



International Planetary Data Alliance

Registry Development and Coordination Project Report



7th IPDA
Steering Committee Meeting

July 13, 2012



Topics

- Project Overview and Goals
 - An overview of the IPDA Registry Project.
- Approach/Architecture
 - The approach and architecture guiding the project effort.
- Registry Service and the Harvest Tool
 - An overview of the PDS software utilized by the project.
- Status and Future Work
 - Progress to date and future plans.
- Demonstration
 - Demonstrate the Registry Service functionality.

Overview

- This project was initiated at the 6th IPDA Steering Committee meeting in September 2011.
- The intent was to build on the success of the previous year's Registry Implementation Project.
- That project stood up an IPDA Service and Tool Registry utilizing the PDS Registry Service and a custom web-based interface.
 - <http://planetarydata.org/services/registry>

Original Project Goals

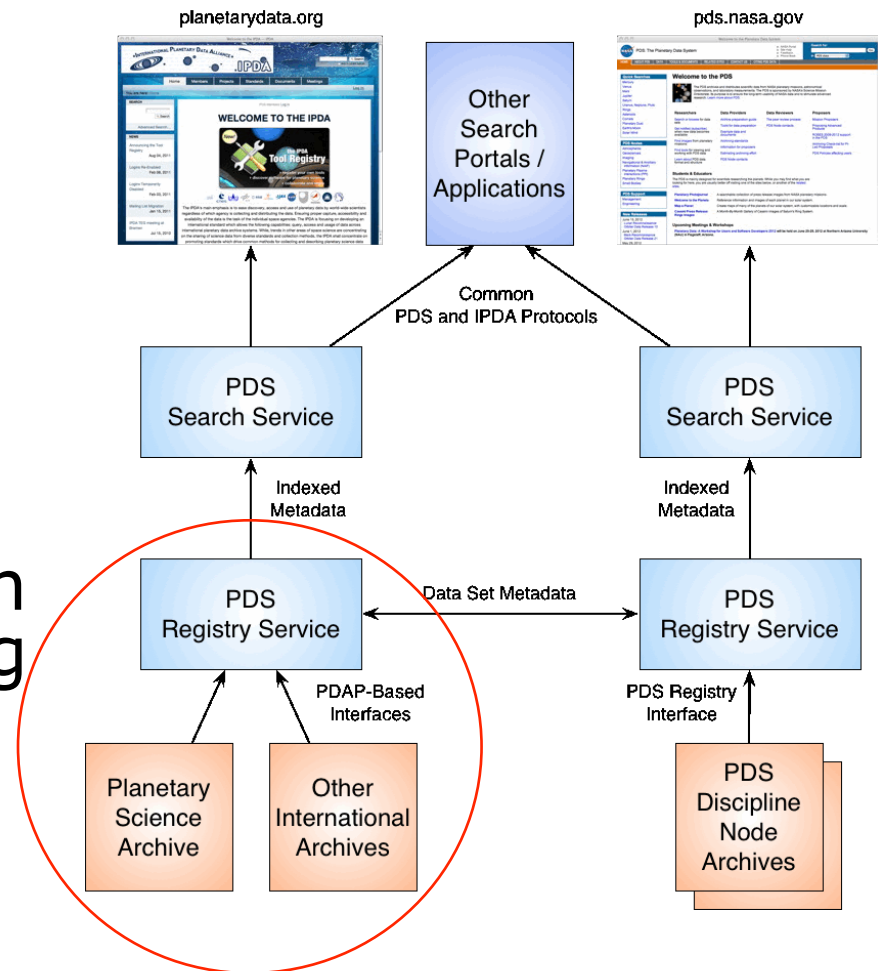
- Maintain existing registry implementation including upgrades to support the latest version of the PDS Registry Service.
- Develop an interface for managing, searching and viewing data set registrations.
- Develop an automated mechanism for maintaining data set registrations.
- Populate the registry with Venus Express data set information.

Approach

- Build the software system based on generic common software and common protocols for accessing that software.
 - PDS Registry Service with its REST-based API is the main component.
 - PDS Search Service based on Apache Solr provides support for high performance facet-based search.
- Utilize the PDS4 data model for data object definitions and to configure the software where appropriate.
 - The model defines the key context objects (i.e., Data Set, Instrument, etc.).

Architecture

- This project focused on deployment of the Registry Service and the harvesting of PSA data set metadata.
- Registry Service provides a common model for capturing metadata across multiple agencies.

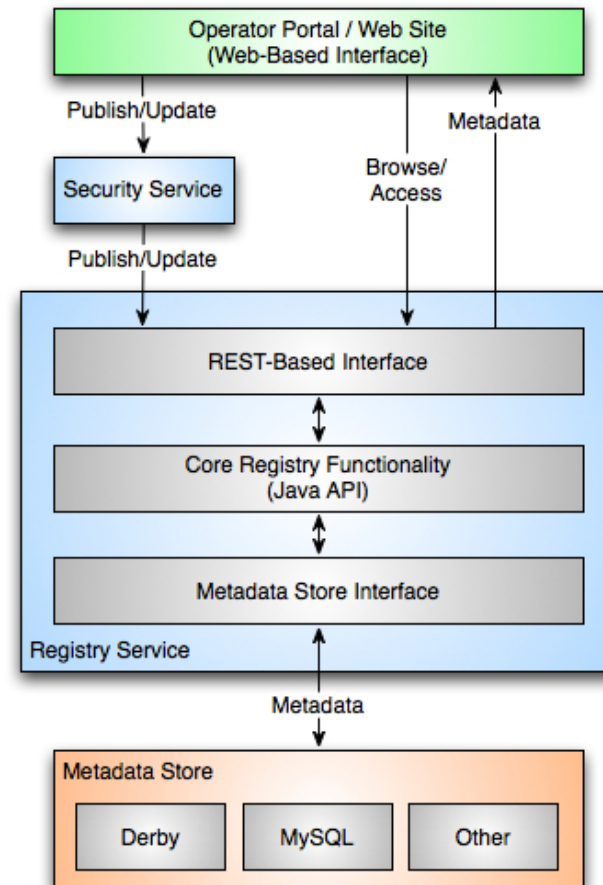


Registry Service

- Provides functionality for tracking, auditing, locating, and maintaining artifacts within the system.
 - Artifacts include data products, data dictionary element definitions, service descriptions and project documents.
- Provides a common implementation for registry service instances based on the Registry Reference Model effort which in turn is based on ebXML.

Registry Architecture

- REST-based API over HTTP for registration and retrieval of metadata.
- Internals developed in Java with an API for manipulating registry objects.
- Metadata store interface allows for multiple database solutions.



Registry REST-Based API

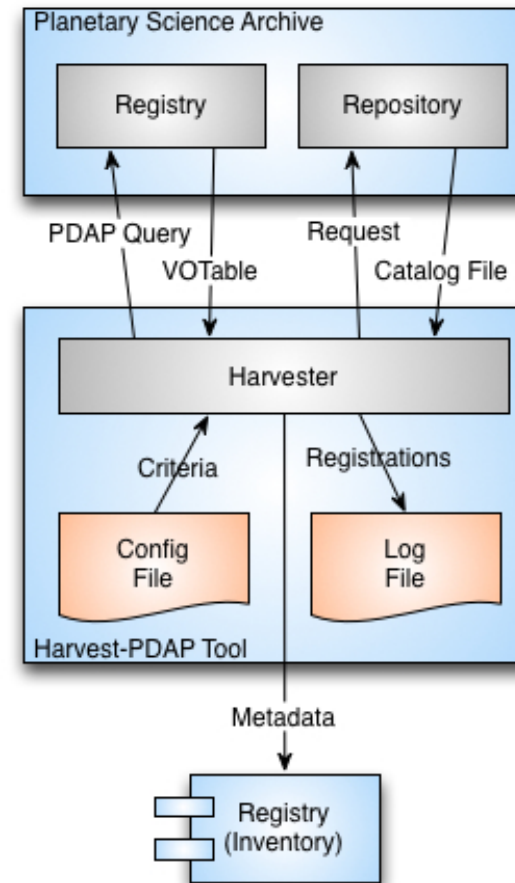
- This interface delegates all functions involving a product:
 - <http://ipda.jpl.nasa.gov/registry/extrinsics/>
 - GET: Retrieves a paged list of products from the registry.
 - POST: Publishes a product to the registry.
- This interface acts on a specific product (lid stands for logical identifier):
 - <http://ipda.jpl.nasa.gov/registry/extrinsics/logicals/{lid}/>
 - GET: Retrieves the product from the registry.
 - POST: Updates the product in the registry.
 - DELETE: Removes the product from the registry.

Harvest Tool

- Crawler-based tool for capturing and registering product metadata.
- Provides a PDS-specific interface to the Registry Service for registering products.
- Allows for periodic or on-demand registration of products.
- Configurable to support registration of products residing in PDS3 and PDS4 archives.
- Provides the first line of metadata harvesting within the system in order to facilitate tracking of and access to products.

PDAP-Specific Harvest Tool

- Based on the PDS Harvest Tool.
- Currently accesses PSA via its PDAP interface.
- Can be extended to support other archives with PDAP interfaces.



Status

Original Project Goals

- Maintain existing implementation ...
 - Supported updates for three PDS builds.
- Develop a data set interface ...
 - Decided not to build a web-based interface at this point in time.
- Develop an automated mechanism ...
 - The Harvest Tool can be configured to run periodically to pick up new data sets.
- Populate the registry ...
 - Instead of limiting to just Venus Express, the Harvest Tool registers all available data sets from the PSA archive.

Status

Other Stuff

- Updated and improved the Python-based client library that the Tool Registry uses to access the Registry Service.
- Deployed the latest version of the Registry Service for IPDA.
 - <http://ipda.jpl.nasa.gov/registry/>
- Developed a Harvest Tool to extract data set metadata from PSA via the PDAP interface.
- Populated the IPDA registry with PSA metadata.
- Updated the Registry Service Design Specification document to correspond with the latest release of the PDS Registry Service.
- Authored the Registry Service Protocol and Harvest Tool Design Specification documents.

Future Work

- Work with other agencies to gain access to their data set metadata in order to populate the IPDA registry.
- Expand the support beyond data sets to encompass investigation, instrument, etc. information.
- Develop a procedure/process for keeping this extracted metadata up-to-date.
- Expand the IPDA web site interface to support management and search beyond tools and services.

Demonstration

Overview

- This demonstration exercises the IPDA instance of the Registry Service and its REST-based API.
- The protocol for interfacing with the service is captured in the Registry Service Protocol document.
 - <http://planetarydata.org/projects/active-projects-for-2011-2012/registry-development-and-coordination/registry-service-protocol>

Report Status

- Navigate to the report endpoint:
 - <http://ipda.jpl.nasa.gov/registry/report>

- The following should appear in the browser:

```
<ns2:report xmlns:ns2="http://registry.pds.nasa.gov"
registryVersion="1.2.0"
packages="6"
classificationNodes="70"
classificationSchemes="2"
services="14"
extrinsics="4282"
associations="4356"
serverStarted="2012-07-09T17:25:36.216-07:00"
status="OK" />
```

The “extrinsics” class includes the PSA Data Set products and their associated Resource products which reference the online URL of the data set.

Search for Data Sets

- Navigate to the extrinsic endpoint:
 - [http://ipda.jpl.nasa.gov/registry/extrinsics?
name=*VENUS EXPRESS*&
objectType=Product Data Set PDS3](http://ipda.jpl.nasa.gov/registry/extrinsics?name=*VENUS EXPRESS*&objectType=Product Data Set PDS3)
- The following should appear in the browser:

```
<ns2:response xmlns:ns2="http://registry.pds.nasa.gov"
  numFound="82" start="1">
<ns2:results>
  <ns2:extrinsicObject
    versionName="1.0"
    status="Submitted"
    objectType="Product_Data_Set_PDS3"
    name="VENUS EXPRESS VENUS VRA 1/2/3 NMP 0125 V1.0"
    lid="urn:nasa:pds:data_set.VEX-V-VRA-1/2/3-NMP-0125-V1.0"
    home="http://localhost:8080/registry"
    guid="urn:uuid:0255c483-54e8-4561-adc7-0af13b964395">
  ...
```

There were 82 data sets found where the "name" included "VENUS EXPRESS".

List Services

- Navigate to the services endpoint:
 - <http://ipda.jpl.nasa.gov/registry/services>
- The following should appear in the browser:

```
<ns2:response xmlns:ns2="http://registry.pds.nasa.gov"
  numFound="14" start="1">
  <ns2:results>
    <ns2:service
      versionName="1.0"
      description="The PGS allows for searching data in datasets,
        based on geometry."
      status="Submitted"
      objectType="Service"
      name="PSA Geometry Search (PGS)"
      lid="urn:ipda:tbd:23"
      home="http://localhost:8434/registry"
      guid="urn:uuid:0189e46e-13d0-42eb-b69c-c471b2782c75">
    ...
```

Wrap Up

- The IPDA instance of the Registry Service is hosted along side the prototype IPDA web site.
 - Once the DNS switch is made, both will be accessible via <http://planetarydata.org>.
- All endpoints are accessible to the outside world for GET requests.
- All other request types (e.g., POST, DELETE, etc.) are restricted to local machine access.
 - If external access is required in the future, we can configure it for authenticated access.

Questions / Comments