



# A Tour of NASA's Data Universe for a Space-Apps Audience

**Justin Gosses**

S.A.I.C. senior data scientist supporting  
Office of the Chief Information Officer  
Transformation & Data Division



# Contents of this talk

The premise of this talk is that by telling you a little about why different open-data sites exist and how they relate to one another, you'll be better prepared to find datasets.

1. Introduction to open-innovation data sites (nasa-wide data aggregators)
2. The NASA Data Universe:
  - i. Government Mandates
  - ii. Harvesting Relationships
  - iii. The range of sites: Open-innovation program, Science Archives, Others
3. Tips
4. Examples of finding datasets

# Open-Innovation Program

Run by Office of the Chief Information Officer (OCIO),  
Open-Government Mandate Driven,  
& Agency-wide

*and what I help support*

# Open-innovation program

A series of websites acting as the agency-wide collecting place for NASA's public data

including:

- a. **API.nasa.gov**
- b. **Code.nasa.gov**
- c. **Data.nasa.gov**
- d. **Open.nasa.gov** = General open-gov information & datanauts (an outreach program)

Who is this site for?



Citizen Scientist



Developer



Citizen Activist



Govvie

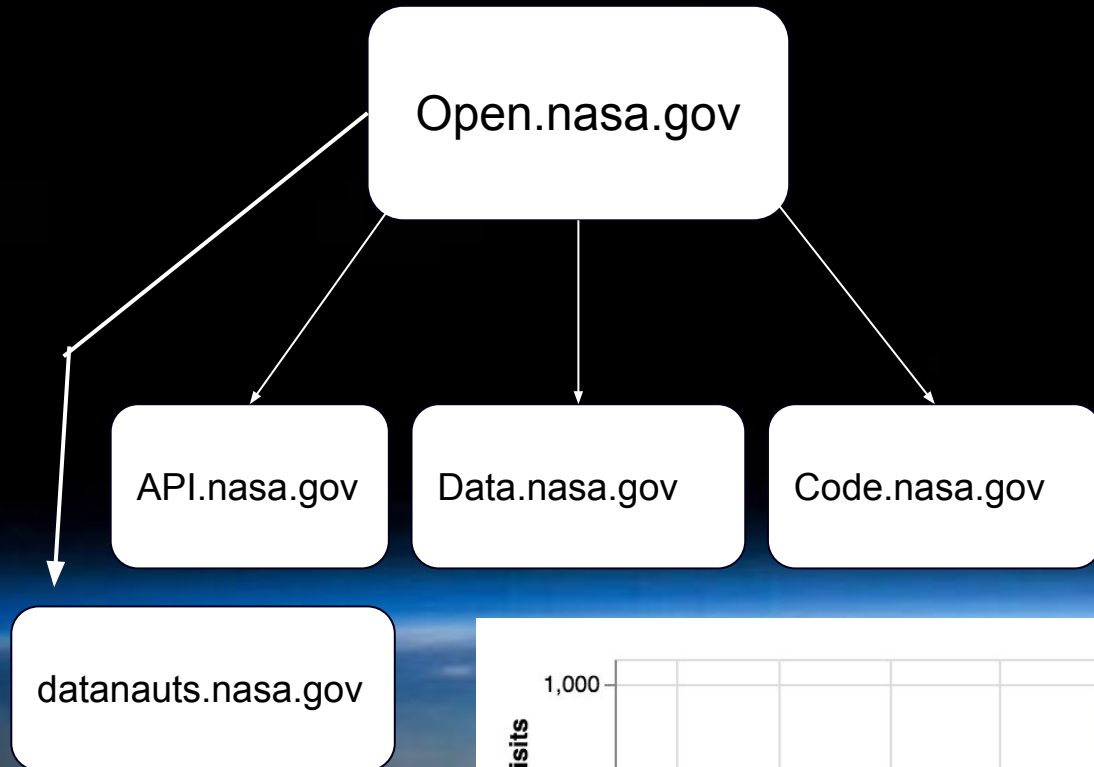


Curious

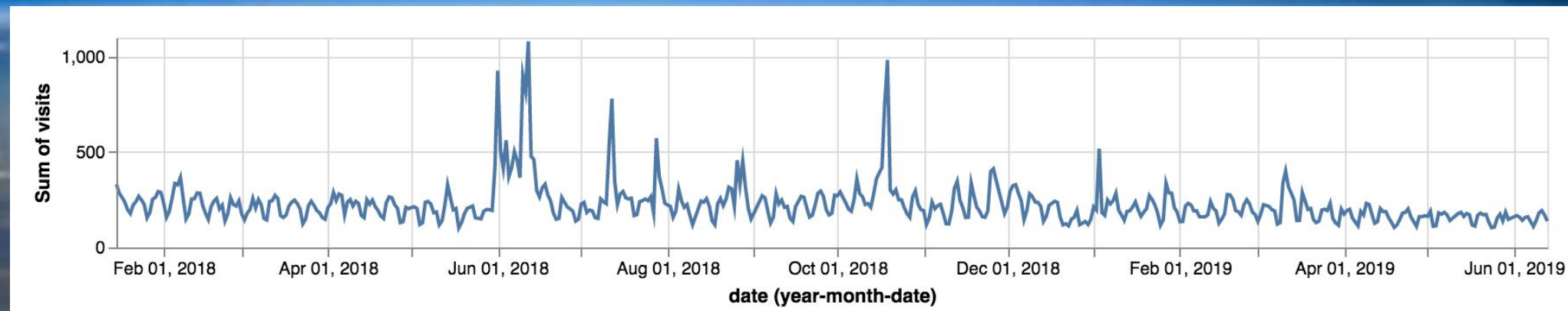
# OPEN.nasa.gov

The front-door for all other open-innovation sites.

Blog + About + News + Links



[Visits to open.nasa.gov](#) from analytics.usa.gov's API.



# API.nasa.gov (a passthrough service with tracking by api.data.gov)

*A.P.I. = Application Programming Interface (write code to get back data)*

There are many other APIs available not listed on this site! This page serves as a central easy to find location for NASA's easier to use A.P.I.

## STATISTICS:

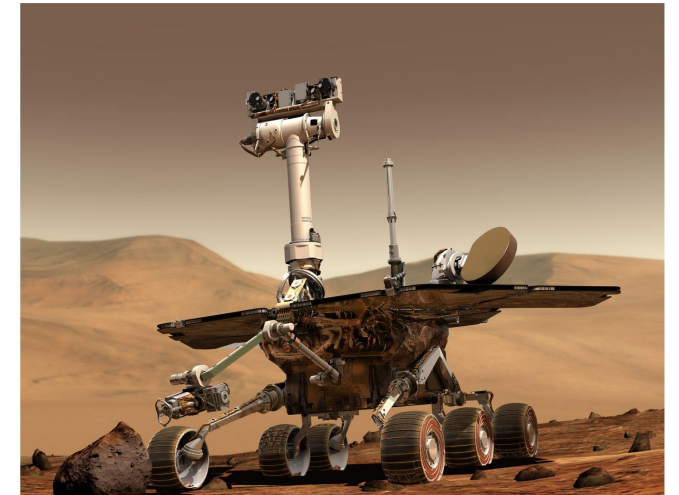
- **17 APIs:** patents, exoplanets, satellite imagery, hand camera imagery, Mars, etc.
- **55,000 API key** owners since 2015.
- **9 Million hits** in May, 2019

Learn how to build an Astronomy Picture of the Day App with the NASA API and React + Redux (Part...



<https://www.freecodecamp.org/news/learn-how-to-build-astronomy-picture-of-the-day-app-with-nasa-api-and-react-redux-part-ii-83f15970d0e3/>

Texting robots on Mars using Python, Flask, NASA APIs and Twilio MMS



<https://www.twilio.com/blog/2017/04/texting-robots-on-mars-using-python-flask-nasa-apis-and-twilio-mms.html>



# CODE.nasa.gov

**555 open-source projects**

Fed from software that has gone through Software Release System run by Office of Chief Engineer

Most but not all code is also on [github.com/nasa](https://github.com/nasa)

**Table shows the open-source projects with the most interaction on GitHub using GSA's pre-built scripts**

<https://observablehq.com/d/d2f8833d92e1ea66>

name	owner	issues	forks	stargazers	watchers	forkCount
"openmct"	"nasa"	"1378"	"663"	"7612"	"273"	"704"
"NASA-3D-Resources"	"nasa"	"15"	"218"	"1582"	"220"	"234"
"earthdata-search"	"nasa"	"9"	"136"	"463"	"58"	"147"
"sunpy"	"sunpy"	"968"	"314"	"423"	"45"	"328"
"WebWorldWind"	"NASAWorldWind"	"506"	"166"	"397"	"58"	"185"
"trick"	"nasa"	"637"	"110"	"387"	"64"	"122"
"worldview"	"nasa-gibs"	"1075"	"103"	"377"	"39"	"109"
"visionworkbench"	"visionworkbench"	"40"	"139"	"342"	"38"	"221"
"WorldWindJava"	"NASAWorldWind"	"109"	"180"	"324"	"82"	"192"
"api-docs"	"nasa"	"89"	"52"	"223"	"45"	"61"
"XPlaneConnect"	"nasa"	"95"	"111"	"215"	"47"	"123"
"StereoPipeline"	"NeoGeographyToolkit"	"224"	"68"	"191"	"41"	"107"
"code-nasa-gov"	"nasa"	"41"	"43"	"189"	"22"	"49"
"Common-Metadata-Repository"	"nasa"	"6"	"46"	"179"	"25"	"50"

# DATA.nasa.gov

- ▶ Started in July 2014
- ▶ 20K unique users
- ▶ 40K active datasets
- ▶ 23K page views this month (so far)

The largest number of datasets. Harvests data from other sites. Get's harvested into data.gov.

The screenshot displays the NASA's Open Data Portal interface. At the top, there's a navigation bar with 'NASA's Open Data Portal', 'Data Catalog', and a 'Create' dropdown. A search bar is prominently featured. Below the search bar, a sidebar on the left offers navigation through 'Categories' (Aerospace, Applied Science, Apps, Earth Science, Management/Operatio..., Show All...) and 'View Types' (Calendars, Charts, Data Lens pages, Datasets, External Datasets, Files and Documents, Filtered Views). The main content area shows '24414 Results' and a 'Sort by' dropdown set to 'Recently Added'. A blue banner at the top of the results area encourages users to 'Feature your best content'. Three data entries are listed, each with a title, a brief description, a 'More' link, tags, and creation/view statistics. Each entry also includes an 'External Link' icon.

Category	View Type	Title	Description	Tags	Created	Views	External Link
Earth Science	Datasets	TES/Aura L2 Water Vapor Nadir Special Observation V008	TES Level 2 data contain retrieved species (or temperature) profiles at the observation targets and the estimated errors. The geolocation, quality and other data (e.g., surface characteristics for nadir observations) are also...	earth science, atmosphere, atmospheric water vapor, atmospheric temperature, ngda, and 1 more	June 17, 2019	1	External Link
		TES/Aura L2 Summary Profiles Special Observation V008	TES Level 2 data contain retrieved species (or temperature) profiles at the observation targets and the estimated errors. The geolocation, quality and other data (e.g., surface characteristics for nadir observations) are also...	earth science, atmosphere, atmospheric chemistry, air quality, ngda, and 1 more	June 17, 2019	0	External Link
		TES/Aura L2 Peroxyacyl Nitrate Nadir V008	TES Level 2 data contain retrieved species (or temperature) profiles at the observation targets and the estimated errors. The geolocation, quality and other data (e.g., surface characteristics for nadir observations) are also...	earth science, atmosphere, air quality, atmospheric chemistry, ngda, and 1 more	June 17, 2019	0	External Link



# Who Puts All This Together?

1000s of NASA & contractor staff who contribute code projects, APIs, and datasets

2.5 developers who maintain the open-innovation sites

You! a lot of our code for these sites is on public github.com repositories and we accept pull requests.

- <https://github.com/nasa/api-docs>
- <https://github.com/nasa/code-nasa-gov>



# The NASA Data Universe

Why so many places to find data?



# Mandates & Requirements Drive Dataset Storage Diversity

- **Congress mandates:**
  - Mandate agency-wide open-innovation sites (code,api,nasa) that harvest data other sites.
- **Scientific grants:**
  - Specify certain scientific domain archives.
- **Data consumed by other IT systems with requirements:**
  - Websites, tools, etc. sometimes required being stored in specific systems (databases, APIs).

## Examples:

- Congressional mandate: [Federal Data Strategy](#)
- Science data mandate for specific domain: [Earth Science](#)
- Data consumed by other IT systems: [3D moon rocks with volumetric scans without an open-source rendering engine.](#)

# Harvesting Relationships

US Government-wide

**data.gov**

NASA-wide

**data.nasa.gov**

Small unique datasets

**csv file on plants**

**api.nasa.gov**

Project Specific

**APOD website**

**earthdata.nasa.gov**

Domain Specific Archives





# Domain-specific NASA Data Sites:

1. [Planetary Data System](#) (data from other planets)
2. [Earth Observations](#) (mostly satellite-based Earth data)
3. [Astromaterials](#) (samples of: dust, comets, meteorites, moon-rocks + 3d models + impacts)
4. [Solar Power datasets & APIs](#)
5. [Global Near Earth Observations](#) (global maps of solar flux, rain, etc.)
6. [Soil Moisture](#)
7. [NODIS](#) (NASA Online Directives Information Service)
8. [Scientific and Technical Information](#) or *STI* (papers, reports, etc.)
9. [NASA Technology Transfer Office](#) (software, patents, tools, small business)
10. [USGS](#) & [NOAA](#) also have Earth Imagery
11. [USGS](#) & [NOAA](#) also have Climate data!

# Suggestions for Finding NASA datasets



# Most used datasets are easiest to work with!

## ***Most reused Code***

1. [openmct](#)
2. [earthdata-search](#)

## ***Clones***

281 clones last 2 weeks  
212 clone last 2 weeks

## ***Datasets***

1. [Meteorite Landings:](#)
2. [NASA Facilities:](#)
3. [Global Landslide Catalog Export:](#)
4. [Surface Meteorology and Solar Energy:](#)

## ***Total Downloads***

49,000  
34,000  
29,000  
22,000

## ***API***

1. APOD (astronomy photo of the day)
2. neo (near earth objects)
3. mars-photos/

## ***Downloads in May***

9,520,397  
508,207  
195,953

# Find Starter Code!

saves time on dataset finding & prepping

## Github

[Searching for the terms on github](#) will often provide some open-source licensed code you can reuse! 'NASA' returns 10,000 results!

Datanauts is a program where members of the public work with NASA open-data. [The datanauts github org](#) is a great place to find starter code.

[Search through 5000 past SpaceApps projects](#) using this app by datanaut Alexandre Belloni Alves

## Live JavaScript Code Collections

Observable Notebooks are like Jupyter notebooks but JavaScript, live, editable, & forkable on the web! Search for terms or [check out this NASA collection](#).

Bl.ocks is a site that collects live d3 visualizations. [You can put 'nasa' into the search and get back things that use NASA data!](#)

# Consider whether you're finding or discovering datasets?


*Use [data.nasa.gov](https://data.nasa.gov) or other sites where you can search via titles, names, and other things that work well with string matching.*



**Dataset Finding** = “You know the dataset exists and what it is called”

**Dataset Discovery** = “You don't know the name, whether it exists, what it looks like until you see it, or how you'll use it until you see it.”

*Look at previous code projects or websites that only hold specific types of data. These are more likely to have visual representations of data that help you determine what exists and how you might use it.*





# Consider Discoverability vs. Data Site Type

Sites Type	Example	Meta-data	Interfaces Built-for	Discoverability Type
Harvest generated site	<a href="https://data.nasa.gov">data.nasa.gov</a>	generic with links to more metadata	General public & search engines	String matching in descriptions or titles.
Domain specific site	<a href="https://pds.nasa.gov">pds.nasa.gov</a>	science field specific metadata	Scientists, engineers & developers who need authoritative files	Filter content types & location & format filtering
Dataset specific site	Insight Mars Weather API	dataset specific metadata	dataset users	See example use-cases

# Open-Innovation Site Specific:

## API.NASA.GOV

1. [API documentation may not be on the first page you find about a dataset.](#)
  - 1.1. Be sure to check out all links. API documentation often has some example code but not always easy to find.
2. This site holds easy to use APIs. It does not hold all APIs!
3. Data.nasa.gov datasets sometimes have their own APIs.

## CODE.NASA.GOV & Github.com/nasa

1. Not everything on code.nasa.gov is on github. Some projects on [gitlab.com](#) or other.
2. Some projects under project specific org accounts, not 'nasa'.
3. All data used to build [code.nasa.gov](#) is in the code.json and can be accessed via [API at code.gov](#).

## DATA.NASA.GOV

1. Data.nasa.gov does not have a good ability to filter by data format as most people do not upload datasets but merely link to other data sites, so do not think you have all the CSVs if you search for things in CSV format!
2. The 'tags' and 'types' are human-populated and should not be thought of as 'complete'.

# Example

## Of Finding Data



# Example 1: Lunar Sample Evaluation

## THE CHALLENGE

You are the astronaut/robotic mission lead tasked with bringing valuable specimens from the Moon back to Earth for further study. How will you evaluate lunar samples quickly and effectively before or while still on the mission? How will you differentiate samples of potential scientific value from less interesting material?

## Suggestion 1: Look for org pages of NASA groups that do related work:

- Johnson Space Center Astromaterials Group: They have a [database](#) of all the Apollo samples.

## Suggestion 2: Search for papers/descriptions of past NASA work in [data.nasa.gov](#) & [sti.nasa.gov](#):

- STI has a paper on this challenge: [GeoLab Concept: The Importance of Sample Selection During Long Duration Human Exploration Mission](#)
- Data.nasa.gov has a report about a study similar to this task: [SAND-E](#): Semi-Autonomous Navigation for Detrital Environments - Integrating robotic terrain analysis, geochemistry, and sedimentology





**Any Questions?**

**Best of Luck with SpaceApps!**

Presentation video: <https://youtu.be/NA3J8NPu-U4>