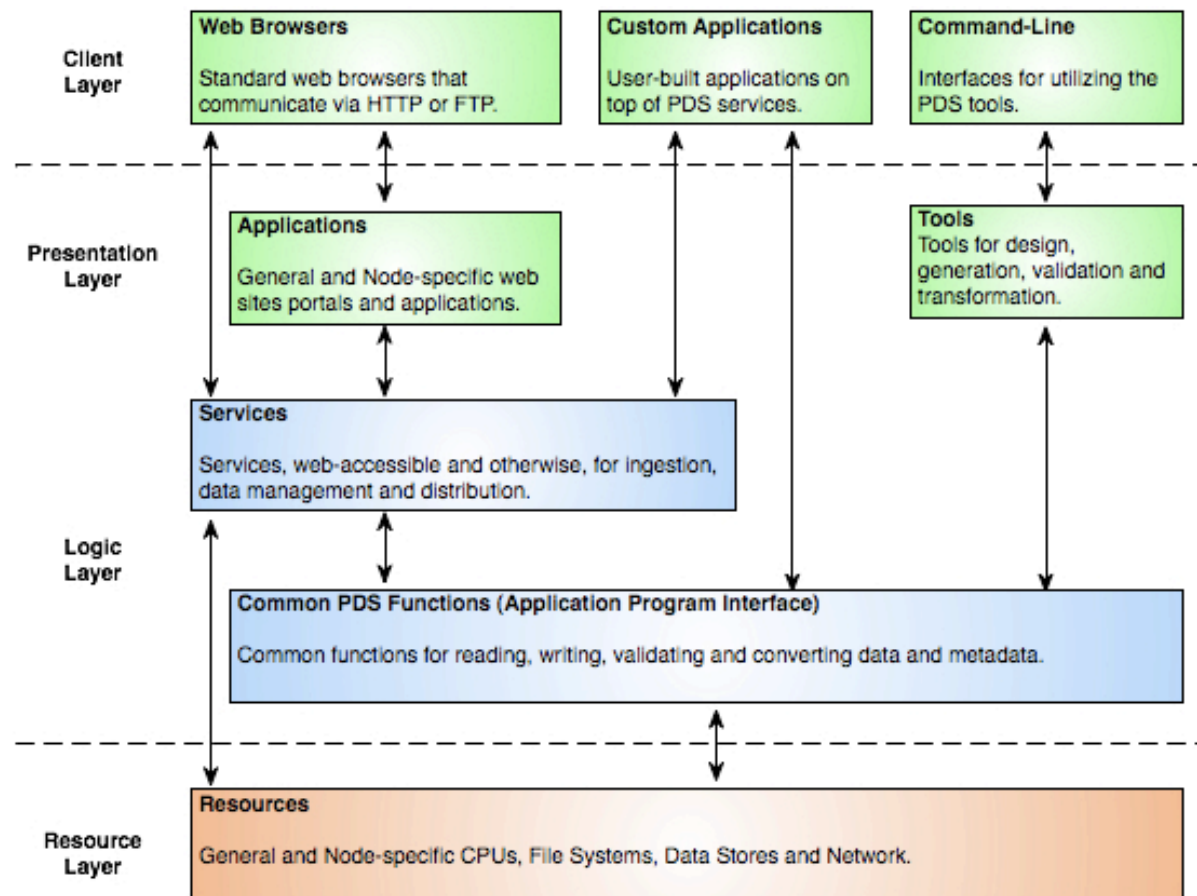
# PDS 2010 Registries



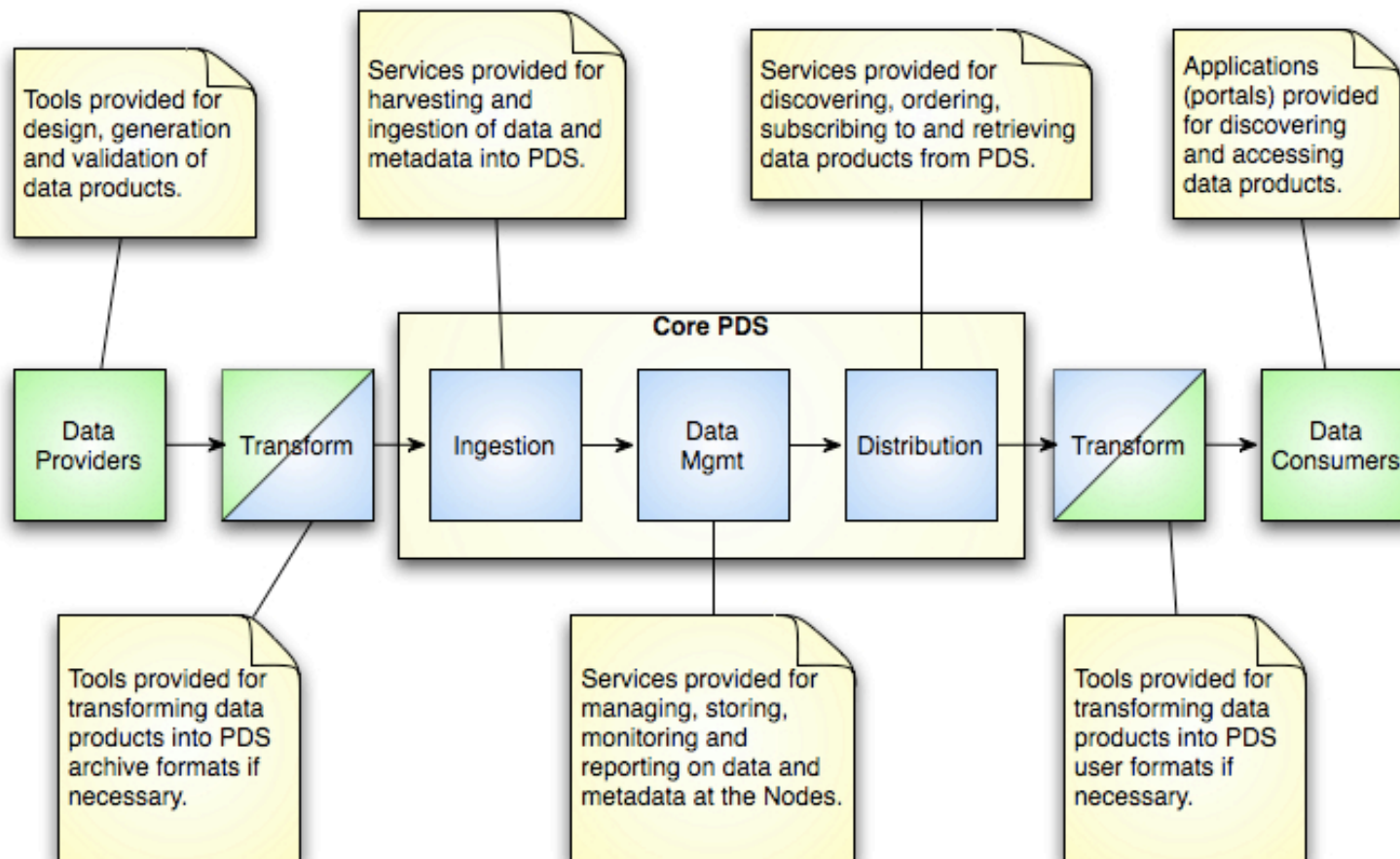
# System Overview

- The design of the system reflects the following:
  - Design goals and constraints set forth by the initial working groups.
  - PDS level 1, 2, and 3 requirements.
  - Best practices for science data and archive systems.
  - Decisions previously discussed.
- The slides that follow provide a couple of different views of the overall system.
  - The layered architecture view details how the different components (services, tools and applications) in the system interact and build on each other.
  - The information flow view details how those same components come into play throughout the flow of the system.

# Layered Architecture View

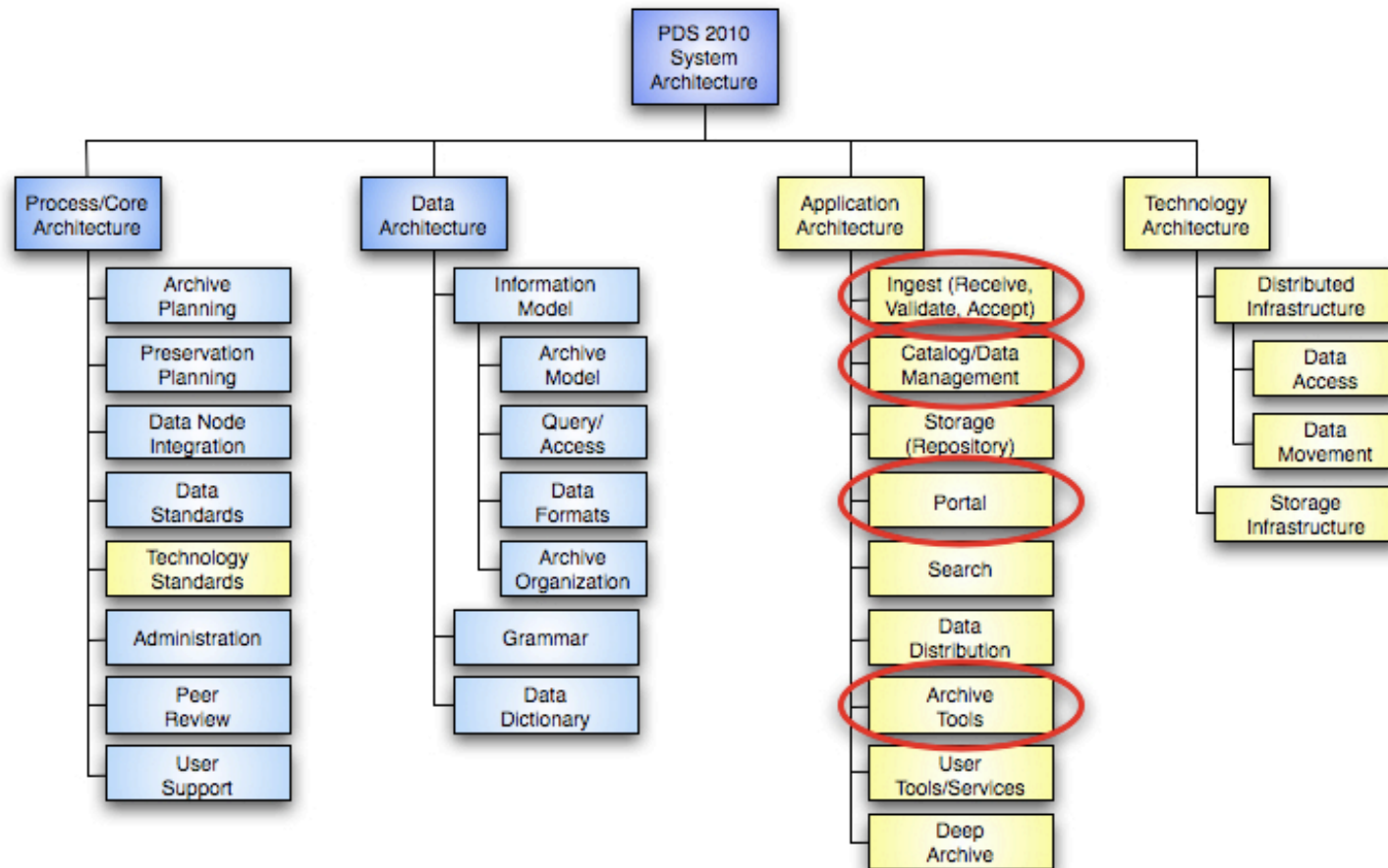


# Information Flow View



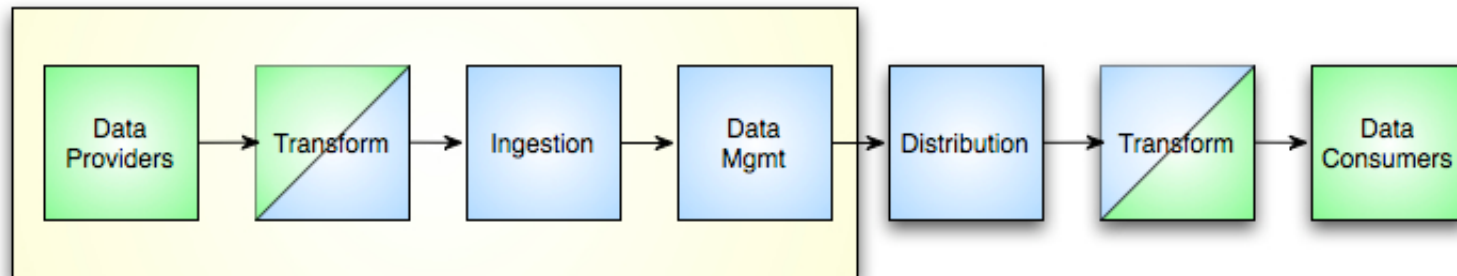
# Ingestion and Distribution Flows

# System Elements Related to Ingestion



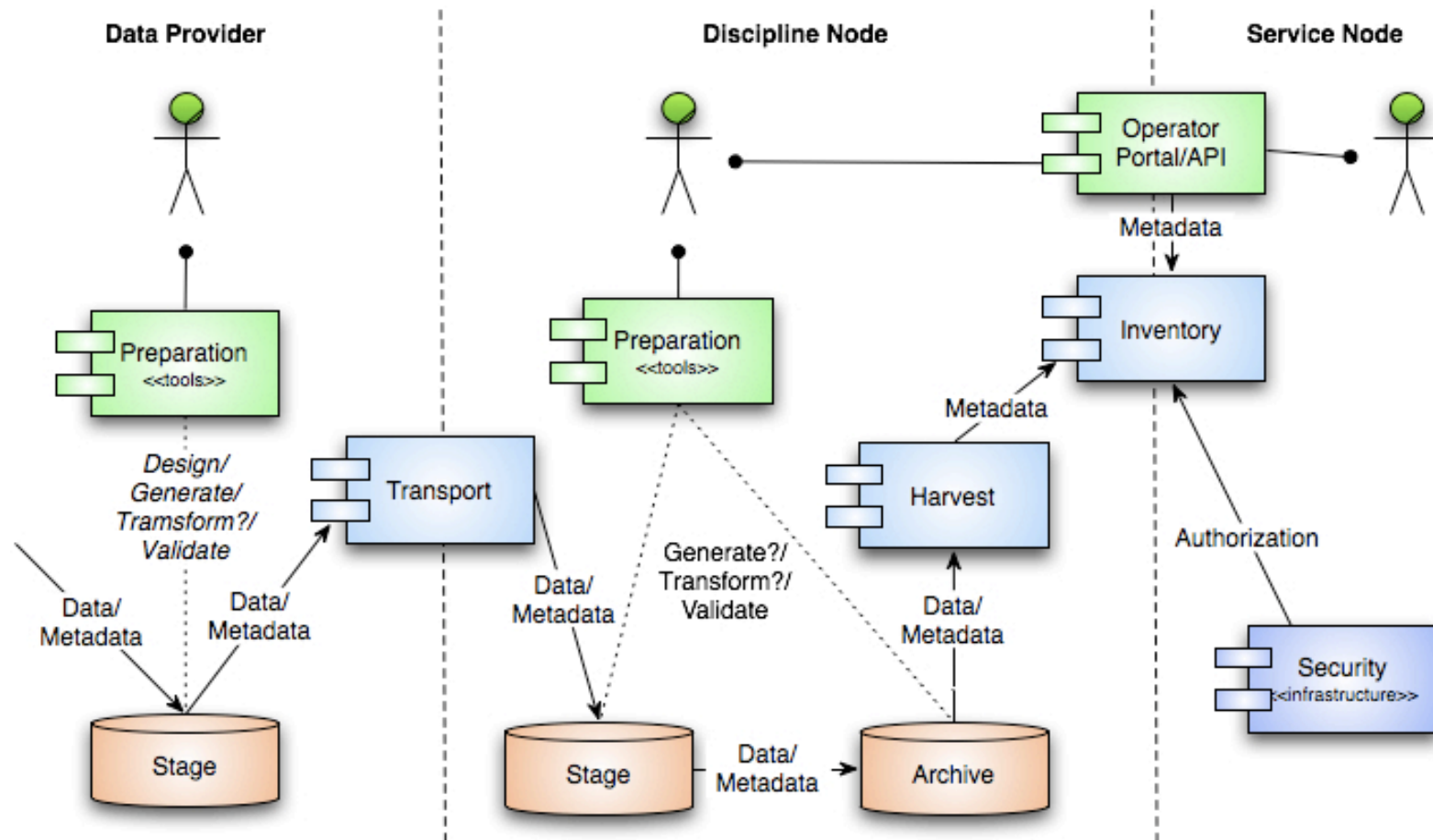


# Focus on Ingestion



- The ingestion scenario covers ingestion of catalog and data product metadata and its associated data into the PDS.
- A tool-based interface is offered to Data Providers to facilitate adoption of and interfacing with the PDS 2010 system.
- Transformation of incoming data/metadata is shown as a possible function for the Data Provider or the Discipline Node via a PDS provided tool.
- A Harvest service is introduced for capturing and registering catalog and data product metadata.
- An Inventory service is introduced for tracking metadata submissions and their associated data products.

# Ingestion Flow



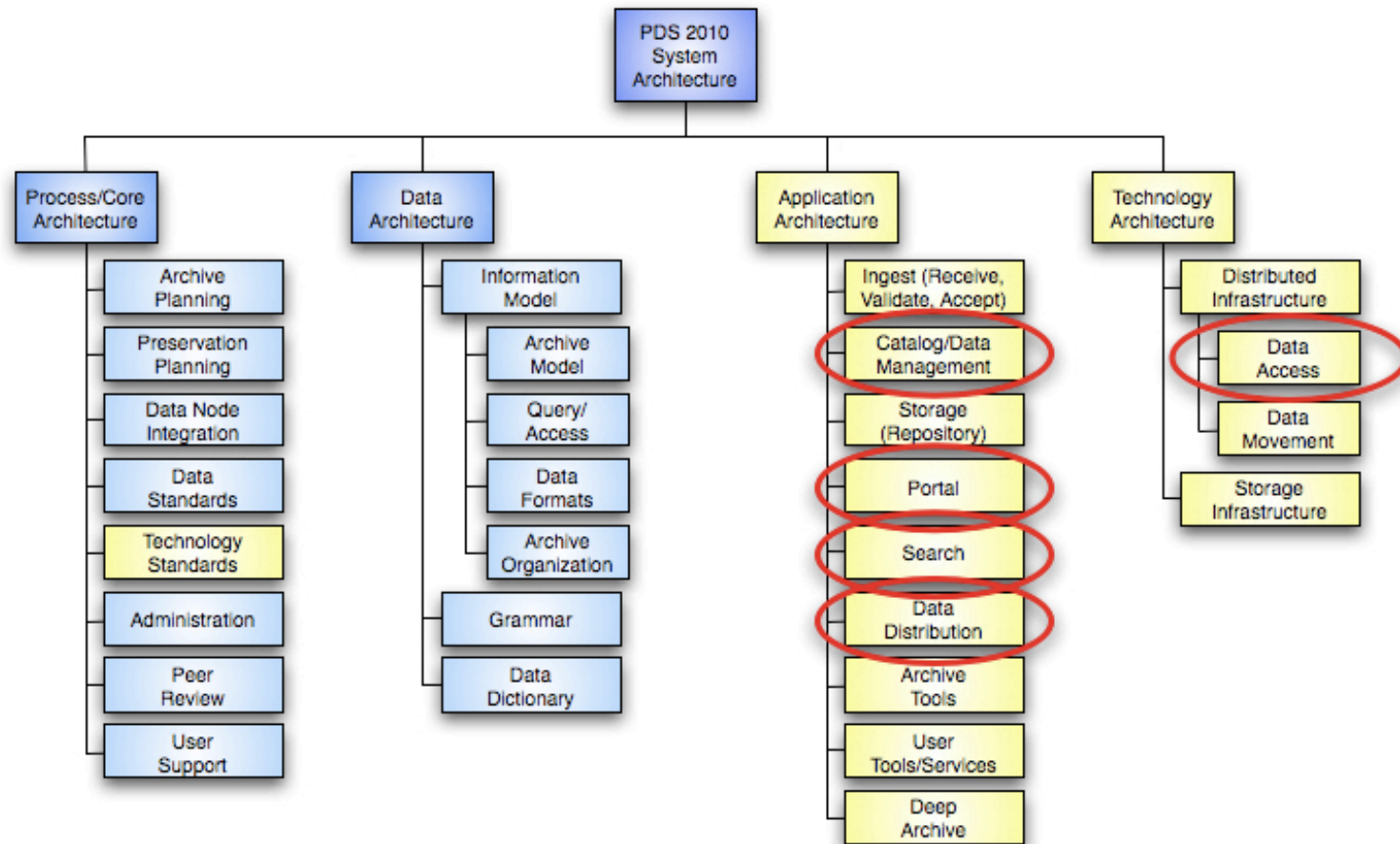
# Ingestion Flow Details

1. Data is received from the source.
2. PDS provided tools are utilized to prepare the data for submission.
3. Transformed/labeled data is submitted to the DN via an agreed interface (e.g., FTP, Data Brick, etc.).
4. Data/metadata are received from the Data Provider and staged.
5. PDS tools or tools based on a common library are utilized to prepare the data for archive.
6. Once in the archive, metadata (catalog and data product) are harvested and registered in the Inventory.
7. Metadata registrations are authorized by the Security service.
8. An Operator Portal is provided for managing housekeeping information and/or augmenting metadata for search enhancement.

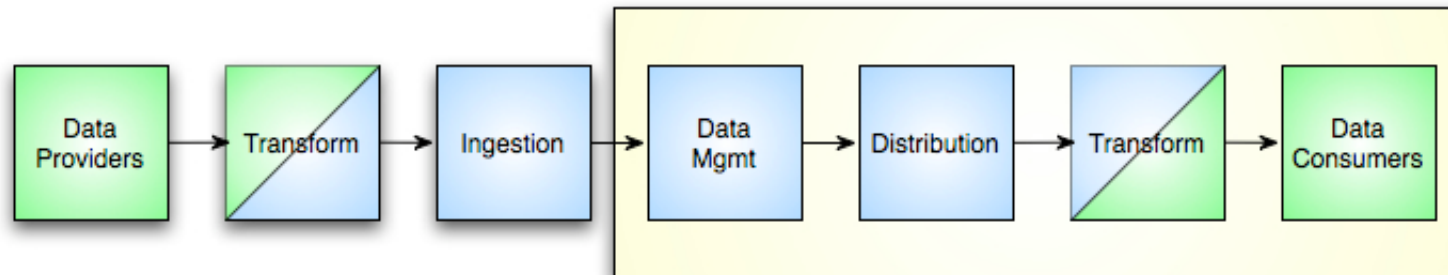
# Phased Approach for Ingestion

- Utilize tool-based interfaces for design, generation and validation of data product submissions.
  - Allows for existing Node processes and procedures to be utilized for ingestion of data products.
  - Minimizes up-front interface changes for Data Providers.
- Facilitate ingestion of catalog and data product metadata using the Harvest service.
  - Allows for periodic or on-demand ingestion of metadata.
  - Adaptable to different directory structures and data models allowing for ingestion of PDS3 and earlier data.
- A web-accessible Ingest service is planned for later on in the development cycle.
  - Leveraged from capabilities developed for the tools and the Harvest service.

# System Elements Related to Distribution

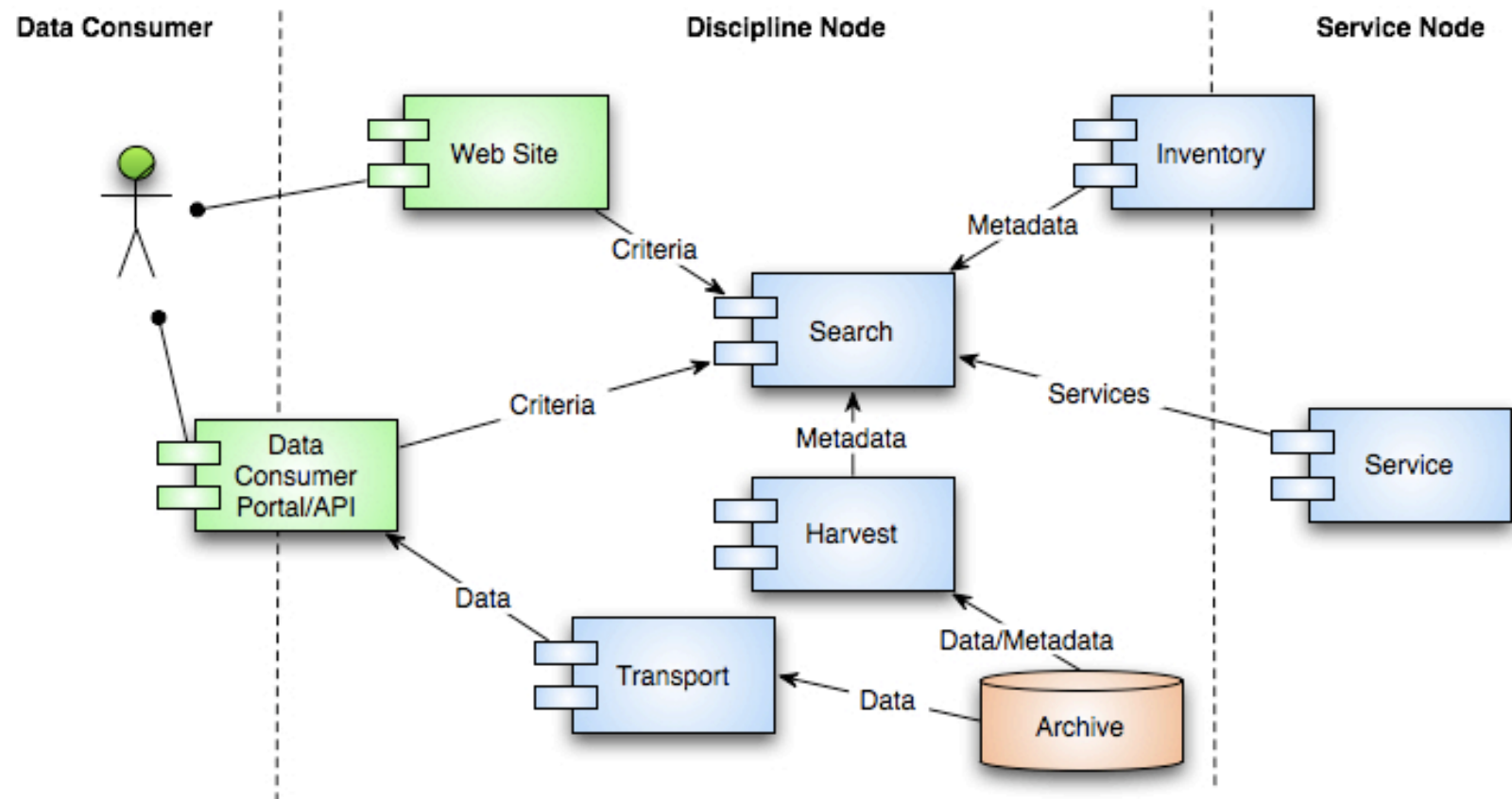


# Focus on Distribution



- The distribution scenario covers search of the catalog and data product metadata and distribution of associated data.
- A common Search service is introduced for interfacing with the services hosting catalog and Node-specific data product metadata.
- An extensible but common, Transport service is introduced to facilitate access and usability (i.e., transformation) of data products.
- Introduction of a Data Consumer portal/API for discovering and retrieving data/metadata.
  - The portal can be a Node developed application or user developed.

# Distribution Flow



# Distribution Flow Details

1. Data Consumer submits a query for data through a portal/API interface.
2. The query is forwarded to the local Search service.
3. The Search service utilizes the Service registry to discover the appropriate Inventory service(s) for the specified query.
4. Results are returned to the portal interface with options for retrieving data product(s) that match the query criteria.
5. A request is made to the Transport service for delivery of the data product(s).



# **Transition to Implementation**

# Implementation Plan

- Capture use cases and requirements for each service, tool and application.
  - Several efforts in the recent past have captured use cases for PDS.
  - Use these as the basis for the current effort.
- Develop a design for each service, tool and application.
- Hold design reviews with designated members of the Technical Staff.
- Carry out implementation/integration according to the project schedule and the phased build plan.

# Near-Term Activities and Plans

- Activities
  - Prepared the system architecture and high-level system design presented today.
  - Conducted a number of evaluations of COTS and Open Source software related to the core services.
- Plans
  - Focus on use cases and requirements for the core services (e.g., Security, Dictionary, Report and Registries).
  - Generate a detailed design for each of the core services, have them reviewed and begin implementation.

# Deployment Schedule

Component	Description	Deployment
Dictionary, Service and initial Search Services	This represents the rest of the core services.	Oct 2010
Design, Generation and Validation Tools	This represents the Data Provider tool suite for preparing data product submissions.	
Inventory, Document and Harvest Services	Services supporting capture and management of data product metadata.	
General Search Applications and Web Site Integration	Initial delivery of distribution services supporting discovery and access to PDS data.	Sep 2011
Transport Service	Additional mechanisms for transporting data.	
Order and Subscription Services along with DN Application Integration and Transformation Tools.	Enhanced capabilities for discovering and accessing PDS data along with support for DN-specific applications.	Jun 2012
Ingest and Storage Services	Includes a service-oriented ingestion capability and investigation of a Storage infrastructure.	

**Questions/Comments**