

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

Planetary Data System

System Distribution

Technical Session
Pasadena, CA
February 28 – March 2, 2011

Sean Hardman

Topics

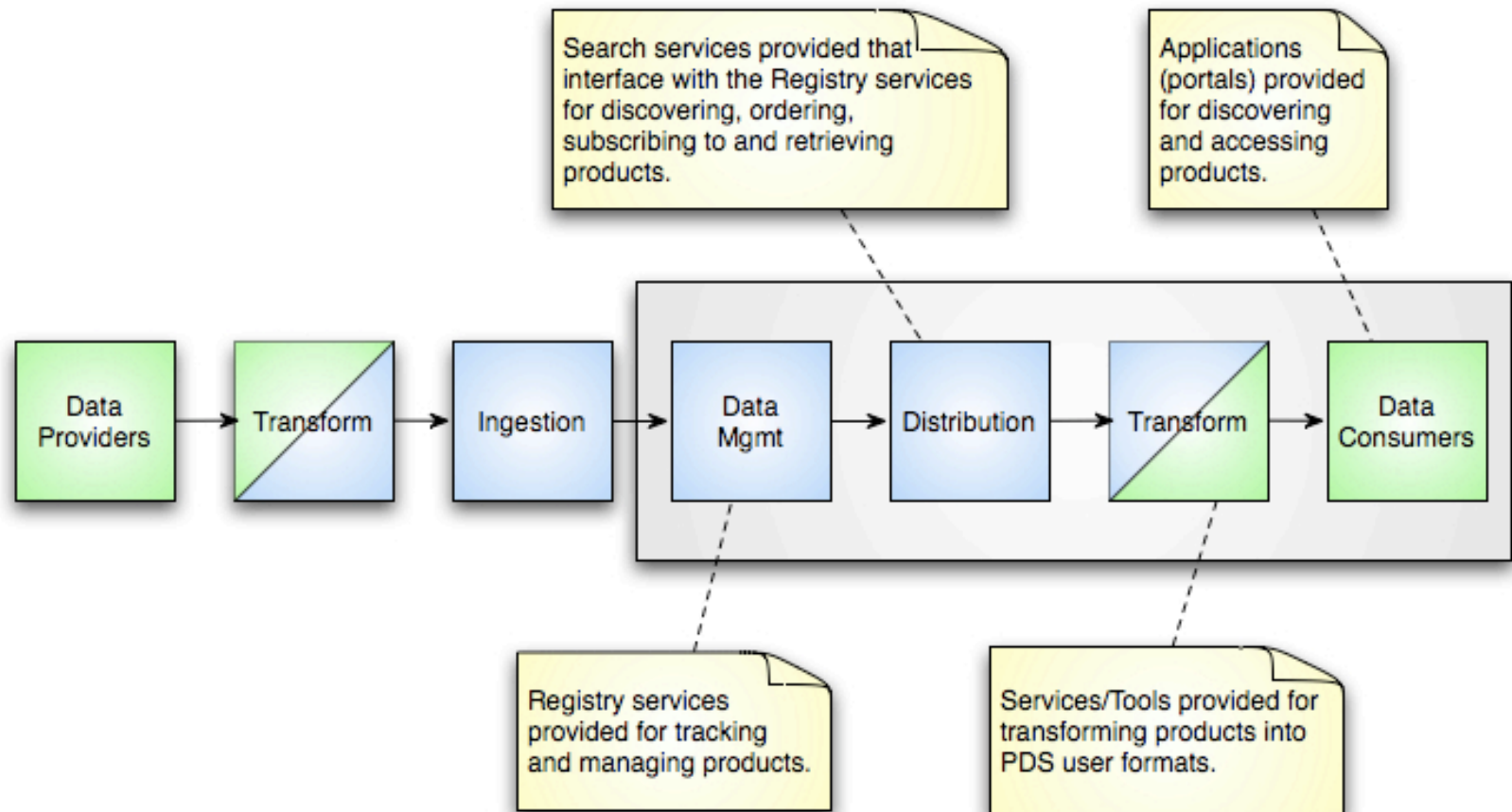
- Overview
- Distribution Related Components
- Distribution Flow
- Search Architecture and Protocol
- Search Scenarios

Overview

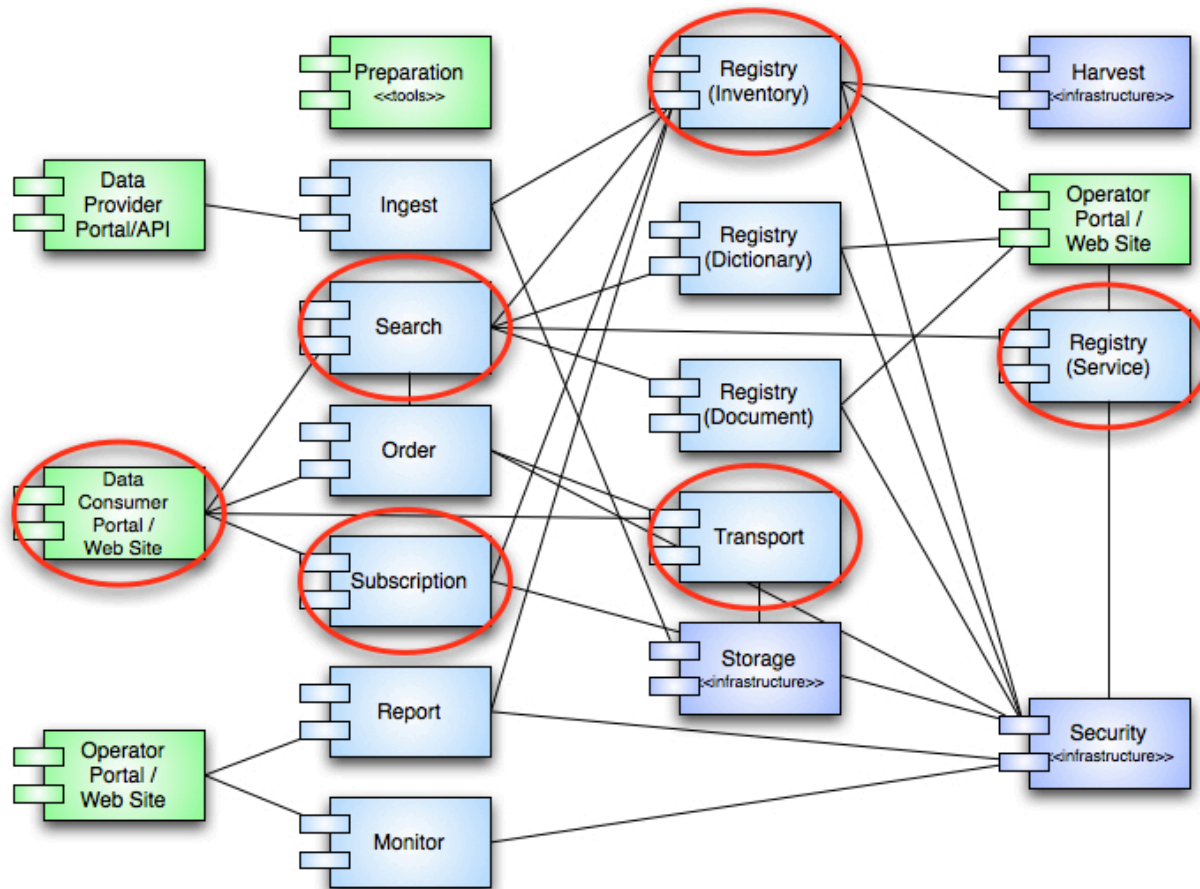
- Distribution involves discovery and retrieval of products.
 - The focus through builds 1 and 2 is on discovery.
- The Search component provides the infrastructure for facilitating all levels of product discovery.
- The SDWG and the DDWG are in the process of identifying and capturing search scenarios.

Distribution

(Discovery and Distribution of Products)



Distribution Related Components



Distribution Related Components

Data Consumer Portal, Subscription and Transport Services

- Data Consumer Portal
 - Integrate the PDS-wide portal (<http://pds.nasa.gov/>) with the Search Service.
- Subscription Service
 - Provides subscription to data, document and software release announcements.
 - Replace the current implementation with one built on the Registry Service.
- Transport Service
 - Integrate existing delivery mechanisms (e.g., FTP, HTTP, etc.).

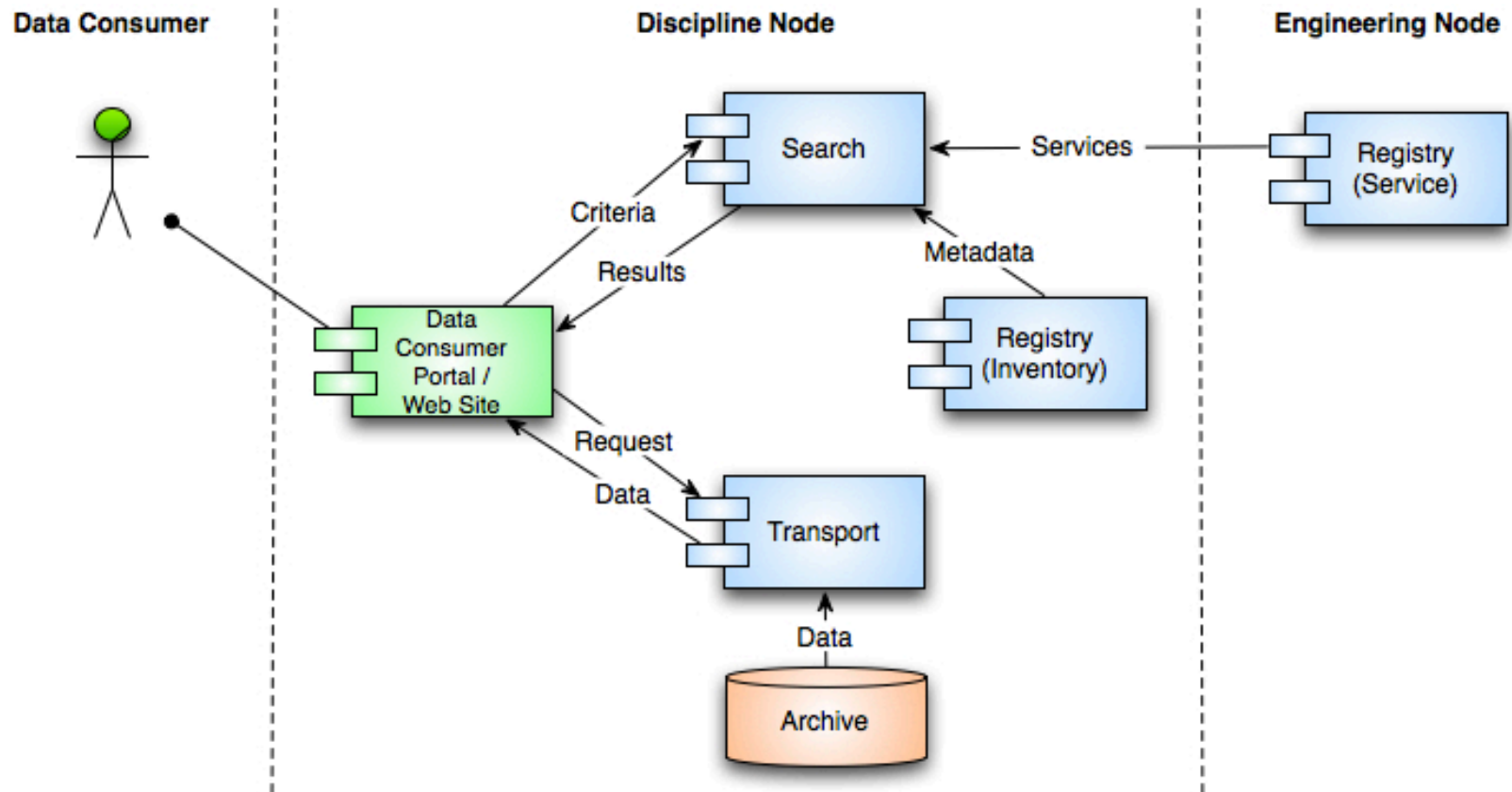
Distribution Related Components

Search Service

- This service is a deployable component that accepts queries for data and returns a set of matching results.
- Provides the public interface (REST-based over HTTP) to the metadata contained in the federated registries.
- Provides the second line of metadata harvesting within the system in order to facilitate discovery of products.
- Generation of search indices from registry metadata supports multiple query formats and is tailor-able for customized search interfaces.

Distribution Flow

Request Initiated at the Discipline Node



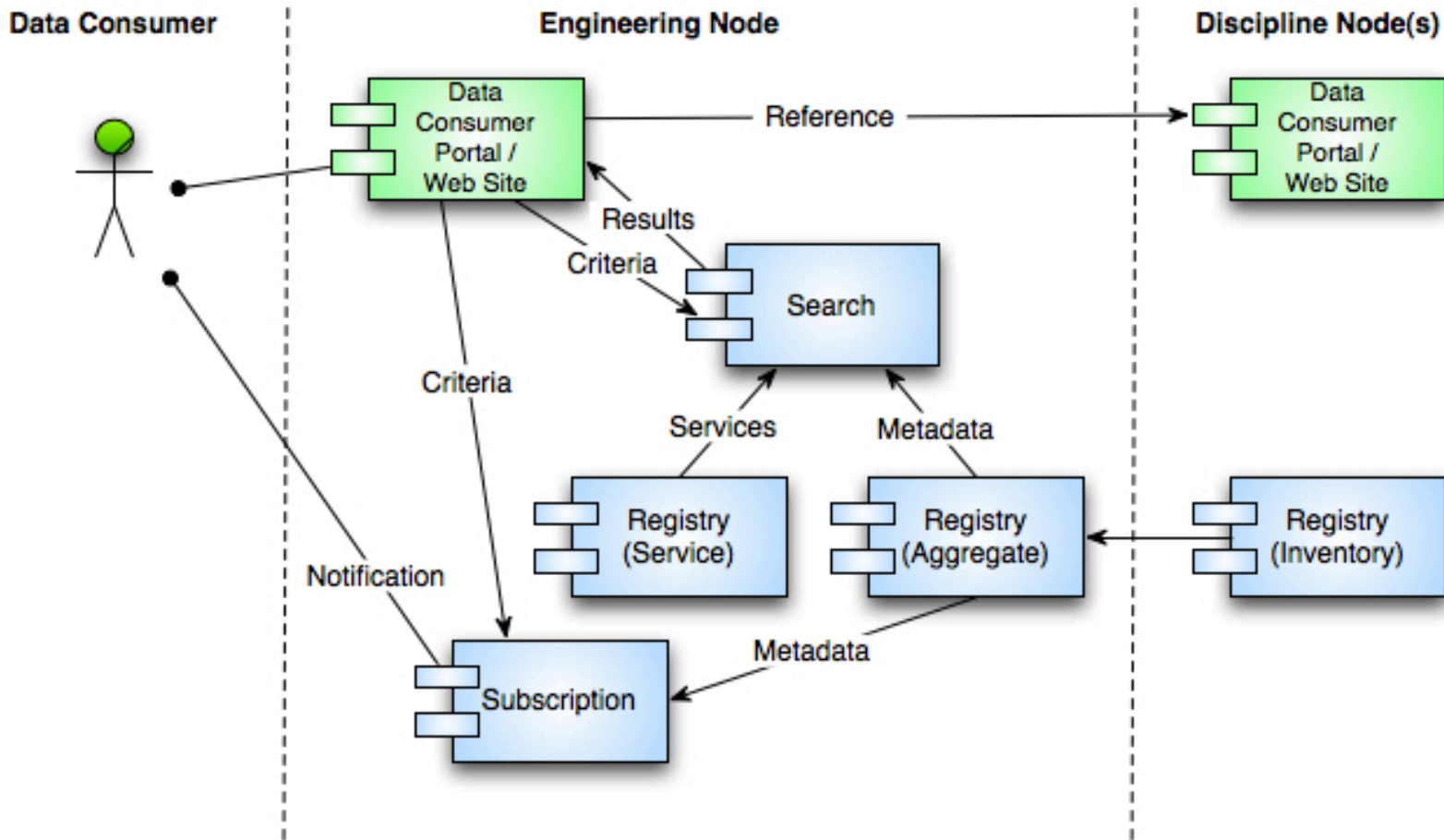
Distribution Flow Details

Request Initiated at the Discipline Node

1. Search service generates a search index utilizing the Service registry to discover the appropriate Registry service(s) for obtaining the metadata for the index. Tailoring of the search index enables support for the Node-specific search tools.
2. Data Consumer submits a query for data through a portal / web site interface.
3. Portal / web site interface forwards the query to the local Search service.
4. Search service returns results to the portal / web site interface with options for retrieving product(s) that match the query criteria.
5. Data Consumer makes a request to the Transport service for delivery of the product(s).

Distribution Flow

Request Initiated at the Engineering Node



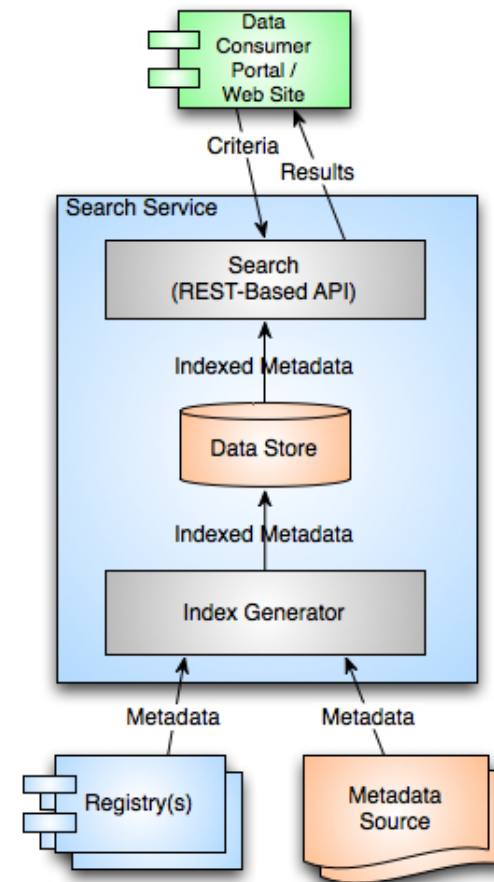
Distribution Flow Details

Request Initiated at the Engineering Node

1. Search service generates a search index utilizing the Service registry to discover the appropriate Registry service(s) for obtaining the metadata for the index.
2. Data Consumer submits a query for data through a portal / web site interface. The Data Consumer may also subscribe to release information via the Subscription service.
3. Portal / web site interface forwards the query to the Search service.
4. Search service returns results to the portal / web site interface with options for retrieving product(s) that match the query criteria.
5. Data Consumer makes a request to the Transport service from the appropriate Node for delivery of the product(s).

Search Architecture

- Search indexes built from multiple sources.
- Allows for annotation of archive metadata.
- Customizable for a discipline-specific search interface.



Search Protocol

- Currently designing the PDS search protocol implemented as a REST-based interface over HTTP.
 - <http://pds.nasa.gov/search?target=mars>
- Supports return of paged results in a defined structure (e.g., XML or JSON).
- Implementation of this protocol across PDS facilitates parameter passing and integration.
- The architecture allows support for other protocols.
 - For example, IPDA's Planetary Data Access Protocol (PDAP)

Search Scenarios

- In the process of collecting user search scenarios for the Search Service.
- The service design focuses on the search infrastructure.
- Scenarios will feed into query model development as well as metadata capture.
- Requesting that the Nodes provide these scenarios.
 - Hopefully captured for existing search interfaces.

Scenario Example

A science user wants to select a time sequence of both **wide and narrow angle frames** obtained when an associated storm was visible on the daylight side of **Saturn**. Storms drift around the planet at various rates. What they want to do is to execute a sequence of searches. First, set the gate to include **images** within **+/- one month** and retrieve a time-ordered list of **start time**, max and min **latitude** (0-20 deg) min and max **longitude** (25-45 deg), **camera name**, **filter** and **exposure time**. Hopefully, after they process selected frames from this data they can derive a longitudinal drift rate for the storm system and return with a series of bracketed searches to isolate the storm and study its evolution over a maximum length of time.

Catalog-Level Search

- Build 2 will replace the backend of the PDS Data Search capability at the EN.
 - <http://pds.nasa.gov/tools/data-search/>
- Heavily dependent on PDS3 catalog migration.
 - Cross-references are key to defining the relationships, for example:
 - Instrument > Instrument Host > Investigation
 - This feeds the facet-based approach of the search interface.
- Deployment Approach
 - The migrated catalog products are registered with an instance of the Registry Service.
 - The Search Service extracts metadata from the registry to generate search index(es).
 - The search interface is modified to utilize the new index(es).

Product-Level Search

- Coordination among Discipline-Specific model development.
 - Reuse of classes and attributes
 - Promotion to a broader namespace where possible
- Facilitates cross-Node search where appropriate.
 - This will never be completely satisfied by the data model.
- Useful to the Nodes whether they deploy the Search Service or not.

Desiderata

- Consistent population of the Identification Area.
 - Except for <product_class>, this hasn't been much of an issue.
- Consistent population of the Cross-Reference Area.
 - Preferred method for relating context information.
 - Subject Area is nice and readable but is less precise.
- Working with the DDWG to firm up definitions of Bundles, Collections and use of Context information.
- Need to incorporate taxonomies into the data model.

Wrap Up

- Design and development of the Search Service has focused on an architecture that supports the varied PDS search scenarios.
- The search architecture currently employed for PDS, where catalog-level search directs users to available data and discipline-specific search interfaces, lives on in the PDS 2010 system.

Questions / Comments