



Jet Propulsion Laboratory
California Institute of Technology



Jet Propulsion Laboratory

Cartography and Imaging Sciences Node

U.S. Geological Survey

PDS Data Services Workshop:
*Lessons Learned from
PDS Imaging Node Data Services*

Rishi Verma

Jet Propulsion Laboratory,
California Institute of Technology

Agenda

- 1 Overview of Data Services
- 2 Case Studies
 - Label mapping tool
 - Metrics
 - Photojournal
 - Atlas / w10n
- 3 Recap / Community Recommendations



Cartography and Imaging Sciences Node

Jet Propulsion Laboratory

U.S. Geological Survey

Imaging Node (IMG) is home to over 700 TB of digital image archives and is a partnership between JPL and USGS

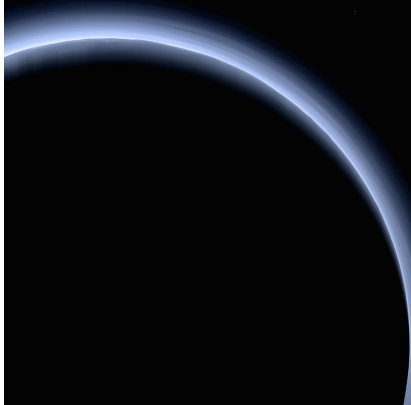
Data characteristics:

- Orbital vs. landed missions
- Original raw experiment data vs. derived products
- Differing coordinate systems

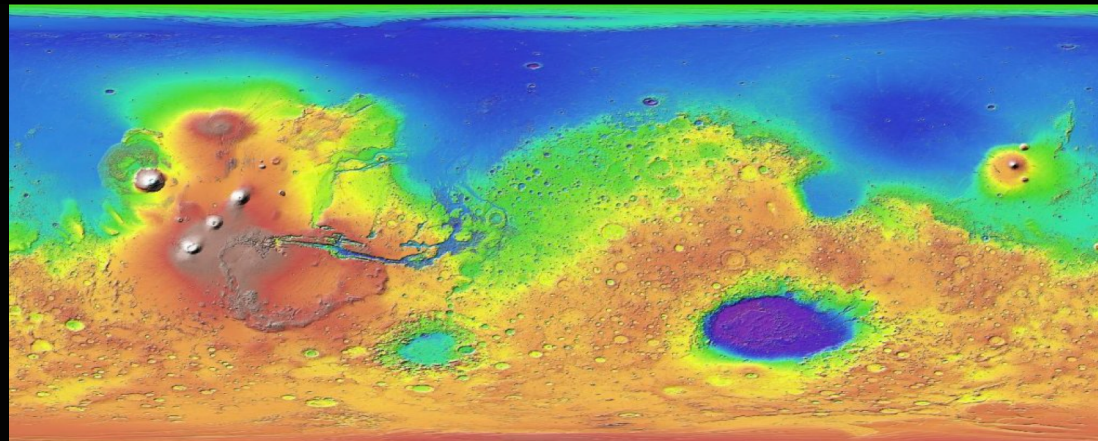
Users:

- Scientists
- Public
- Developers

New Horizons spacecraft:
Plutonian Atmosphere



Mars Global Surveyor: Mola Global Colorized Hillshade



Curiosity Rover Hazcam Image of
Drilling on Jan. 13, 2015

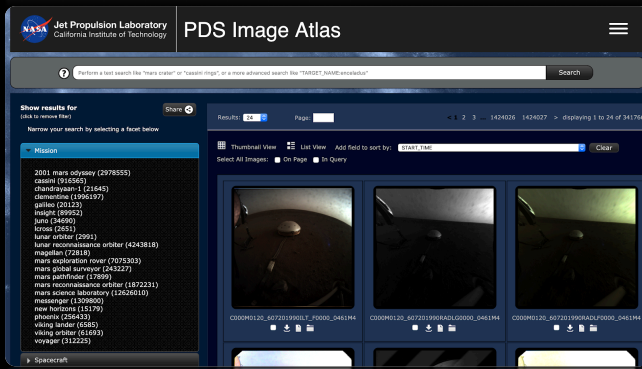


Overview of Data Services



Data Archive Portal (public)

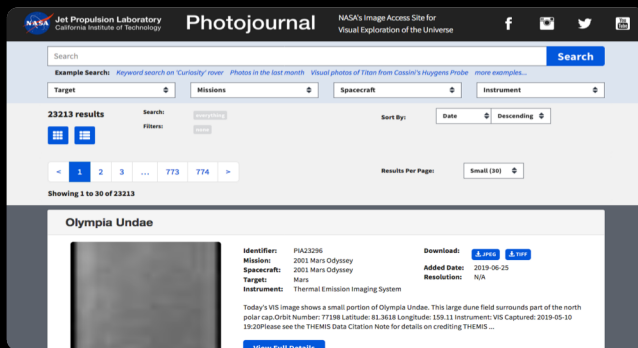
- HTTP(s) access to JPL-based 70+ TB of data
- w10n (webification) services



PDS Image Atlas (public)

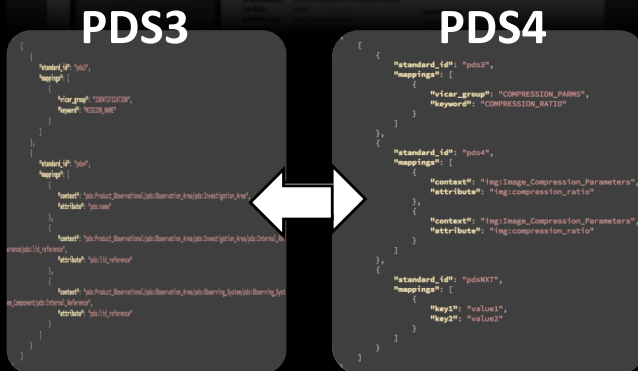
- Fine-grained product search by type, mission, name, location, constraints, features, etc.
- Apache Solr based RESTful API

Overview of Data Services



Photojournal Beta (public)

- Fine-grained *public-interest* product search by target, mission, instrument, etc.
- Apache Solr based RESTful API



Label Mapping Tool (private)

- Provides mapping services between PDS3 ODL keywords and PDS4 XML Xpaths
- Python Flask-based RESTful API

Overview of Data Services

Sawmill



splunk>

AMP

Metrics (private)

- Website metrics for Atlas and Photojournal
- Client & server logging metrics
- Data archive metrics and queries



Automation (private)

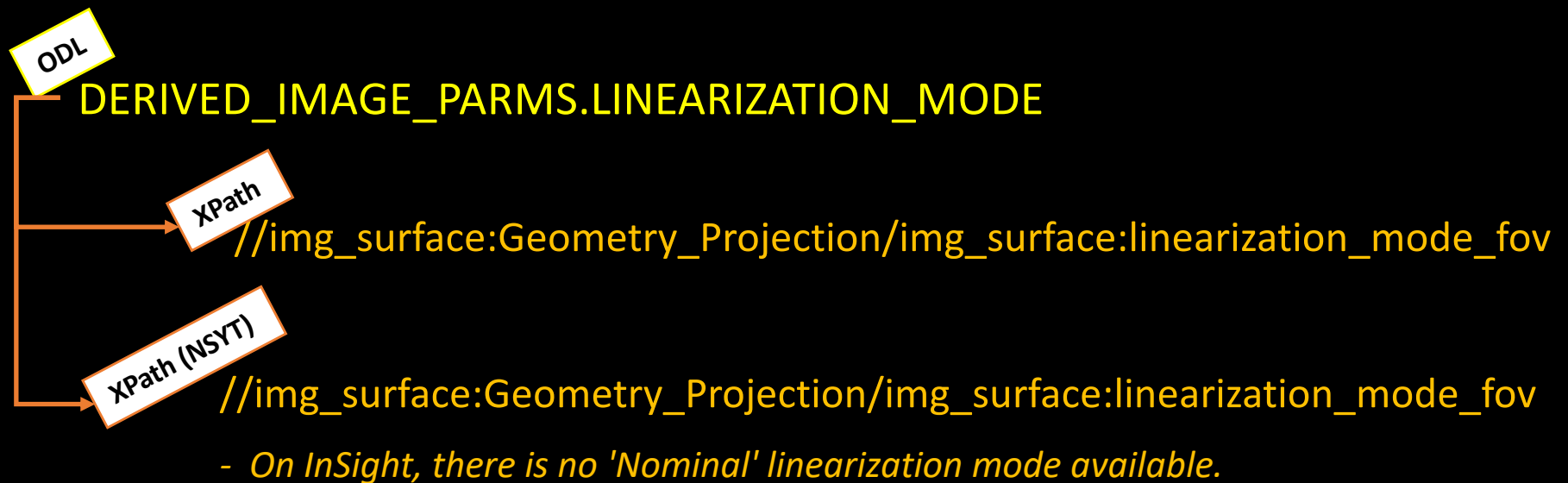
- Automatic software deployment, continuous integration, interactive volume/bundle release pipelines, code management, fault-tolerant computing
- HTTP(s) RESTful APIs

Case Studies



Label Mapping Tool (prototype)

What's the problem we're trying to solve?



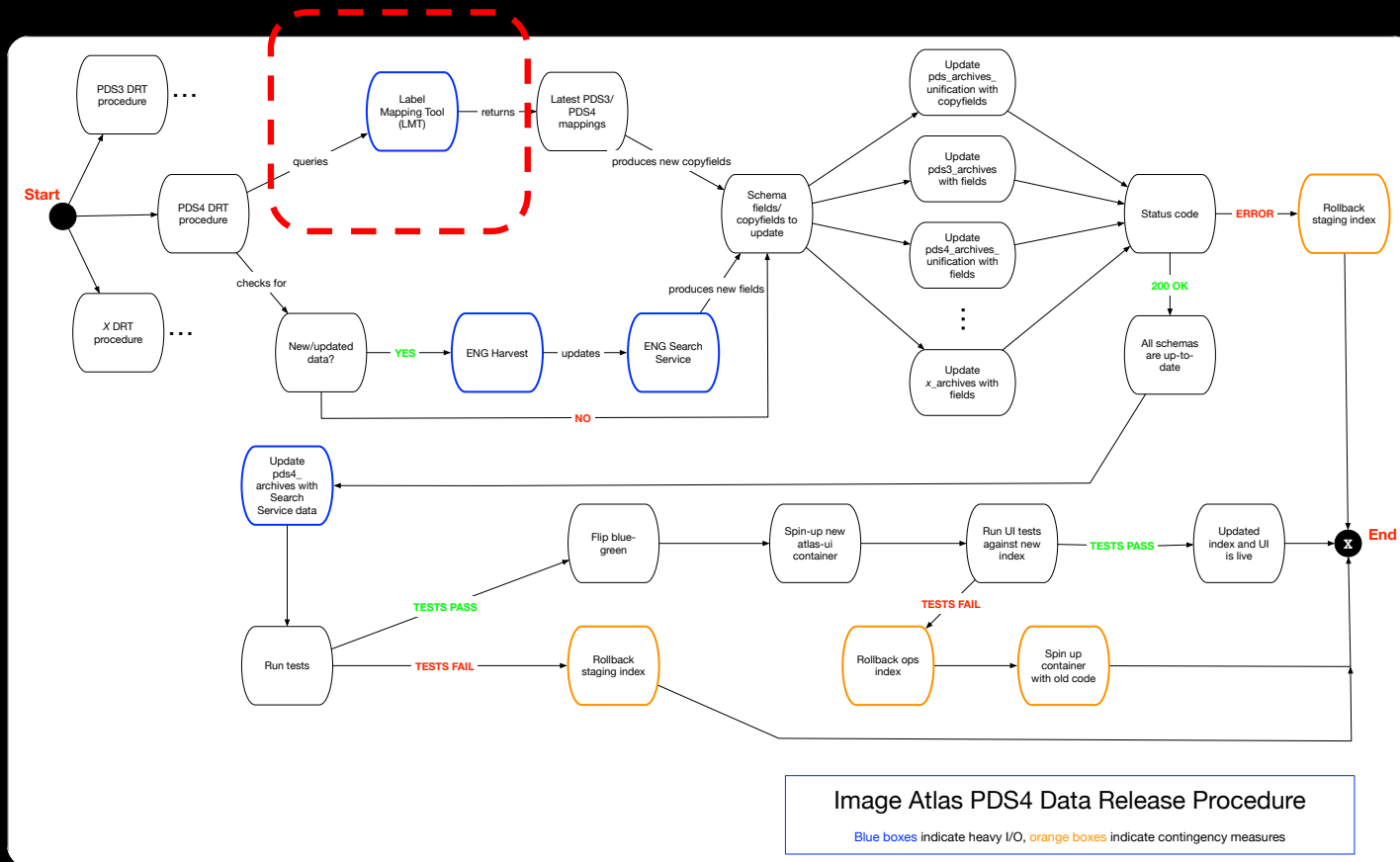
Label Mapping Tool (prototype)

Design Goals

Complexity Reduction in
PDS3 ODL ↔ PDS4 XPath
Mappings

Enable Automation by
Providing Mappings by way
of a Service Oriented
Architecture

Label Mapping Tool (prototype)



Credit: Kevin Grimes, JPL

Image Atlas PDS4 Data Release Procedure

Label Mapping Tool (prototype)

Data Flow

PDS4 Property Maps File
(current implementation)



Mapping Database



Web Service Wrapper



Client Queries



Documentation Queries

API Documentation

Label Mapping Tool (prototype)

Lessons Learned



Dependency on JSON Property Maps File was not originally intended for LMT use, and thus introduces artifacts that add complexity to LMT queries



Usage of automated API documentation generation (Swagger / Open API) reduces manual work in keeping data service documentation up-to-date

Label Mapping Tool (prototype)

Next Steps / Topics of Discussion:



Centrally host LMT on a publicly accessible database and provide a user interface to enable data modifications



Open source LMT code for greater ability to identify and respond to community needs



Discussion point: what is the intended relationship between LMT's database and the PDS4 Information Model?

Metrics

What's the problem we're trying to solve?

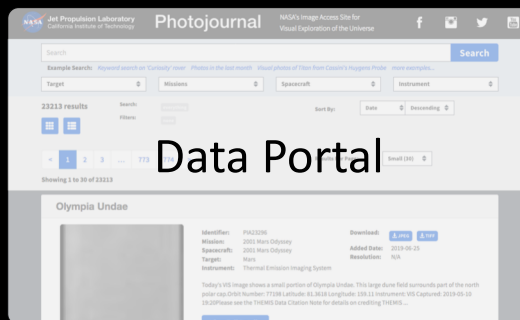


User interaction



Website User Metrics

- Behavioral analytics
- Query profiling
- Artifact popularity



Data Portal



Data Archive(s)



Server Operational Metrics

- Diagnosing system faults
- Download metrics

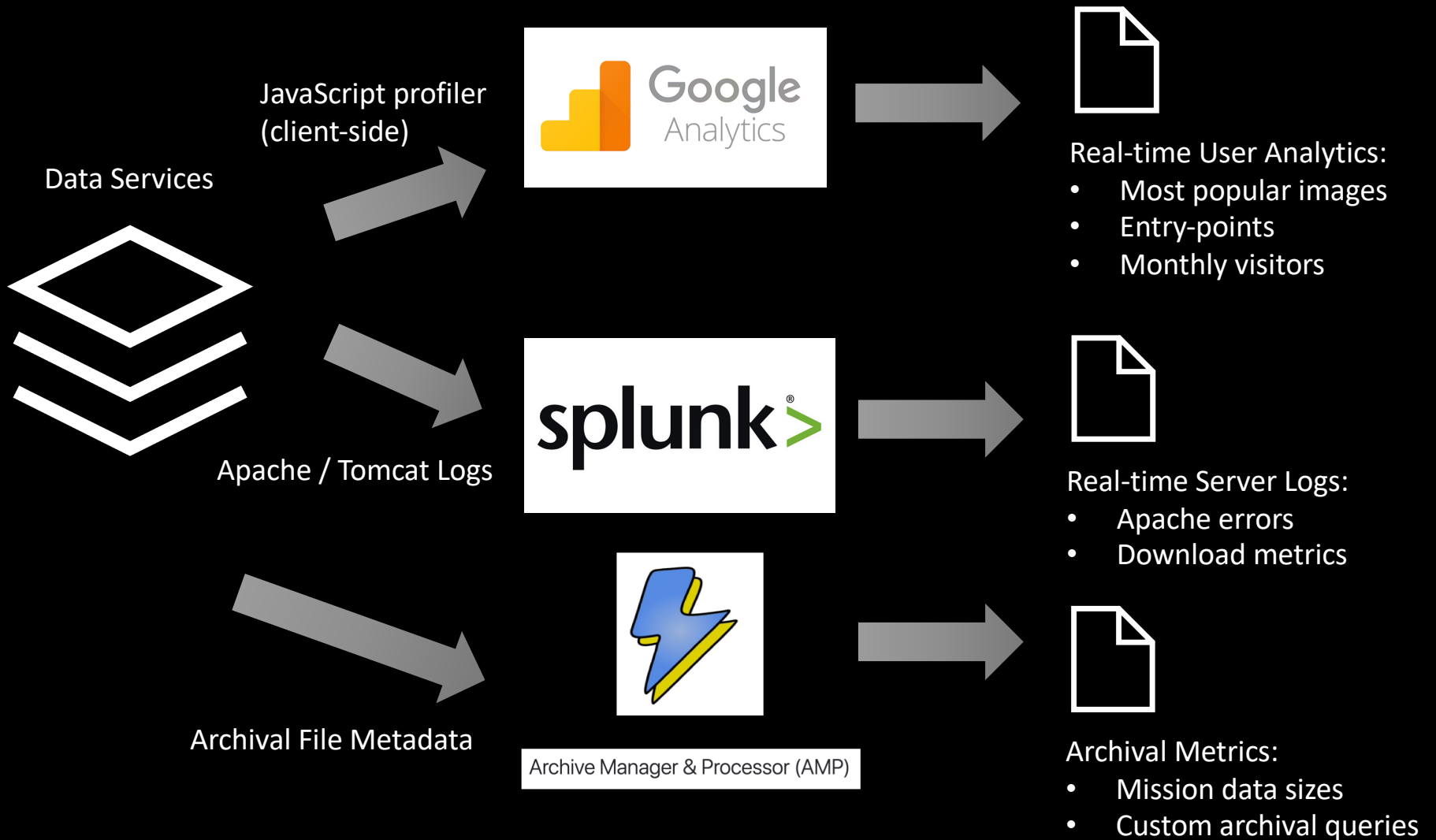


Archival Metrics

- De-duplication, identifying corruption etc.
- Mission-specific metrics

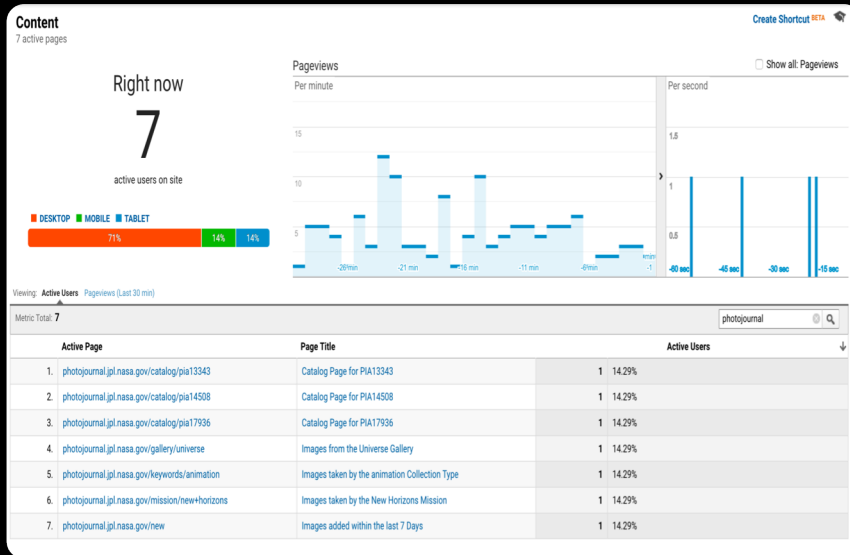
Metrics

Which services help us gather metrics?

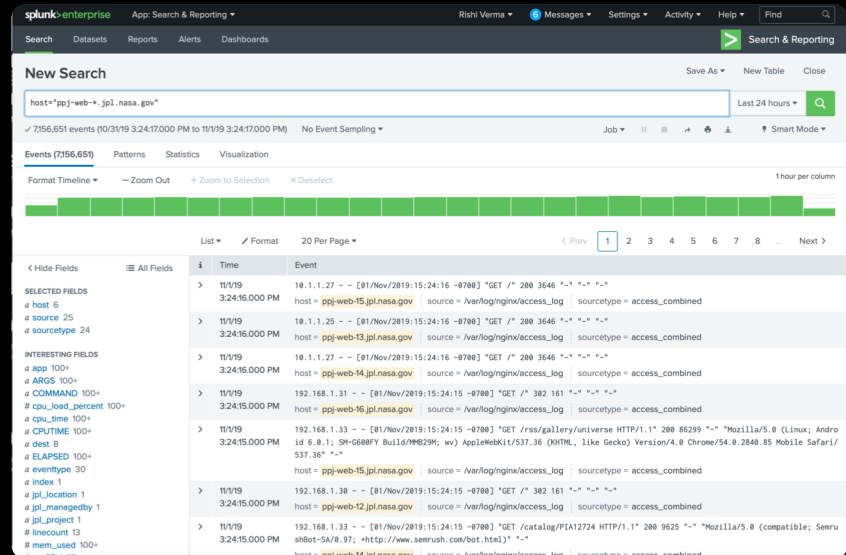


Metrics

PDS IMG Google Analytics Dashboard



PDS IMG Splunk Dashboard



Metrics

Lessons Learned



Correlating client-side data portal metrics and behavioral statistics with server-side logs is a hard problem. Ideally – we'd have an integrated system



Co-locating server-side logs with system status, database, automation, and other logs allows the possibility of faster diagnosing of faults and errors

Metrics

Next Steps / Topics of Discussion:



Alerts and contextual reports from Splunk based on real time data flows to users and overall system status



Centralized PDS dashboards on Google Analytics?



Forwarding logs outside IMG node to a central PDS-wide system is a good idea, but we need the ability to support fast, custom queries that can exclude / include custom metric parameters

¹<https://archive-manager-and-processor.github.io/>

Photojournal

<https://photojournal.jpl.nasa.gov/beta>

What data service problem are we attempting to solve?

- Allow users AND automated programs to search public-interest planetary surface images, videos, and media
- Support *faceted* searching
 - Targets
 - Missions
 - Spacecrafts
 - Instruments
- Provide a well-documented API

The screenshot displays the NASA Photojournal website interface. At the top, it features the NASA Jet Propulsion Laboratory logo and the text 'Photojournal NASA's Image Access Site for Visual Exploration of the Universe'. Below the header is a search bar with a 'Search' button. An example search is provided: 'Keyword search on 'Curiosity' rover Photos in the last month Visual photos of Titan from Cassini's Huygens Probe more examples...'. Below the search bar are four faceted search filters: 'Target', 'Missions', 'Spacecraft', and 'Instrument'. The main content area shows '23213 results' with a search filter set to 'everything' and a sort order of 'Date' descending. A pagination bar shows 'Showing 1 to 30 of 23213' with page numbers 1, 2, 3, ..., 773, 774. Below this, a specific result for 'Olympia Undae' is displayed. It includes a thumbnail image of the dune field, a metadata table, and a 'View Full Details' button.

Identifier:	PIA23296	Download:	JPG TIFF
Mission:	2001 Mars Odyssey	Added Date:	2019-06-25
Spacecraft:	2001 Mars Odyssey	Resolution:	N/A
Target:	Mars		
Instrument:	Thermal Emission Imaging System		

Today's VIS image shows a small portion of Olympia Undae. This large dune field surrounds part of the north polar cap. Orbit Number: 77198 Latitude: 81.3618 Longitude: 159.11 Instrument: VIS Captured: 2019-05-10 19:20 Please see the THEMIS Data Citation Note for details on crediting THEMIS ...

Photojournal

<https://photojournal.jpl.nasa.gov/beta>

Example: *Find visual photographs of Titan from Cassini's Huygens Probe*

From a web-browser



<https://photojournal.jpl.nasa.gov/beta/?search=insight+arm#q=Cassini%20AND%20Titan%20AND%20Huygens%20AND%20Visual>

Photojournal

<https://photojournal.jpl.nasa.gov/beta>

Example: *Find visual photographs of Titan from Cassini's Huygens Probe*

From a RESTful Query:



<https://solr.apache.org>

<https://photojournal.jpl.nasa.gov/beta/solr/photojournal/select?q=Cassini%20AND%20Titan%20AND%20Huygens%20AND%20Visual&facet=true&facet.field=target&facet.field=mission&facet.field=spacecraft&facet.field=instrument&facet.limit=-1&sort=record-modification-time%20desc&rows=30&wt=xml>

Photojournal

Lessons Learned



Solr API is powerful, but requires additional documentation beyond third-party supporting documentation for users to be motivated to use API services with ease

Photojournal

Next Steps / Topics of Discussion:

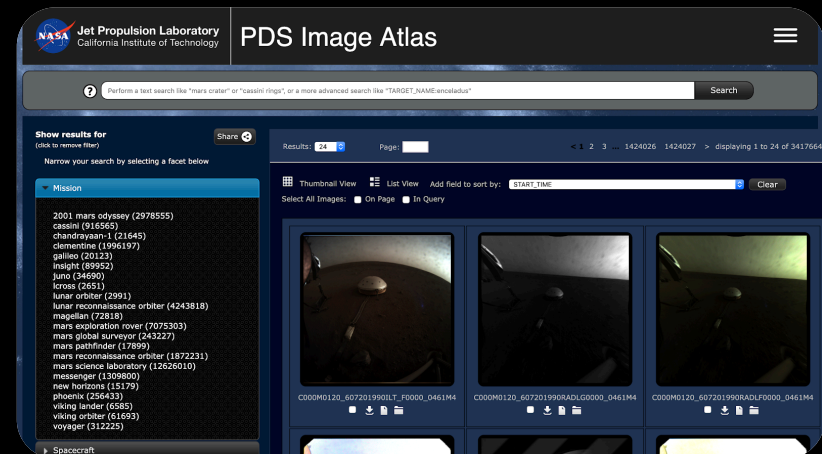


Solr supports the ability to translate search queries via “endpoint” modules – this can interoperate with a global PDS search query syntax while preserving localized, specialty searches

PDS Image Atlas

<https://pds-imaging.jpl.nasa.gov/search>

- Search images from over 20 NASA missions
- Fine-grained search of low-level image metadata
 - Product type
 - Target name
 - Location (latitude/longitude, etc.)
 - Time constraints (sclk, date)
- Search by dynamic features in images
 - Craters (ML)
 - Wheels
- Accessible via web UI or Solr API



PDS Image Atlas

<https://pds-imaging.jpl.nasa.gov/search>

- Search powered by Apache Solr
 - Open source search platform from Apache Foundation
 - Built on Apache Lucene
- Full support for Solr query syntax
 - TARGET:mars^[1]
 - MISSION:"mars science laboratory" START_TIME:[2019-03-22T00:00:00Z TO 2019-03-25T00:00:00Z]^[2]
 - MSL_IMAGE_CLASS:wheel^[3]
 - MRO_IMAGE_CLASS:crater MRO_IMAGE_CLASS:"swiss cheese"^[4]
- API endpoint at <https://pds-imaging.jpl.nasa.gov/solr>
- API documentation at <https://pds-imaging.jpl.nasa.gov/tools/atlas/api/>



<https://solr.apache.org>

Webification (w10n)

<https://pds-imaging.jpl.nasa.gov/w10n>

- File access via RESTful API
- Programmatic access of:
 - GIF versions of VICs/IMGs
 - Subsets of raw raster data
 - Resized/cropped products
- Examples
 - .../NRB_606845760ILTLF0750750NCAM00592M1.IMG/0
/
image[]?output=gif^[4]
 - .../NRB_606845760ILTLF0750750NCAM00592M1.IMG/0
/
resized/64x48[]?output=gif^[5]
 - .../NRB_606845760ILTLF0750750NCAM00592M1.IMG/0
/

Next Steps:

- Open source more data service software: e.g. LMT
- Application Programming Interfaces (API)
 - More formalized guide for developer community
 - API docs for common languages: Javascript, Python, etc.
 - Python based libraries for easier programmatic access to PDS archival data
- Enhanced metadata services
 - Automated caption generation using machine learning
 - Surface feature identification and coordinates

Closing Thoughts

- Need for supporting community service interoperability
 - Search, metadata services, and custom services like LMT
 - Consistent set of RESTful service APIs following a common standard
- Need for developing an information model for “enhanced” product metadata, e.g. generated by machine learning algorithms
- RESTful APIs are often not enough, programming-language specific APIs in Javascript, Python, etc. are highly desired
- Federated vs centralized is a distinction we should abstract:
 - Users care about *speed* and *consistency*
 - e.g. heterogenous data centers in the Internet service model

Acknowledgements / Contributions

Slide contributions

- Kevin Grimes
- Cristina De Cesare
- Anna Waldron
- Jordan Padams
- Paul Ramirez
- Galen Hollins

Other contributors

- Anna Waldron
- Cristina De Cesare
- Rafael Alanis
- Paul Ramirez
- Myche McAuley

Backup Slides

PDS API

- PDS API
 - Vision: Construct a PDS API that integrates individual node API capabilities to answer cross cutting questions and/or provide integrated capabilities.
 - Co-develop (i.e. with several node staff) the API across nodes guided/contributed by EN
 - Focus on integrating Node API's but abstract for user the need to know the architectural organization of PDS
 - Consider if this development occurs as a PDS EN incubator project or attaches to a community effort such as PlanetaryPy which aims to be modeled after Astropy (astropy.org)
 - Facilitate use case generation, adoption, and contributions through community discussion on OpenPlanetary Forum (forum.openplanetary.org) as early as possible
 - Design based on a sample Python client API that users could use to in their own analysis code
 - Example: Submit a single search across for all products from Cassini. Once results are returned capabilities can be provided on certain products (e.g. imaging products can have subsetting capabilities via w10n)

PDS Python API

```
import pds

# Returns results that works across nodes for different
# product types?
results = pds.search(mission_name="Cassini")
print result[0].identifier()

# Build types to support capabilities on Images?
results = pds.search(product_type=pds.Image,
mission_name="Cassini")

# Subsetting that uses w10n?
image = results[0]
image.subset(...)
```

Sample Solr queries

TARGET:mars

The screenshot shows the NASA PDS Image Atlas search results for the query "TARGET:mars". The page header includes the NASA logo, "Jet Propulsion Laboratory California Institute of Technology", and "PDS Image Atlas". A search bar contains the query "TARGET:mars". The results are displayed in a grid of thumbnails. The left sidebar shows a list of missions, with "2001 mars odyssey (2978555)" selected. The main content area shows three thumbnails of the Mars surface, with a tooltip indicating "Click on image to display higher resolution." and a download icon. The page number is 1 of 24, and the total number of results is 25227906.

Jet Propulsion Laboratory
California Institute of Technology

PDS Image Atlas

SEARCH: TARGET:mars

Show results for TARGET:mars (click to remove filter) Share

Results: 24 Page: 1 < 1 2 3 ... 1051162 1051163 > displaying 1 to 24 of 25227906

Thumbnail View List View Add field to sort by: START_TIME Clear

Select All Images: On Page In Query

Mission

- 2001 mars odyssey (2978555)
- insight (89952)
- mars exploration rover (7075303)
- mars global surveyor (243227)
- mars pathfinder (17899)
- mars reconnaissance orbiter (1872231)
- mars science laboratory (12626010)
- messenger (18)
- phoenix (256433)
- viking lander (6585)
- viking orbiter (61693)

Spacecraft
Instrument
Target
Product Type
Lighting Geometry
Filters

Click on image to display higher resolution.

0M0120_607201990RADLG0000_0461M4
C000M0120_607201990RADLF0000_0461M4

Sample Solr queries

MISSION:"mars science laboratory"

START TIME:[2019-03-22T00:00:00Z TO 2019-03-

The screenshot shows the NASA Jet Propulsion Laboratory PDS Image Atlas search interface. The search query is: MISSION:"mars science laboratory" START TIME:[2019-03-22T00:00:00Z TO 2019-03-25T00:00:00Z]. The results are displayed in a table with columns for Start Time and Mission Name. The first result is a thumbnail image of the Mars Science Laboratory rover on the surface, with a start time of 2019-03-24T23:35:52Z and mission name Mars Science Laboratory. The interface includes a sidebar with filters, a search bar, and a table of results.

Search Results Summary:

Start Time	Mission Name
2019-03-24T23:35:52Z	Mars Science Laboratory

Search Query: MISSION:"mars science laboratory" START TIME:[2019-03-22T00:00:00Z TO 2019-03-25T00:00:00Z]

Filters: Mission (mars science laboratory (21847)), Spacecraft, Instrument, Target, Product Type, Lighting Geometry, Filters, Lat/Lon Bounding Box, Time Constraints, Orbital Mission Constraints, Landed Mission Constraints, PDS Archive Constraints.

Table Headers: Start Time, Mission Name


Table Content:

Start Time	Mission Name
2019-03-24T23:35:52Z	Mars Science Laboratory

Image ID: NLB_606740453EDR_F0750750NCAM00354M1

Sample Solr queries

MSL_IMAGE_CLASS:

 **Jet Propulsion Laboratory**
California Institute of Technology

PDS Image Atlas

Perform a text search like "mars crater" or "cassini rings", or a more advanced search like "TARGET_NAME:enceladus"

Show results for
(click to remove filter)

remove all
(x) MSL_IMAGE_CLASS:wheel
(x) MSL_CONFIDENCE:[0.9 TO 1]

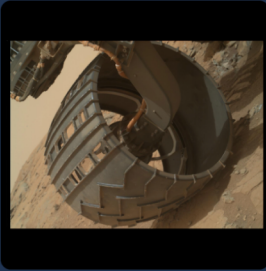



Mission

- mars science laboratory (9150)
- Spacecraft
- Instrument
- Target
- Product Type
- Lighting Geometry
- Filters
- Lat/Lon Bounding Box
- Time Constraints
- Orbital Mission Constraints
- Landed Mission Constraints
- PDS Archive Constraints

Results: 24 Page: < 1 2 ... 364 365 366 ... 381 382 > displaying 8737 to 8760 of 9150


Thumbnail View List View Add field to sort by: START_TIME

Select All Images: On Page In Query

	Start Time ↑ ↓	Mission Name ↑ ↓	Msl Confidence ↑ ↓
 0275MH000259000102983E01_XXXX   	2013-05-15T15:22:55.796Z	Mars Science Laboratory	0.997

Sample Solr queries

MSL_IMAGE_CLASS:

 **Jet Propulsion Laboratory**
California Institute of Technology

PDS Image Atlas

Perform a text search like "mars crater" or "cassini rings", or a more advanced search like "TARGET_NAME:enceladus"

Show results for
(click to remove filter)

remove all
(x) MSL_IMAGE_CLASS:wheel
(x) MSL_CONFIDENCE:[0.9 TO 1]





Mission

- mars science laboratory (9150)
- Spacecraft
- Instrument
- Target
- Product Type
- Lighting Geometry
- Filters
- Lat/Lon Bounding Box
- Time Constraints
- Orbital Mission Constraints
- Landed Mission Constraints
- PDS Archive Constraints

Results: 24 Page: < 1 2 ... 54 55 56 ... 381 382 > displaying 1297 to 1320 of 9150

Thumbnail View List View Add field to sort by: START_TIME

Select All Images: On Page In Query

	Start Time ↑ ↓	Mission Name ↑ ↓	Msl Confidence ↑ ↓
 1682ML0087470020700819E03_XXXX   	2017-04-30T09:15:58.512Z	Mars Science Laboratory	0.987

Sample Solr queries

The screenshot displays the NASA Jet Propulsion Laboratory's MRO HIRISE Image Content interface. The interface is divided into several sections:

- Header:** NASA logo, Jet Propulsion Laboratory, California Institute of Technology.
- Search Bar:** A search input field with a search button.
- Filters:** A list of filter categories on the left side, including Mission, Spacecraft, Instrument, Target, Product Type, Lighting Geometry, Filters, Lat/Lon Bounding Box, Time Constraints, Orbital Mission Constraints, Landed Mission Constraints, PDS Archive Constraints, and Advanced Constraints. The "MRO HIRISE Image Content" filter is currently selected.
- Image Content:** A central panel showing a grayscale image of the Martian surface with numerous red bounding boxes highlighting detected craters. The "Image Content" panel has a "Crater" filter selected.
- Results:** A table on the right side showing the results of the search. The table has a "Crater Counts" column and a "Crater" column. The table shows 86 craters for the "Orbiter" mission.

The "Crater Counts" table is as follows:

Crater Counts	Crater
86	Orbiter

Sample Photojournal Queries

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
--<response>
- <lst name="responseHeader">
  <int name="status">0</int>
  <int name="QTime">53</int>
- <lst name="params">
  <str name="q">Cassini AND Titan AND Huygens AND Visual</str>
  <str name="facet.limit">-1</str>
- <arr name="facet.field">
  <str>target</str>
  <str>mission</str>
  <str>spacecraft</str>
  <str>instrument</str>
</arr>
  <str name="sort">record-modification-time desc</str>
  <str name="rows">30</str>
  <str name="facet">>true</str>
  <str name="wt">xml</str>
</lst>
</lst>
- <result name="response" numFound="89" start="0">
- <doc>
  <date name="publication-date">2018-09-24T16:34:57Z</date>
- <arr name="alt-tag">
  - <str>
    This frame from an animation shows clear bright spots on Saturn's moon, Titan, that have been interpreted as evidence of dust storms.
  </str>
</arr>
- <arr name="keywords">
  <str>ipcat-saturn</str>
  <str>titan</str>
  <str>animation</str>
</arr>
```

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 29,
    "params": {
      "q": "Cassini AND Titan AND Huygens AND Visual",
      "facet.limit": "-1",
      "facet.field": [
        "target",
        "mission",
        "spacecraft",
        "instrument"
      ],
      "sort": "record-modification-time desc",
      "rows": "30",
      "facet": "true",
      "wt": "json"
    }
  },
  "response": {
    "numFound": 89,
    "start": 0,
    "docs": [
      {
        "publication-date": "2018-09-24T16:34:57Z",
        "alt-tag": [
          "This frame from an animation shows clear storms."
        ]
      }
    ]
  }
}
```


Sample Photojournal Queries

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 29,
    "params": {
      "q": "Cassini AND Titan AND Huygens AND Visual",
      "facet.limit": "-1",
      "facet.field": [
        "target",
        "mission",
        "spacecraft",
        "instrument"
      ],
      "sort": "record-modification-time desc",
      "rows": "30",
      "facet": "true",
      "wt": "json"
    }
  },
  "response": {
    "numFound": 89,
    "start": 0,
    "docs": [
      {
        "publication-date": "2018-09-24T16:34:57Z",
        "alt-tag": [
          "This frame from an animation shows clear bright spots on Saturn's moon, Titan, that have been interpreted as evidence of dust storms."
        ],
        "keywords": [
          "ipcat-saturn",
          "titan",
          "animation"
        ],
        "z-dim": 3,
        "browse-size": [
          "256x256"
        ]
      }
    ]
  }
}
```