Open Data Federation

Prepared for LSDA Winter Meeting 2016

Jason Duley   2/12/2016

OCIO T&I Division/ARC

jason.duley@nasa.gov

Topics

- About the HQ OCIO Open Innovation Team
- What is Open Data
- Open Data Federation
- Data.nasa.gov
- Data.gov
- Work in progress
- Questions

About

- Part of HQ OCIO Technology & Infusion Division (T&I)

- The Open Innovation Team has the charter to:
    - Implement OMB Mandates on Open Data
        - M-13-13
    - Create and maintain an Open Data Registry
    - Standup and continue to build a robust API program
    - Other Activites

https://open.nasa.gov/open-data/

What is Open Data

- NASA Missions/Projects/Programs create and publish data making it available to the public as Open Data

  - no ITAR, SBU, PII is made available

- Open Data is made available on domain specific sites (e.g.: GeneLab, EOSDIS, NASA3D) as the authoritative source

- Data is now available for use in academia, STEM, citizen science projects, Hackathons, entrepreneurship, etc.

https://spaceappschallenge.org

Open Data Federation

- NASA maintains a Public Data Listing commonly referred to as the PDL
- The PDL is a machine-readable metadata descriptor that contains "pointers" to existing NASA datasets
- Datasets are described using a federal government-wide standard - Project Open Data v1.1
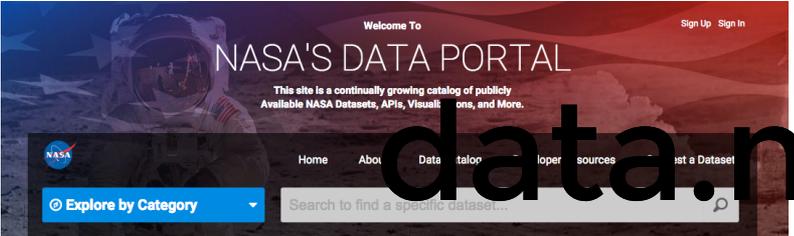 - https://project-open-data.cio.gov/v1.1/schema

Open Data Federation
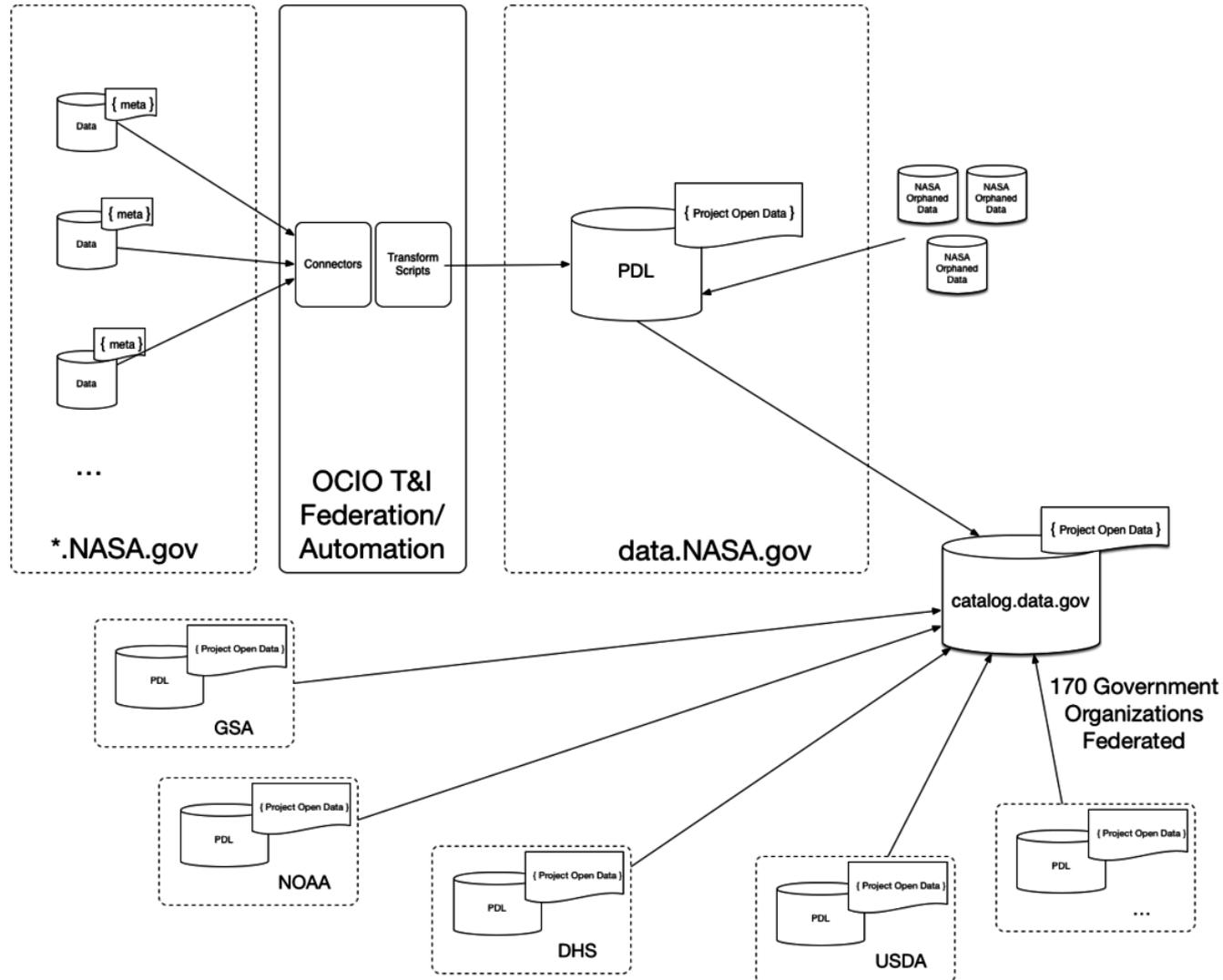
data.nasa.gov

- NASA Open Data Portal provides:

  - A dataset metadata registry
  - Dataset hosting
  - API Wrappers/API Key Infrastructure
  - Visualization
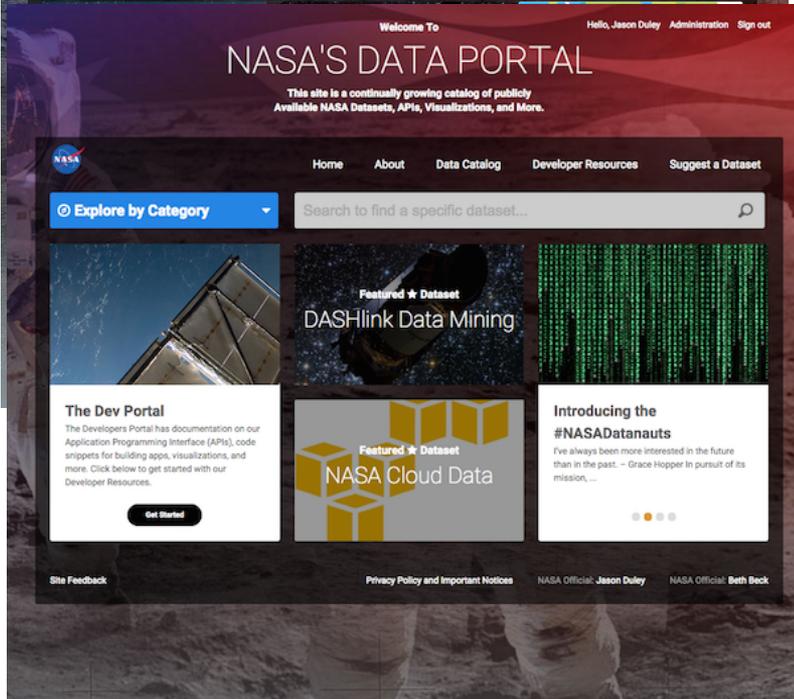 - Citizen Engagement / Feedback

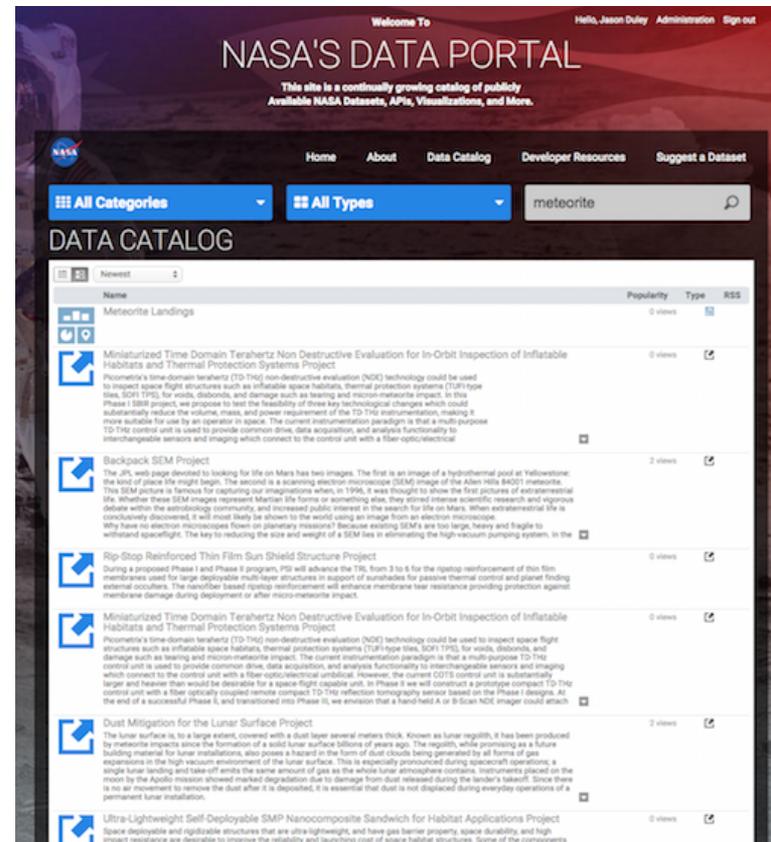https://data.nasa.gov

data.nasa.gov

- data.nasa.gov home page
  - Data Catalog
  - Developer Portal
  - Data Success Stories

- Data Catalog
  - Search by Keywords
  - Filter by Categories/Types
    - Maps
    - Charts
    - APIs
    - Datasets

Quick Demo

data.gov

- Data.gov is a GSA led effort to federate open data across the federal government
- Contains metadata from over 170 federal agencies
- Federated PDLs are consolidated at catalog.data.gov

https://www.data.gov/about

http://catalog.data.gov/dataset

Work in progress

- Our team is leading a Data Science effort to created re-usable services leveraging machine learning and natural language processing technologies
- Two services currently under development to be integrated into data.nasa.gov include:
    - Keyword Generation
    - Creation of Dataset Relevancy
- Goal is to enable the capability end of FY16

Thank you!
Questions
& Discussion

Backup

API Catalog

- One stop shop to get started using NASA APIs
- Promotes citizen contribution

- Site dedicated to NASA APIs
- Sign up for developer keys
- Live code editor integrated with site to experiment with API
- Templates for API documentation

https://api.nasa.gov

API Program

- Why API Management is needed
    - Increasing number of APIs
    - Finer-grained resource allocation
    - API access control
- Deployment vs. Management
    - mgmt layer sits in front of deployment (e.g.: micro-service)
- Features
    - Caching, Rate Limiting, API Keys
    - Provides hooks for analytics