PDS Cartography and Imaging Sciences Node

Jet Propulsion Laboratory

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

PDS

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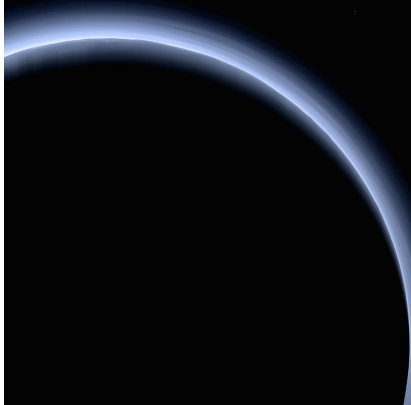
Data characteristics:
- Orbital vs. landed missions
- Original raw experiment data vs. derived products
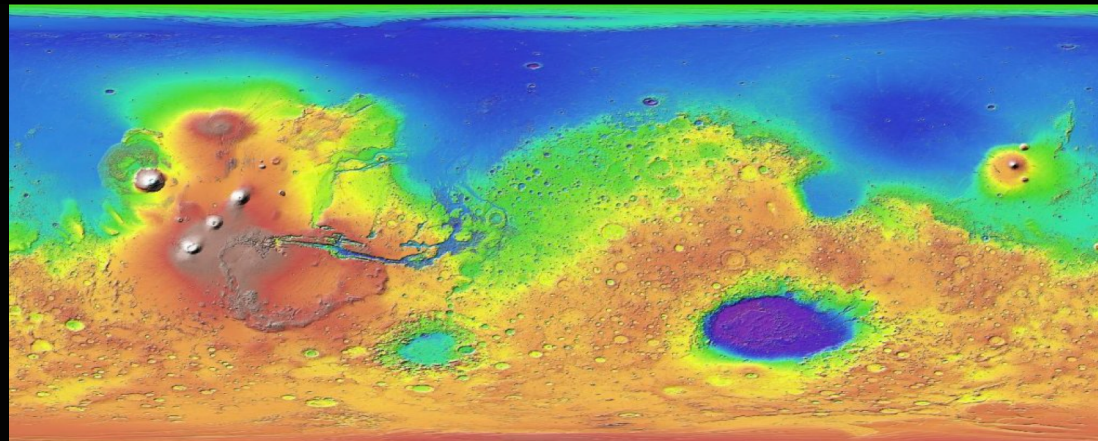- Differing coordinate systems

Users:
- Scientists
- Public
- Developers

New Horizons spacecraft: Plutonian Atmosphere

Mars Global Surveyor: Mola Global Colorized Hillshade

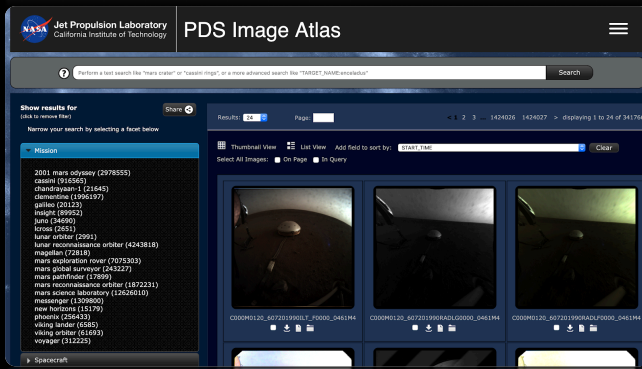Curiosity Rover Hazcam Image of Drilling on Jan. 13, 2015

Overview of Data Services
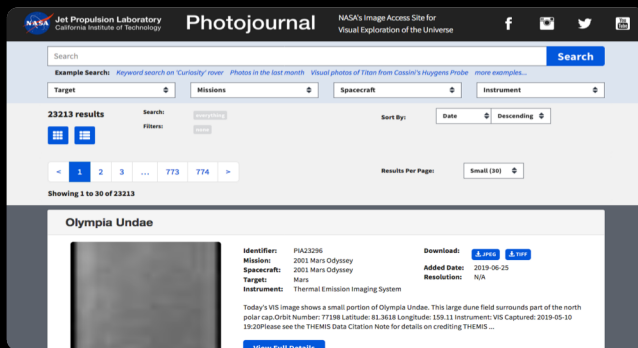

Data Archive Portal (public)

- **HTTP(s) access to JPL-based 70+ TB of data**
- **w10n (webificiation) services**
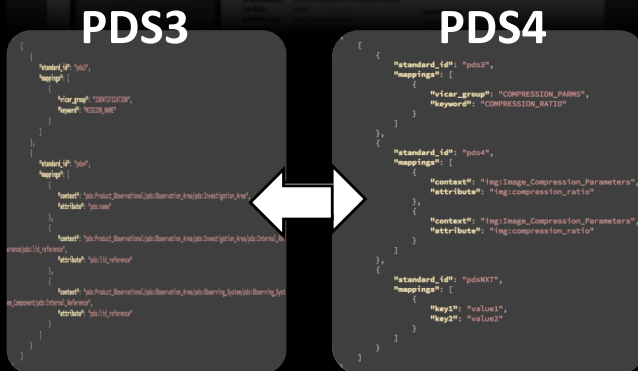
PDS Image Atlas (public)

- **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

Google Analytics

splunk> **AMP**

XebiaLabs RELEASE

ANSIBLE

GitHub   jupyter   kubernetes

Metrics (private)
- **Website metrics for Atlas and Photojournal**
- **Client & server logging metrics**
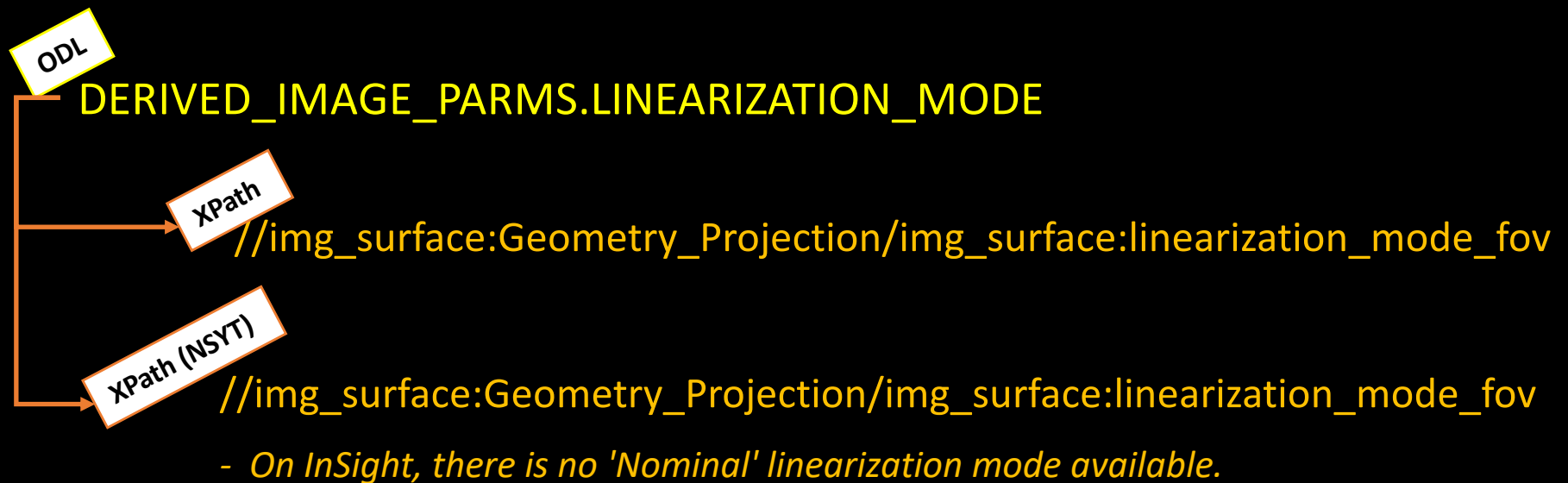- **Data archive metrics and queries**

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

PDS Cartography and Imaging Sciences Node
Jet Propulsion Laboratory   U.S. Geological Survey

Case Studies

Label Mapping Tool (prototype)

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

ODL

DERIVED_IMAGE_PARMS.LINEARIZATION_MODE

XPath

//img_surface:Geometry_Projection/img_surface:linearization_mode_fov

XPath (NSYT)

//img_surface:Geometry_Projection/img_surface:linearization_mode_fov

- *On InSight, there is no 'Nominal' linearization mode available.*

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)

Image Atlas PDS4 Data Release Procedure

Blue boxes indicate heavy I/O, orange boxes indicate contingency measures
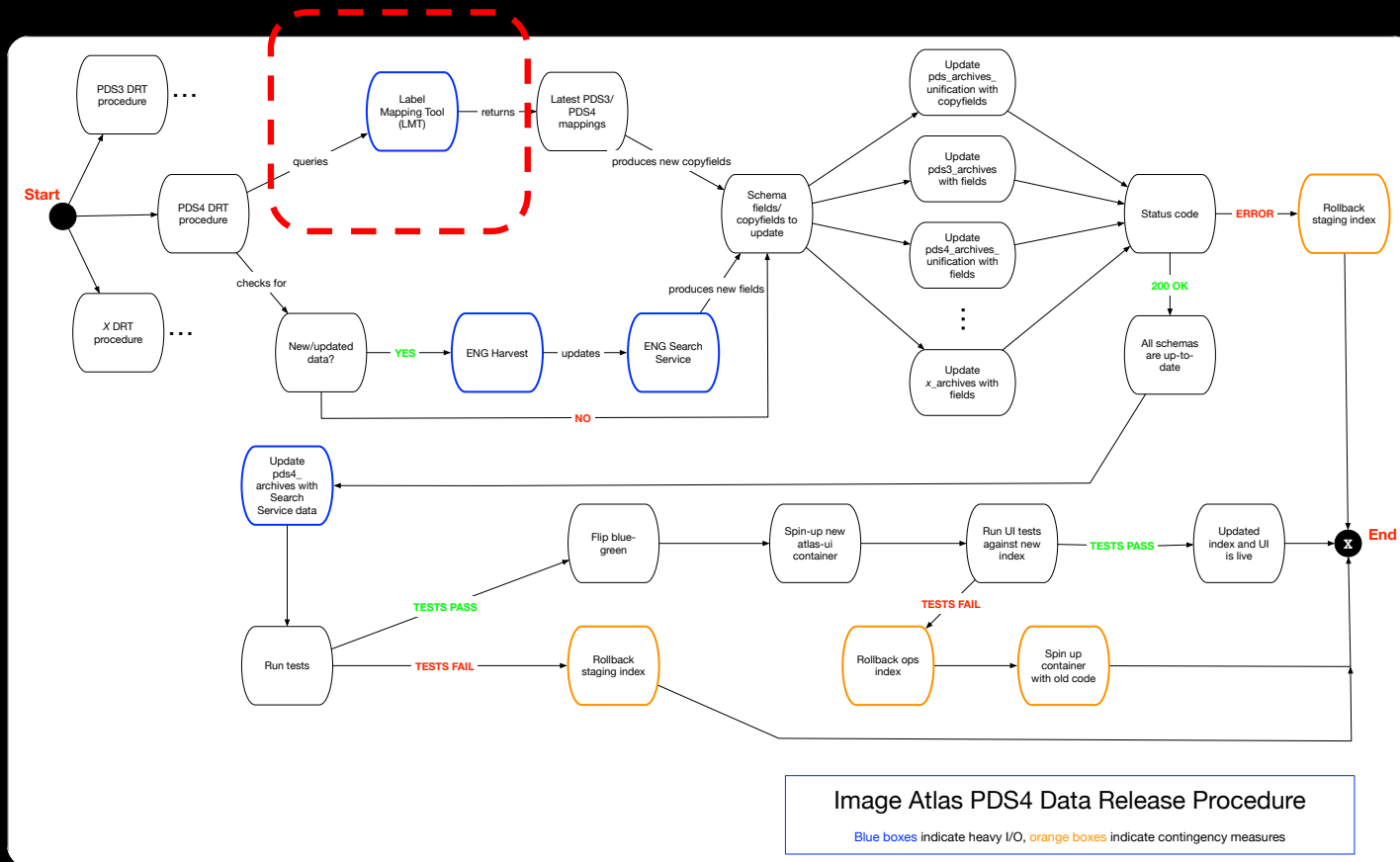
Credit: Kevin Grimes, JPL

Label Mapping Tool (prototype)

Data Flow

PDS4 Property Maps File
(current implementation)

Mapping Database

Web Service Wrapper

Client Queries

API Documentation

Documentation Queries

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

Data Portal

Website User Metrics
- Behavioral analytics
- Query profiling
- Artifact popularity
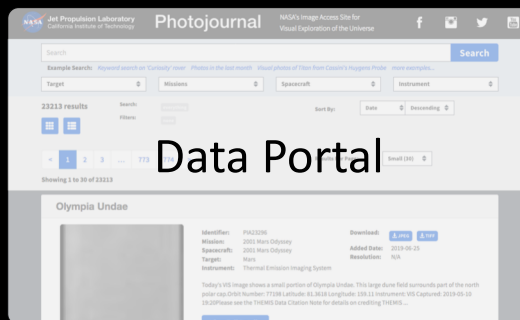
Server Operational Metrics
- Diagnosing system faults
- Download metrics

Data Archive(s)

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

Metrics

Which services help us gather metrics?

Data Services

JavaScript profiler (client-side)

Google Analytics

Real-time User Analytics:
- Most popular images
- Entry-points
- Monthly visitors

Apache / Tomcat Logs

splunk>

Real-time Server Logs:
- Apache errors
- Download metrics

Archival File Metadata

Archive Manager & Processor (AMP)

Archival Metrics:
- Mission data sizes
- Custom archival queries
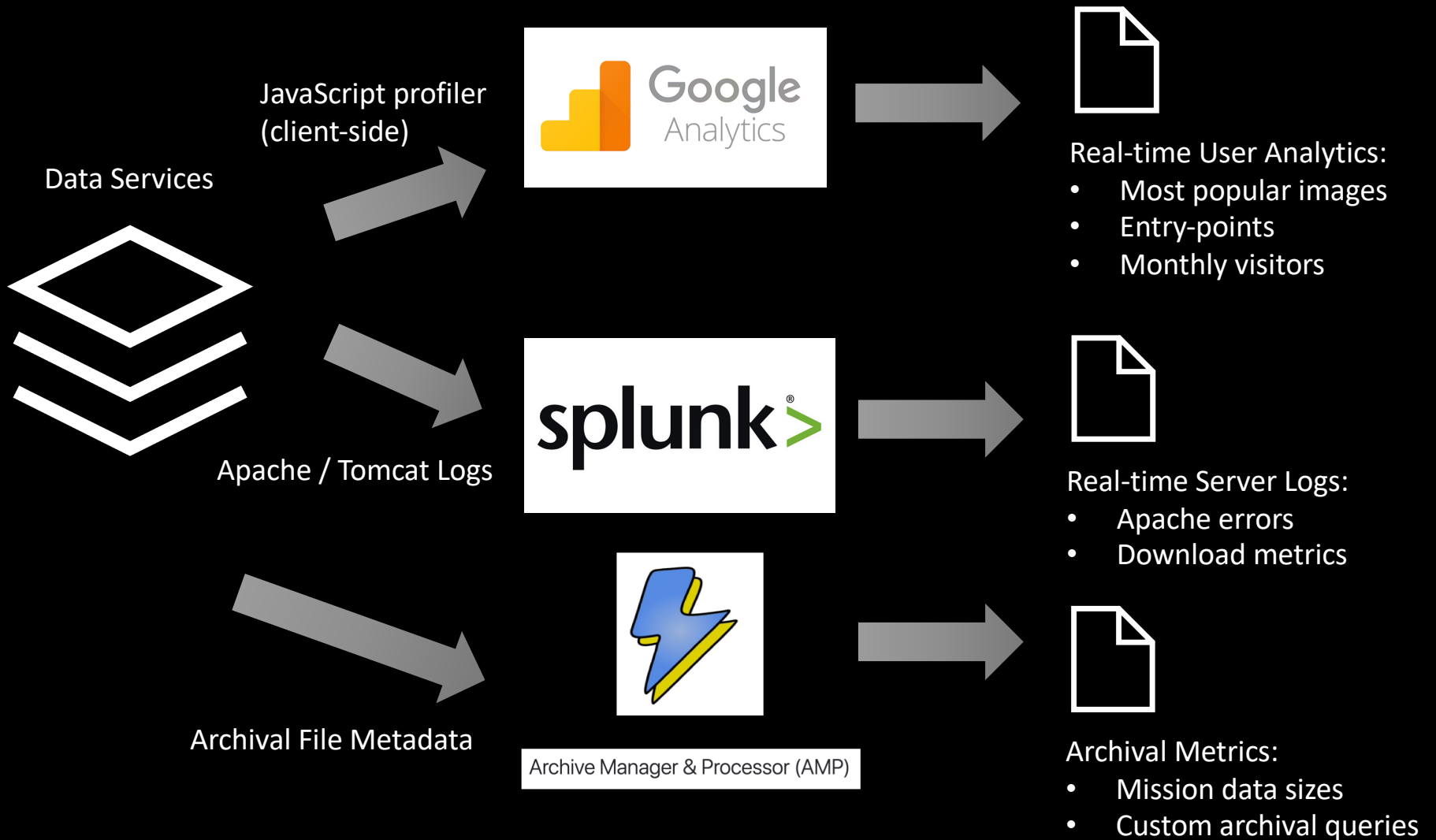
Metrics

PDS IMG Google Analytics Dashboard
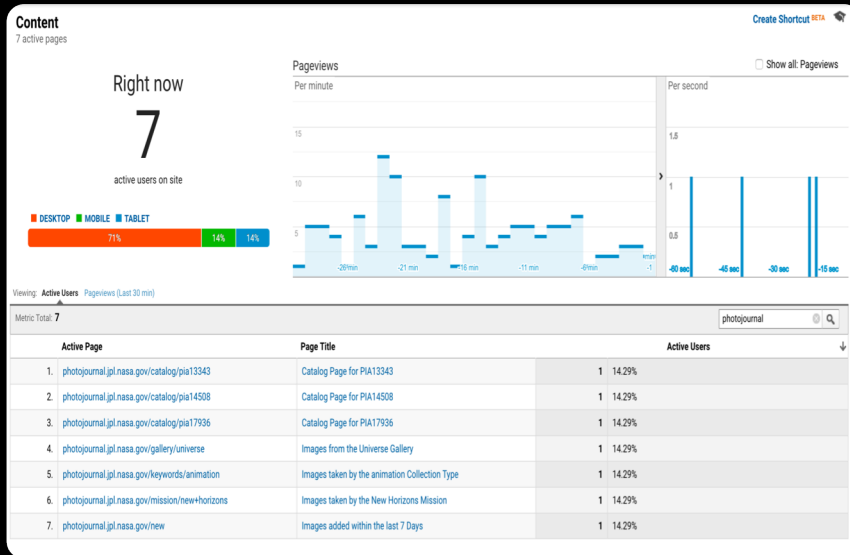
PDS IMG Splunk Dashboard
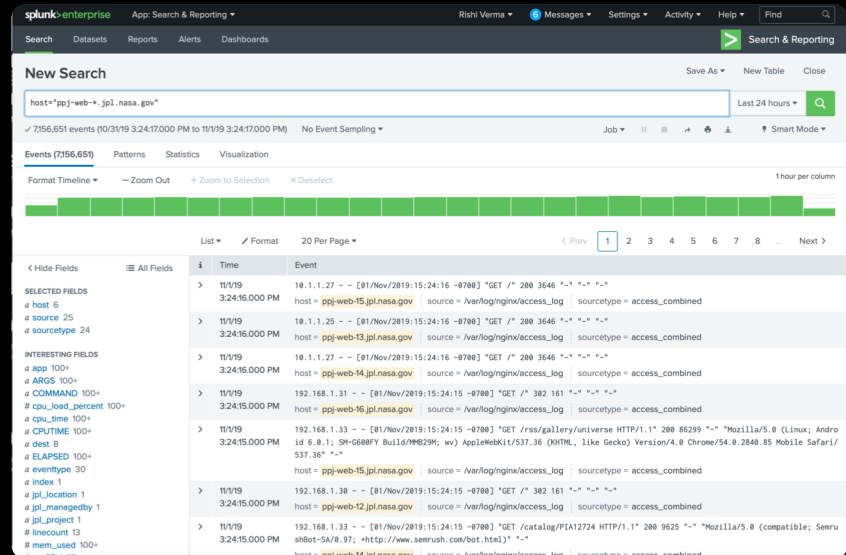
Metrics

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

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

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
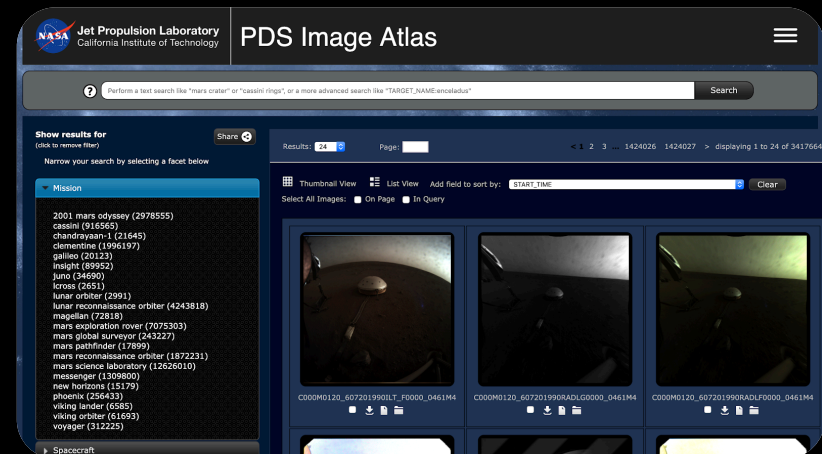
- Search powered by Apache Solr
    - Open source search platform from Apache Foundation
    - Built on Apache Lucene
- Full support for Solr query syntax
    - TARGET:mars[1]
    - MISSION:"mars science laboratory" START_TIME:[2019-03-22T00:00:00Z TO 2019-03-25T00:00:00Z][2]
    - MSL_IMAGE_CLASS:wheel[3]
    - 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)

- 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 capablities via w10n)

PDS Python API

```python
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

Sample Solr queries

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

Sample Solr queries

MSL_IMAGE_CLASS:

Sample Solr queries

MSL_IMAGE_CLASS:

Sample Solr queries

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 dus
          </str>
        </arr>
      - <arr name="keywords">
          <str>ipcat-saturn</str>
          <str>titan</str>
          <str>animation</str>
        </arr>
```

```json
{
    "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"
```

Raw | Parsed