

A horizontal banner image featuring a sequence of celestial bodies from left to right: a blue planet (Earth), a brown planet (Mars), a larger brown planet (Jupiter), and a portion of a white planet (Saturn). The text "Planetary Data System" is overlaid in white on the right side of the banner.

Planetary Data System

# **System Ingestion**

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

Sean Hardman

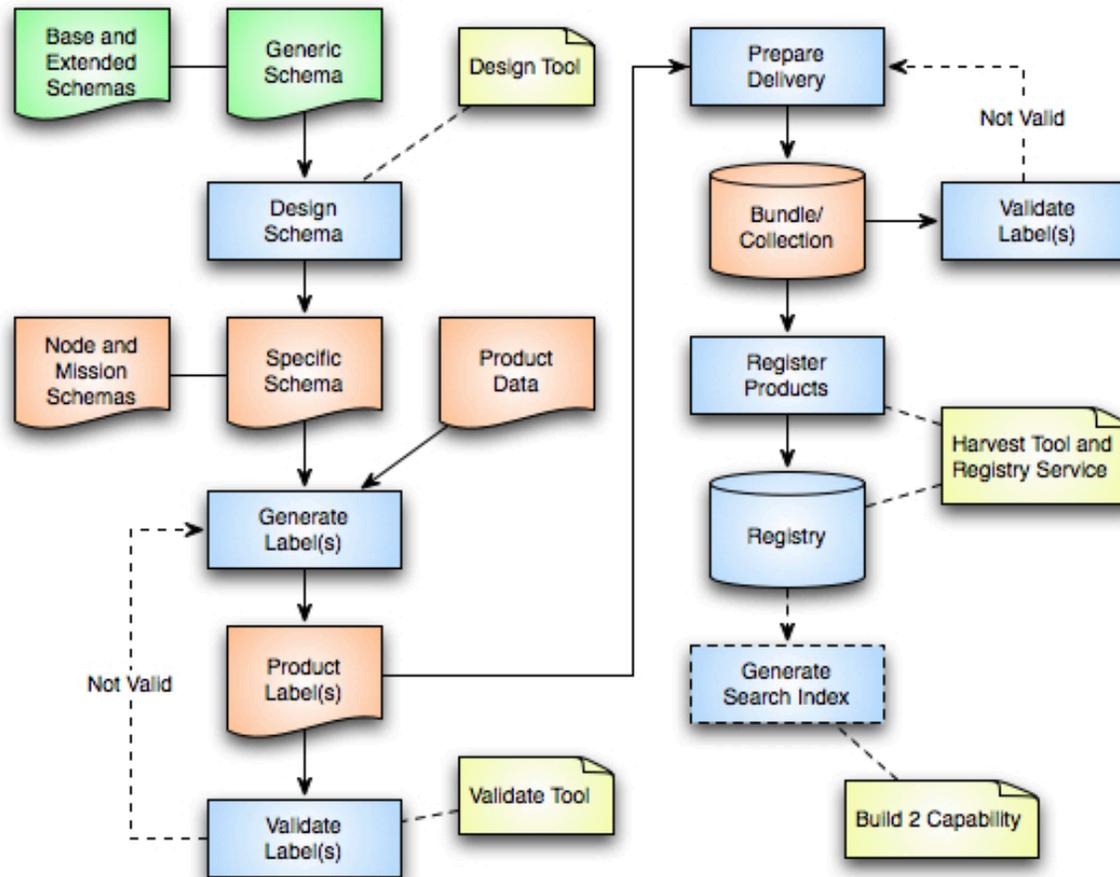
# Topics

- Overview
- Data Product Lifecycle
- Ingestion Related Components
- Ingestion Flow
- Harvested Metadata
- Registry Configuration and Process

# Overview

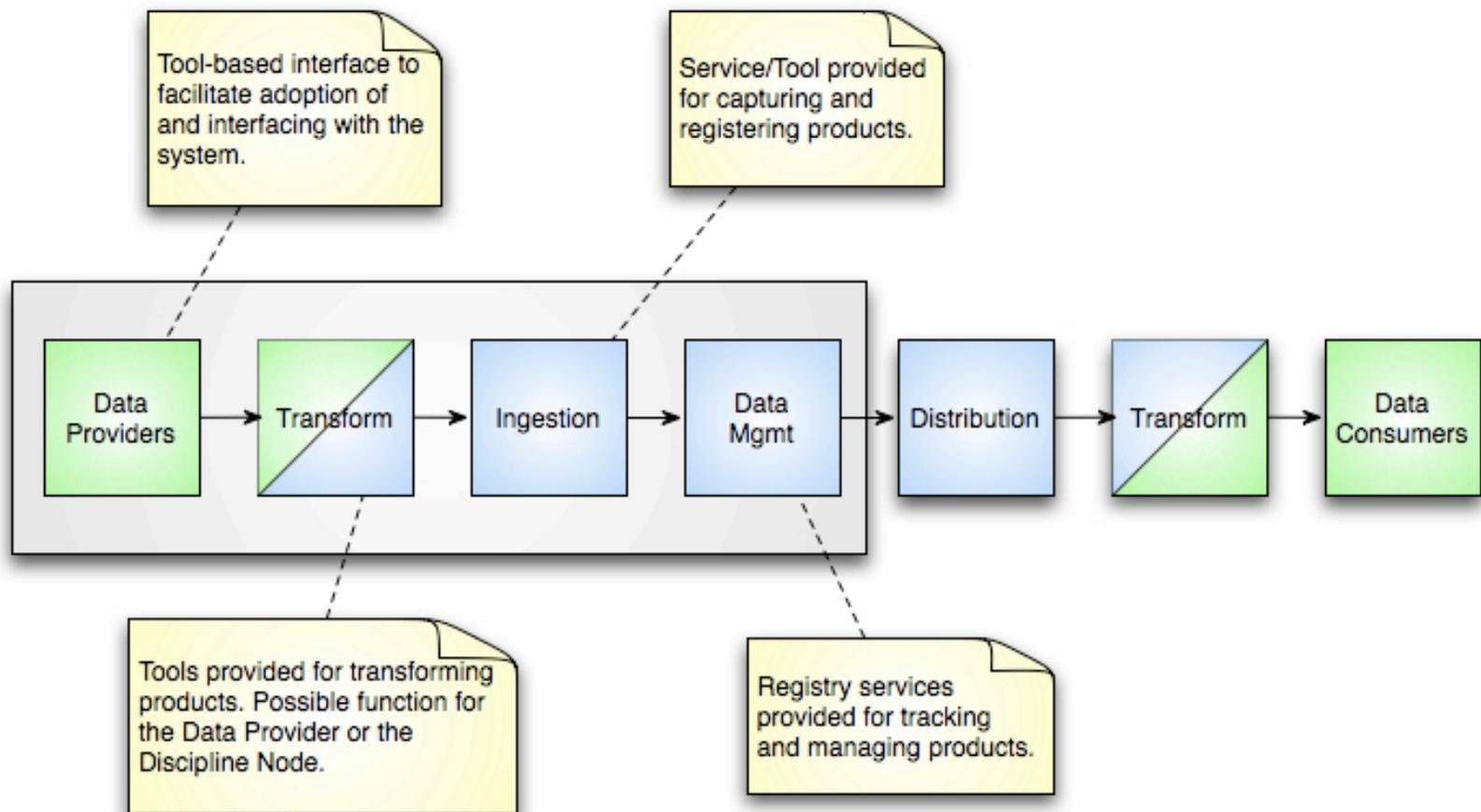
- Ingestion involves preparation of and registration of PDS products.
- Tools are provided for the preparation of these products.
- The Harvest and Registry components facilitate registration of these products.
  - Although they will support PDS3 products, this presentation focuses on PDS4 products.
- The following product lifecycle diagram details where the components fit into the process.

# Product Lifecycle

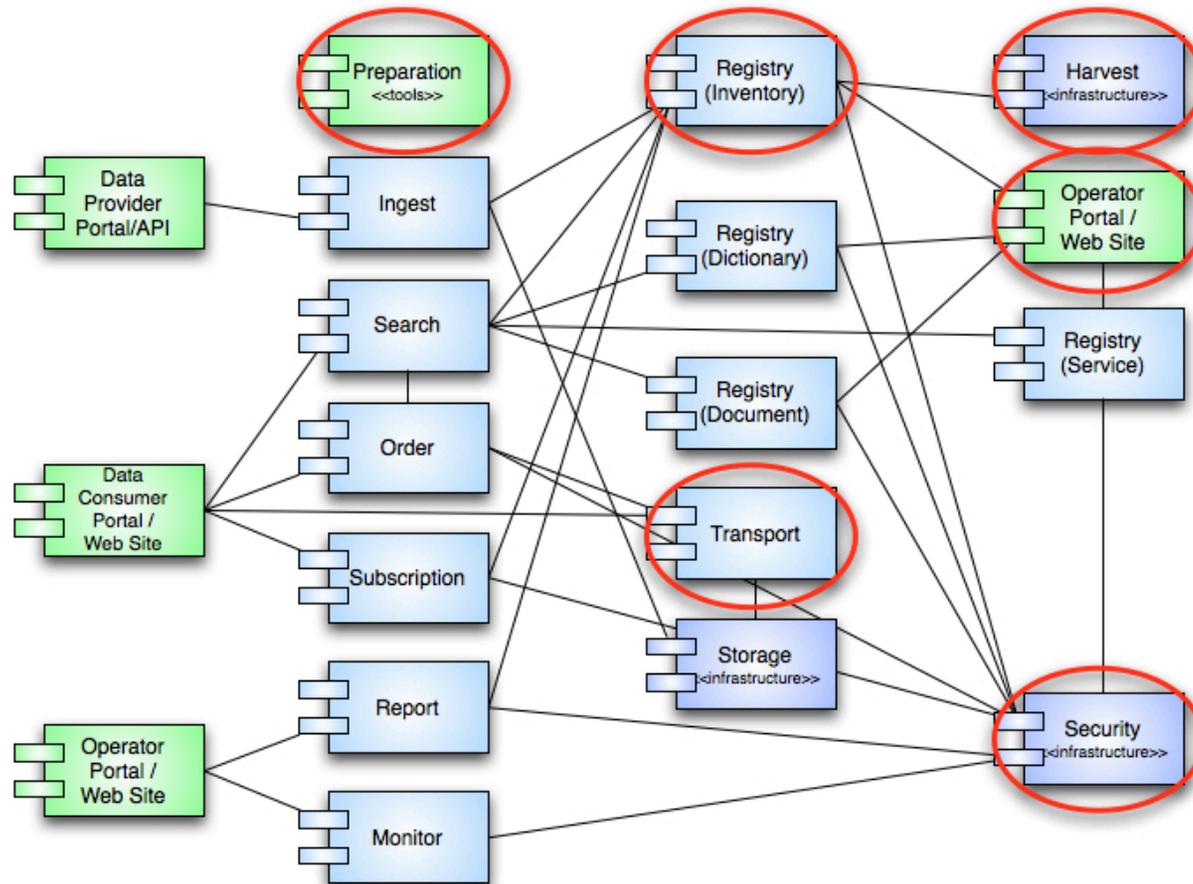


# Ingestion

(Capture and Registration of Products into the System)



# Ingestion Related Components



# Ingestion Related Components

## Preparation Tools and Transport Service

- Preparation Tools
  - Suite of tools for preparing data for ingestion into PDS focusing on design, generation, transformation and validation.
  - Allows for existing Node processes and procedures to be utilized for ingestion of data products.
  - Minimizes up-front interface changes for Data Providers.
- Transport Service
  - Represents continued support for FTP and Data Brick delivery mechanisms.
  - Currently looking into other mechanisms.

# Ingestion Related Components

## Harvest Tool

- Crawler-based tool for capturing and registering product metadata.
- Allows for periodic or on-demand registration of products.
- Configurable to support registration of products residing in PDS3 and PDS4 archives.
- Designed to integrate well with existing Node operations.
- Provides the first line of metadata harvesting within the system in order to facilitate tracking of and access to products.

# Ingestion Related Components

## Registry Service

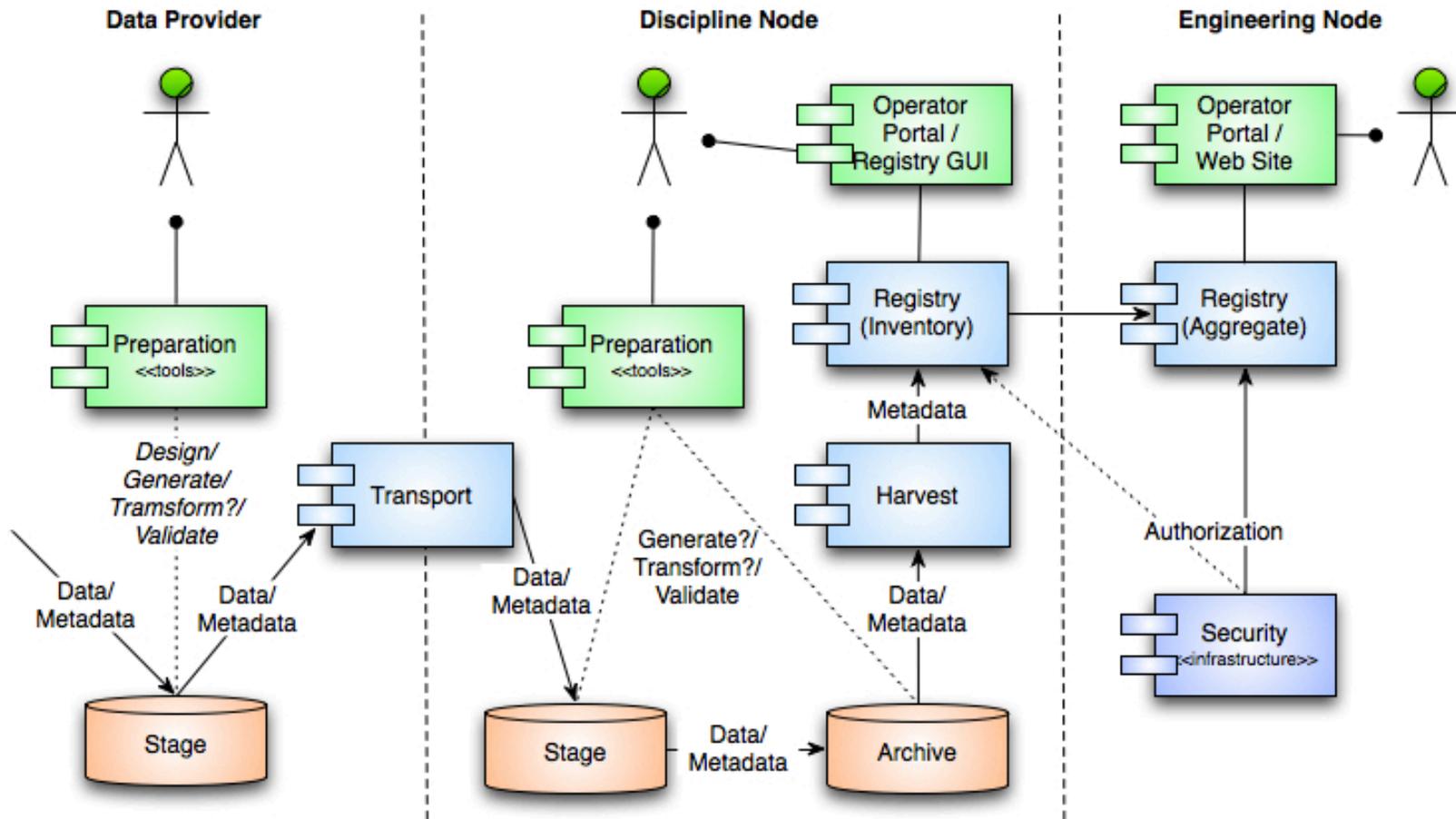
- Provides functionality for tracking, auditing, locating, and maintaining artifacts within the system.
- Provides a common implementation for registry service instances based on the CCSDS Registry and Repository Reference Model.
- Provides a REST-based external interface via HTTP.
- Provides a metadata store interface for supporting different databases.

# Ingestion Related Components

## Operator Portal and Security Service

- Operator Portal
  - A general web-based interface for managing registry policy, content and end-to-end tracking.
  - The interface is deployable for local instances of the Registry service at the Nodes.
- Security Service
  - Provides the authentication and authorization functions for the system.
  - Satisfied with an Open Source product supporting the Lightweight Directory Access Protocol (LDAP).

# Ingestion Flow



# Ingestion Flow Details

1. Data Provider receives data from the source (e.g., Project, Instrument Team, etc.).
2. Data Provider utilizes PDS provided tools to prepare the data for submission.
3. Data Provider submits transformed/labeled data to the Discipline Node via an agreed interface (e.g., FTP, Data Brick, etc.).
4. Discipline Node receives data/metadata from the Data Provider and stages it in local storage.
5. Discipline Node utilizes PDS tools or tools based on a common library to prepare the data for archive.
6. Discipline Node initiates harvesting of the archive, which registers product metadata with the Registry service. Metadata registrations are authorized by the Security service.
7. Metadata is replicated with the Registry service hosted at the EN.
8. Discipline Node manages housekeeping information and/or augments metadata for search enhancement via the Operator Portal.

# Registry Entry Types

- Non-Digital Object Entry
  - Captures a description of a non-digital object.
  - Examples include investigations, instruments, etc.
- Digital Object Entry
  - Captures a description of a digital object that is referenced by an URI.
  - Examples include products, product label schemas, etc.
- Relationship Entry
  - Captures an association between registered objects.
  - Associations are also typed (e.g., member of).

# Harvested Metadata

## Identification Area

- The logical id and version id become the unique identifier for the product.
- Product class is used to classify the object type.
- Title becomes the display name.

```
<Identification_Area_Product>
  <logical_identifier>URN:NASA:
    PDS:data_set.A12A-L-SWS-3-
    SOLAR-WIND-28S-RES-V1.0
  </logical_identifier>
  <version_id>v1.0
  </version_id>
  <product_class>
    Product_Data_Set_PDS3
  </product_class>
  <title>APOLLO 12 ALSEP/SWS
    SOLAR WIND 28-SEC
    RESOLUTION TABLES V1.0
  </title>
  ...
</Identification_Area_Product>
```

# Harvested Metadata

## Cross Reference Area

- A reference entry becomes an association in the registry.
- Relates two registered objects with an association type.

```
<Cross_Reference_Area_Context>
<Context_Reference_Entry>
  <lidvid_reference>
    URN:NASA:PDS:
      investigation.APOLLO_12::1.0
  </lidvid_reference>
  <reference_association_type>
    has_investigation
  </reference_association_type>
</Context_Reference_Entry>
...
</Cross_Reference_Area_Context>
```

# Harvested Metadata

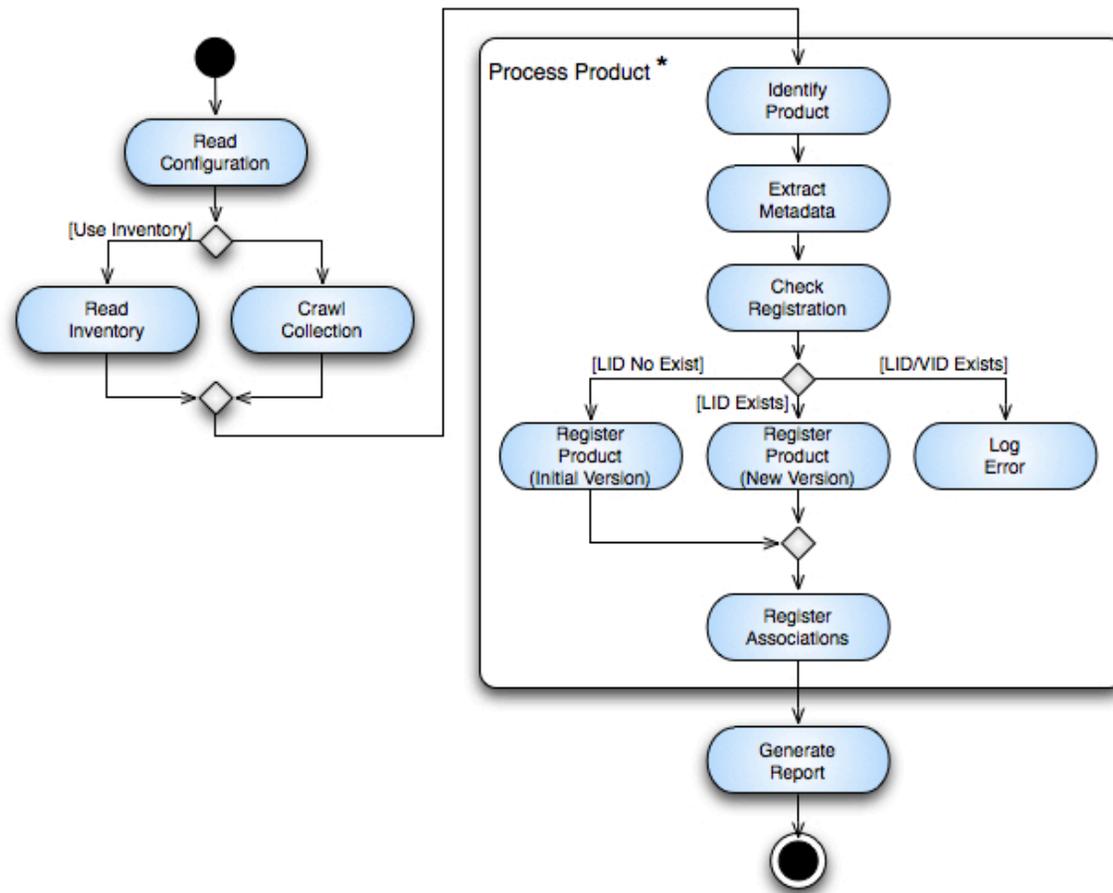
## Other Areas

- Other areas such as the Subject and Data areas can be harvested on demand.
- Depends on the search requirements for the local registry.
- Harvest can be configured to extract specific elements from the product label and place them into slots in the registry.

# Registry Configuration

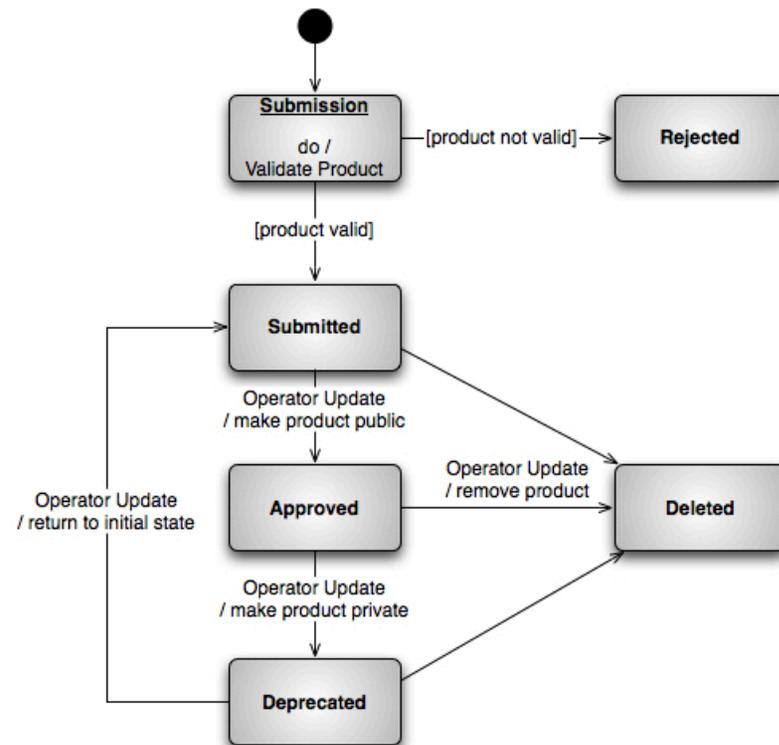
- The data model identifies the PDS product types.
- It will soon identify the common metadata elements for each of the product types.
- This information can be exported in a form to facilitate Registry Service configuration.
- Allows the data model to exert some semblance of control of the contents of the registries.

# Harvest Process



# Registry Status

- A status of "Submitted" indicates successful registration.
- Registered products are not visible to the public until approved.



# Registry Replication

- The EN will host an aggregate Registry Service that periodically pulls registry entries from the distributed set of Node-hosted Registry Service instances.
  - Only entries with a status of “Approved” are aggregated.
- Allows the Nodes to maintain governance of their own registry contents.
- The aggregate registry instance facilitates PDS-wide tracking and reporting.

# Wrap Up

- The ingestion process and flow closely mimic existing processes at the Nodes.
- One goal was to minimize impact on current Node operations.
- The Harvest and Registry components provide the core functionality for the rest of the system to be built upon.
- Facilitates tracking, reporting and ultimately product-level search.

# Questions / Comments