Open NASA Data: From API to Data Analysis

by Noemi Derzsy

NASA Datanauts – April 24, 2017

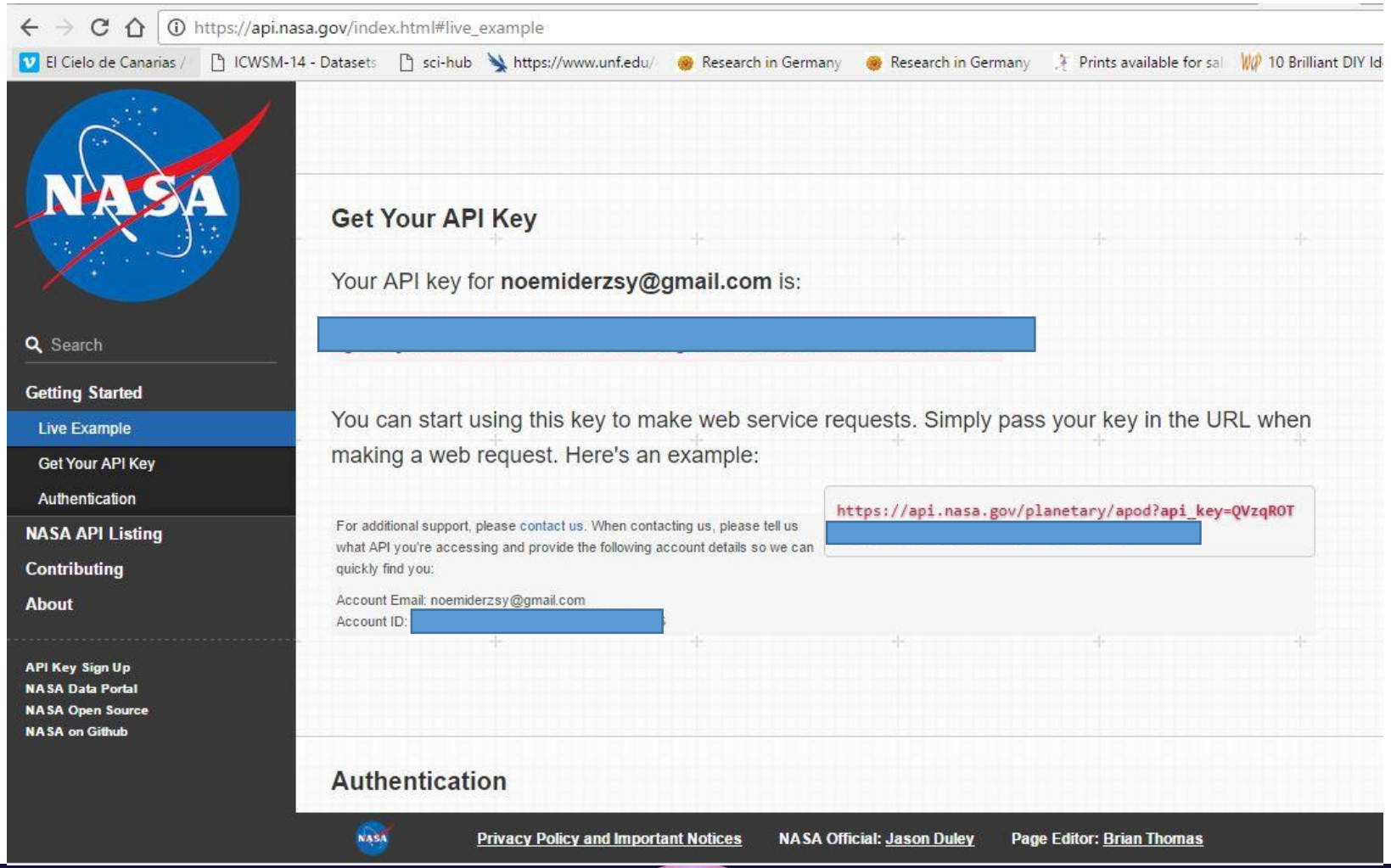
Open NASA Platform

Data: 32,089
Code repositories: 328
APIs: 51

Continuously growing...

Open NASA API

Obtain API Key



Simply pass your key in the URL when making a web request:

https://api.nasa.gov/planetary/apod?api_key=QVzqRO.....


```
{
  "copyright": "Adam Block",
  "date": "2017-03-11",
  "explanation": "Riding high in the constellation of Auriga, beautiful, blue vdB 31 is the 31st object in Sidney van den Bergh's 1966 catalog of reflection nebulae. It shares this well-composed celestial still life with dark, obscuring clouds recorded in Edward E. Barnard's 1919 catalog of dark markings in the sky. All are interstellar dust clouds, blocking the light from background stars in the case of Barnard's dark nebulae. For vdB 31, the dust preferentially reflects the bluish starlight from embedded, hot, variable star AB Aurigae. Exploring the environs of AB Aurigae with the Hubble Space Telescope has revealed the several million year young star is itself surrounded by flattened dusty disk with evidence for the ongoing formation of a planetary system. AB Aurigae is about 470 light-years away. At that distance this cosmic canvas would span about four light-years.",
  "hdurl": "http://apod.nasa.gov/apod/image/1703/vdb31AdamBlock.jpg",
  "media_type": "image",
 "service_version": "v1",
  "title": "Reflections on vdB 31",
  "url": "http://apod.nasa.gov/apod/image/1703/vdb31AdamBlock_s1024.jpg"
}
```

Access OpenNASA Data using Python

pyNASA

https://github.com/bmtgoncalves/pyNASA

- Simple interface to select NASA datasets:
  - pyNASA.meteorite() - Meteorite Landings
  - pyNASA.comets() - Near-Earth Comets - Orbital Elements
  - pyNASA.landslides() - Global Landslide Catalog Export
  - pyNASA.facilities() - NASA Facilities
  - pyNASA.bolide() - Fireball And Bolide Reports
  - pyNASA.comet_discovery() - WISE NEA/COMET DISCOVERY STATISTICS
  - pyNASA.outgassing() - Outgassing Db
  - pyNASA.open_source() - Open Source And General Resource Software
  - pyNASA.patents() - NASA Patents
  - pyNASA.eva() - Extra-vehicular Activity (EVA) - US and Russia
  - pyNASA.candida_albicans() - Candida albicans response to spaceflight (NASA STS-115)
  - pyNASA.sxsw_2016() - SxSW 2016 Leads
  - pyNASA.gene_chip_assay() - A E- GEOD-50881 Gene Chip Assay
  - pyNASA.bexrb_monitor() - BeXRB Monitor Data
  - pyNASA.geod_50881() - S E- GEOD-50881 Study Samples

@bgoncalves

pyNASA

https://github.com/bmtgoncalves/pyNASA

- Simple interface to select NASA datasets
- Returns dataset directly as a pandas DataFrame
- Easy to extend

```python
from pyNASA import pyNASA

if __name__ == "__main__":
    from NASA_accounts import apps
    app = apps["pyNASA"]

    nasa = pyNASA(**app)
    data = nasa.outgassing()

    print(data.shape)
```

@bgoncalves

demo.py

account credentials

- Store the as a dictionary in an external file:

```
apps = {
    "pyNASA": {
        "token": "YOUR_TOKEN",
        "secret": "YOUR_SECRET"
    }
}
```

- This way it easy and convenient to manage and use multiple accounts

```
from NASA_accounts import apps
app = apps["pyNASA"]
```

@bgoncalves

NASA_accounts_stub.py

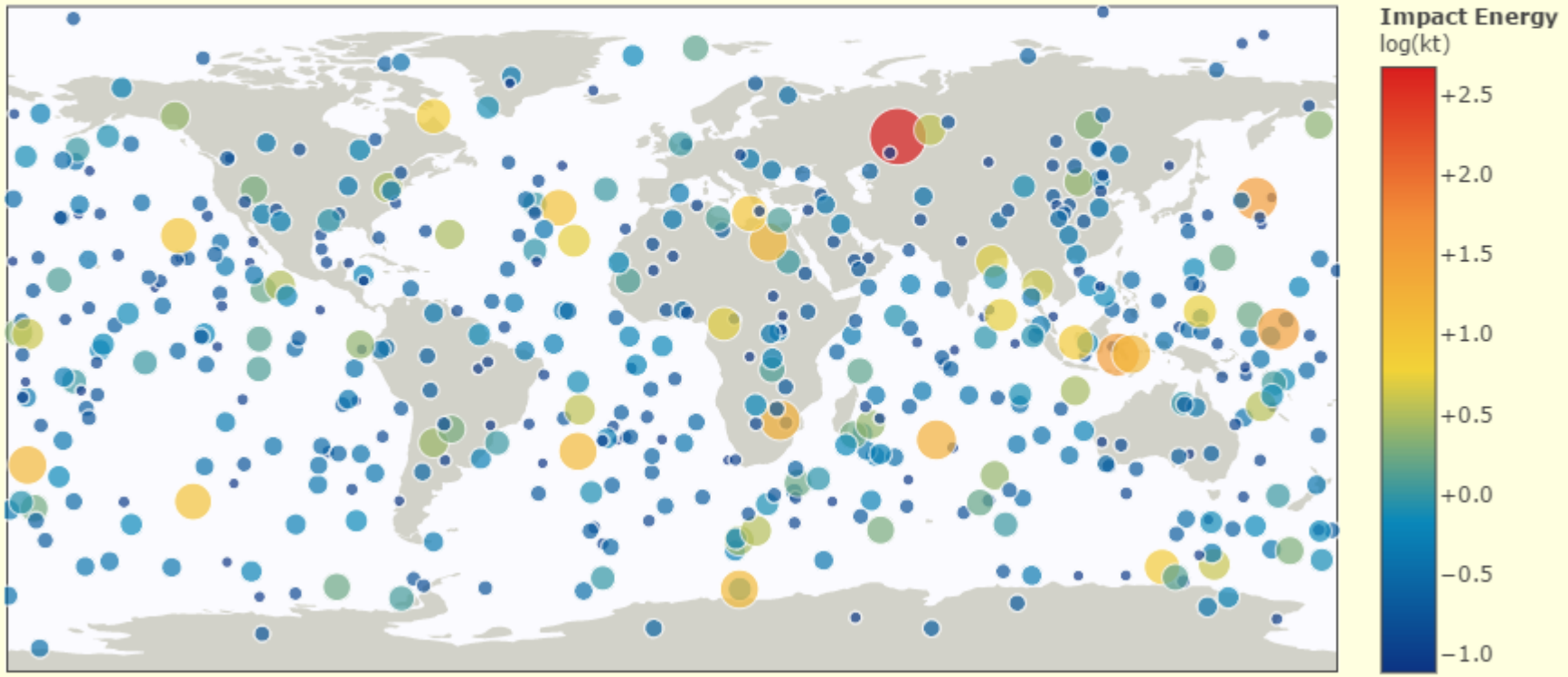
What data can you find and what can you do with them?

Some ideas....

Meteorite Landings

Fireballs Reported by US Government Sensors

(1988-Apr-15 to 2017-Mar-11)

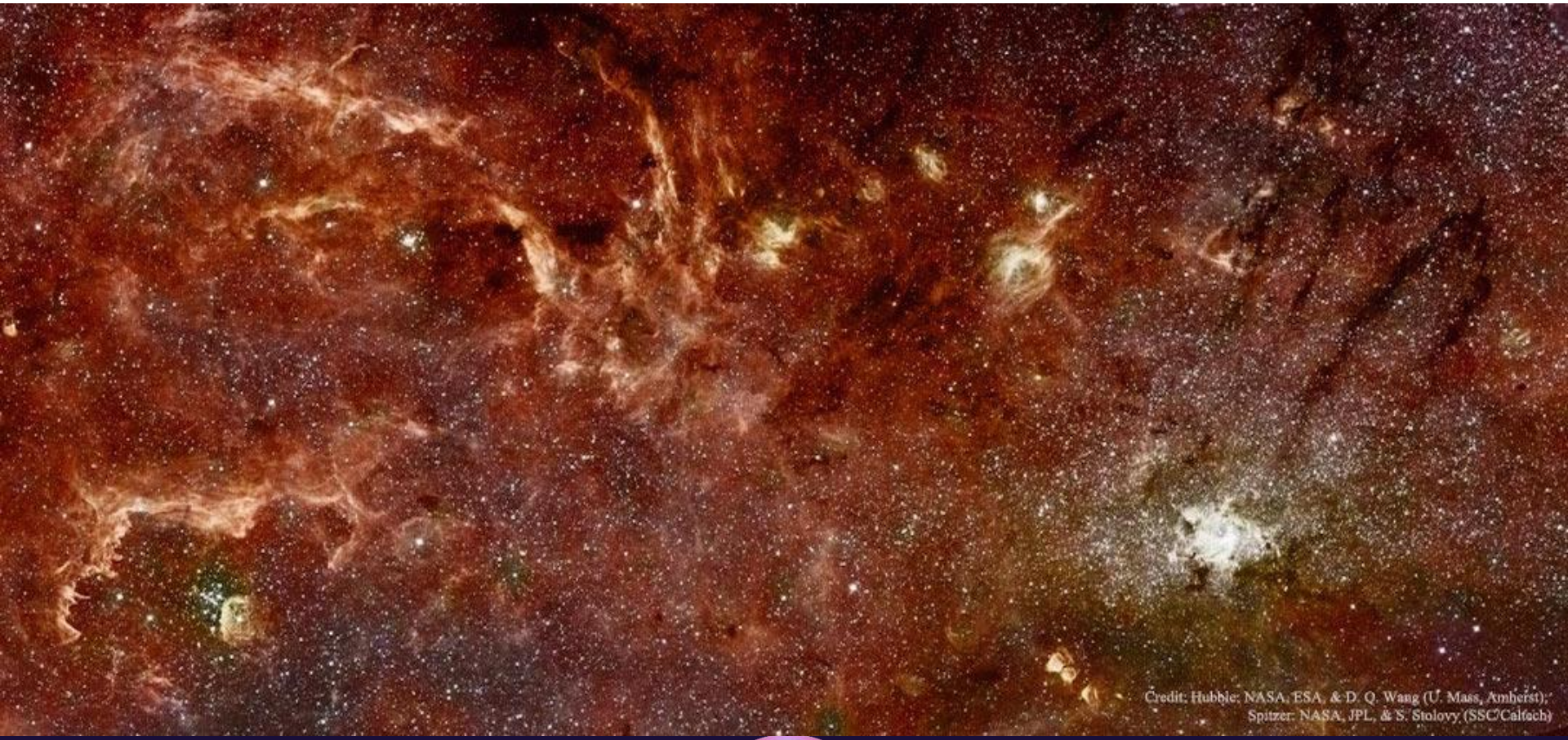


https://cneos.jpl.nasa.gov/fireballs/

Alan B. Chamberlin (JPL/Caltech)

Deep Learning using APOD

APOD - Astronomy Picture of the Day



Noemi Derzsy

DATANAUTS

NASA Datanauts – April 24, 2017

Deep Learning using EPIC

EPIC: information on the daily imagery collected by DSCOVR's Earth Polychromatic Imaging Camera (EPIC) instrument. Uniquely positioned at the Earth-Sun Lagrange point, EPIC provides full disc imagery of the Earth and captures unique perspectives of certain astronomical events such as lunar transits

What else?

- Data visualization: ex. geolocated data (meteorite landings)

- Data analysis: ex. NASA patent portfolio

- Deep learning: ex. Image analysis from data collected by MARS ROVER, or space sound analysis "When we retrieve sounds from far-off planets, we can apply the filter to identify unknown human space colonies. That was a joke. Sort of."

- Hardware projects
- Art projects
- Educational projects

Or anything else that you can think of...

Open NASA Data Collection Analysis

Noemi Derzsy

NASA Metadata

Collection of NASA thesauri, dictionaries, taxonomies and related documents:
https://github.com/nasa/dictionaries

Metadata information:

- Title
- Description
- Organization within NASA
- Keywords
- License
- Location (HTML link)
- etc.

Format: JSON

Open NASA Data - Topic Analysis

Open Government Data (32,000 datasets): https://data.nasa.gov/data.json
JSON (JavaScript Object Notation)

Parsing JSON

```
# Topic Analysis
from urllib import urlopen
import json

u = urlopen('https://data.nasa.gov/data.json')
d = json.loads(u.read().decode('utf-8'))

from pprint import pprint
pprint(d)
```

working with files (load) instead of strings (loads)

Explore… connections between the datasets

list of datasets with description → a list of Python dictionaries

Consolidated data collection

Open NASA Datasets Metadata

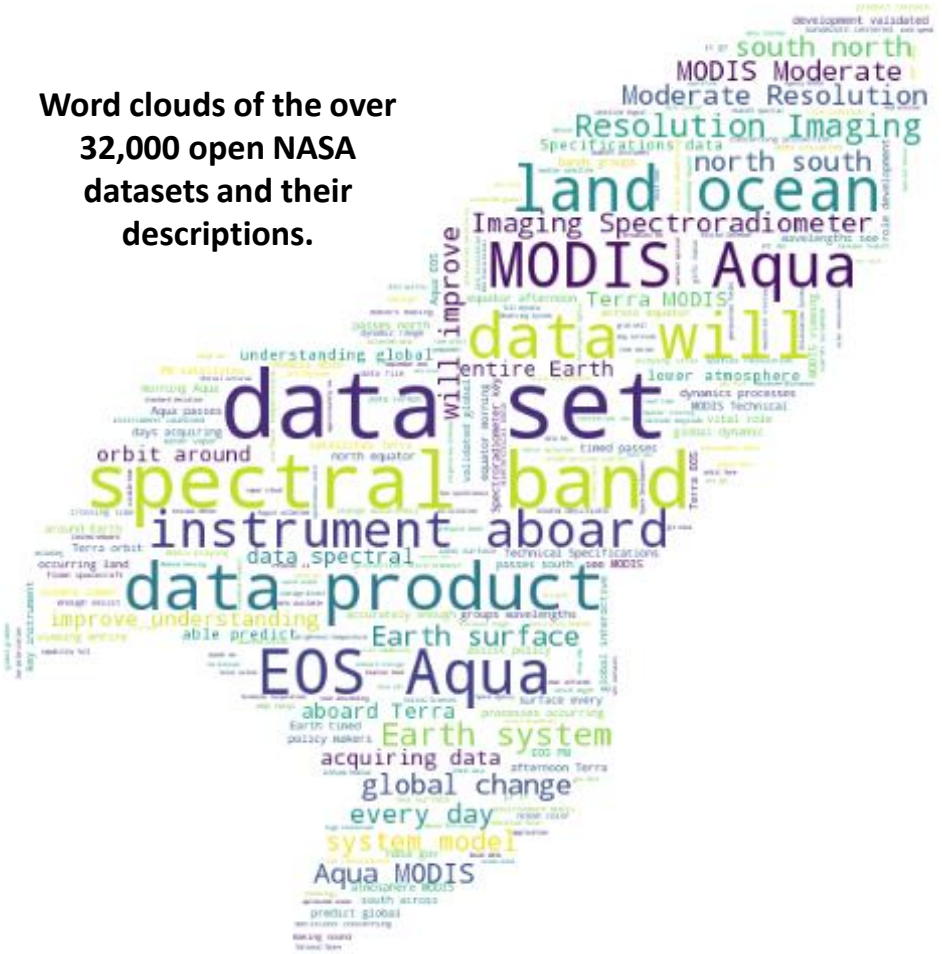

```
" id": {
"@type": "dcat:Dataset",
"accessLevel": "public",
"accrualPeriodicity": "irregular",
"bureauCode": [
"contactPoint": {
"description": "USGS 15 minute stream flow data for Kings Creek on the Konza Prairie",
"distribution": [
"identifier": "C179003030-ORNL_DAAC",
"issued": "2008-12-02T00:00:00.000Z",
"keyword": [
"landingPage": "http://reverb.echo.nasa.gov/reverb?selected=C179003030-ORNL_DAAC",
"language": [
"modified": "2008-12-02T00:00:00.000Z",
"programCode": [
"publisher": {
"spatial": "39.1 -96.6",
"temporal": "1984-12-25T00:00:00Z/1988-03-04T00:00:00Z",
"theme": [
"title": "15 Minute Stream Flow Data: USGS (FIFE)",
"license": "http://www.usa.gov/publicdomain/label/1.0/"
```

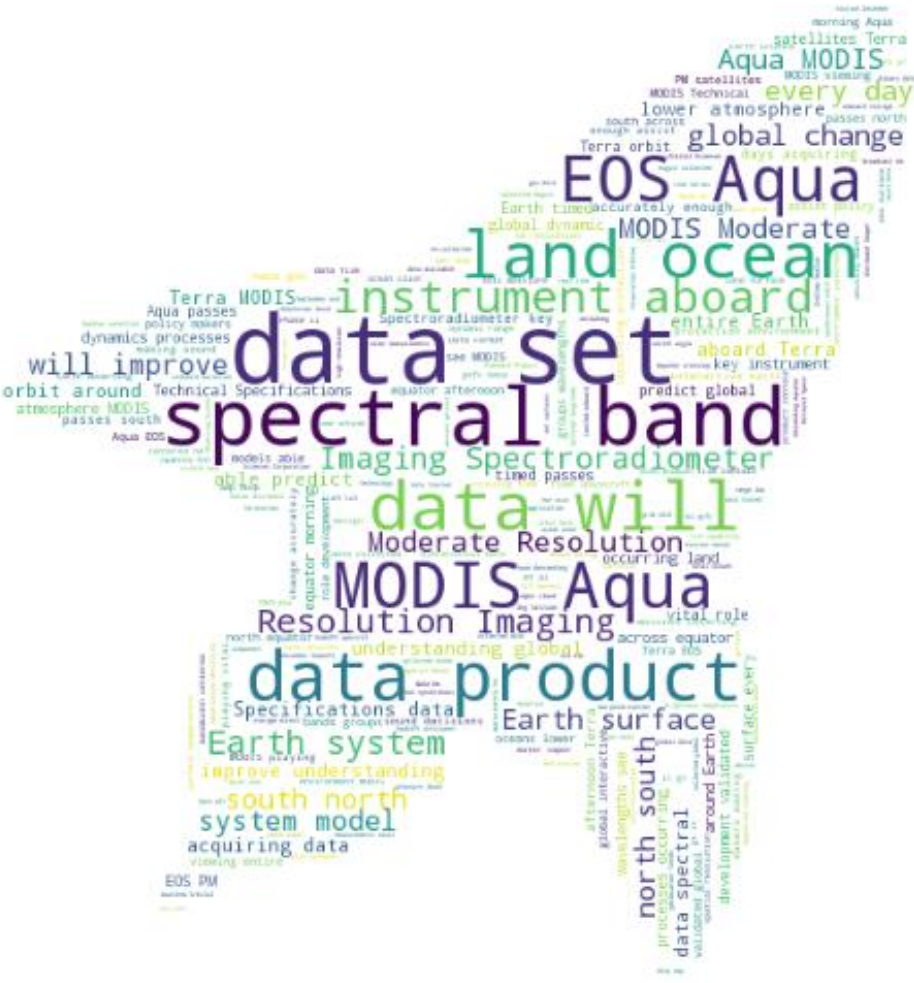
Word Cloud with Python: Title

Word clouds of the over 32,000 open NASA datasets and their titles.

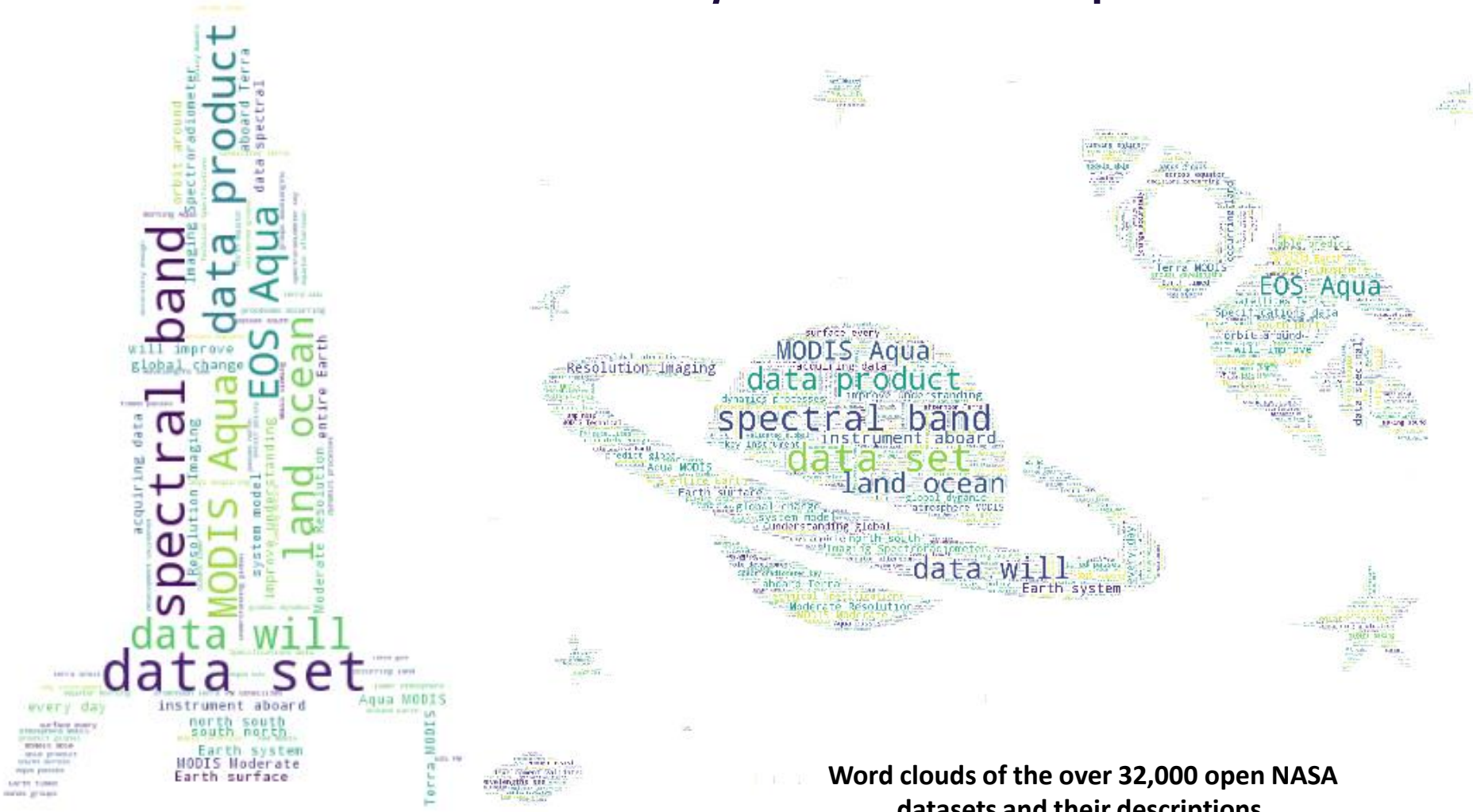
Word Cloud with Python: Descriptions

Word clouds of the over 32,000 open NASA datasets and their descriptions.



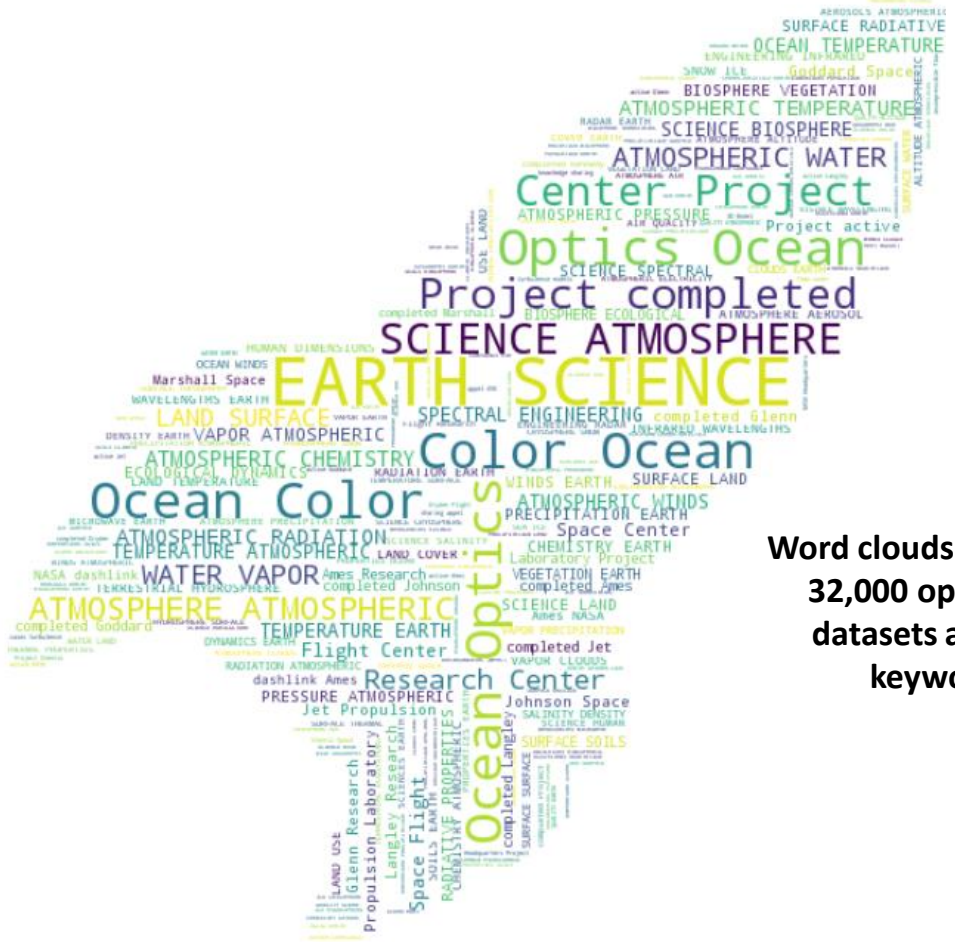


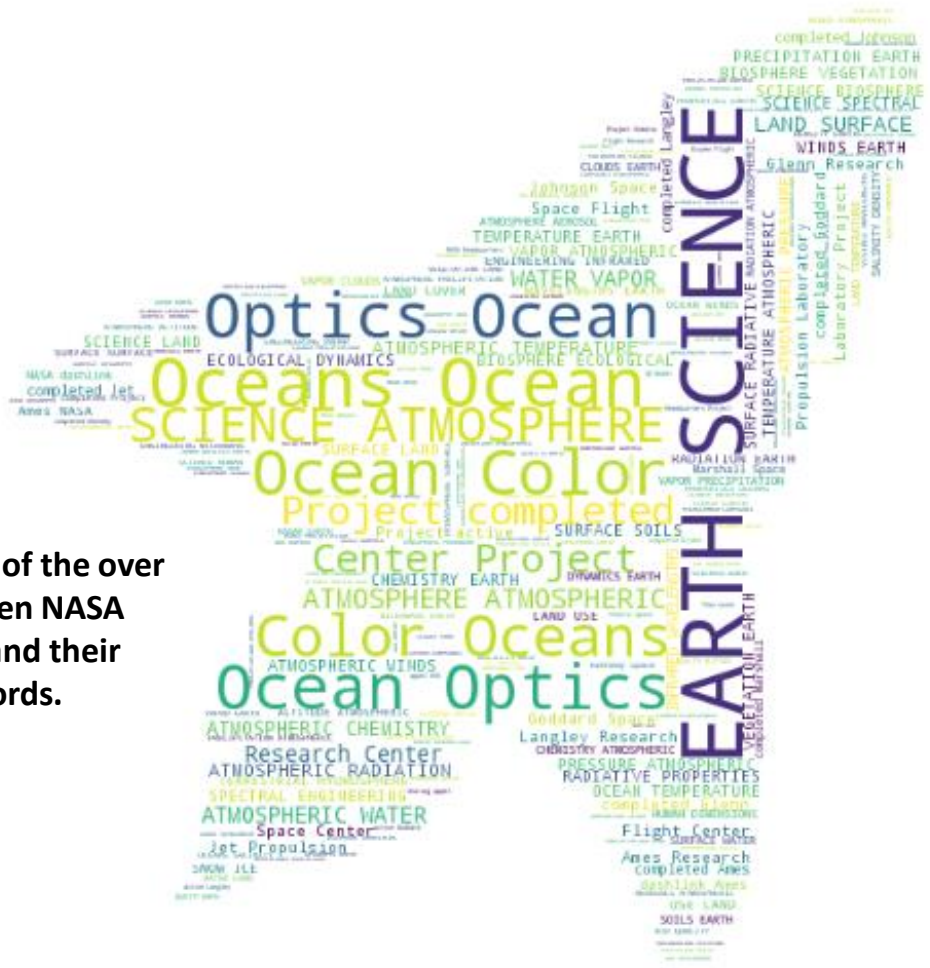
Word Cloud with Python: Descriptions



Word clouds of the over 32,000 open NASA datasets and their descriptions.

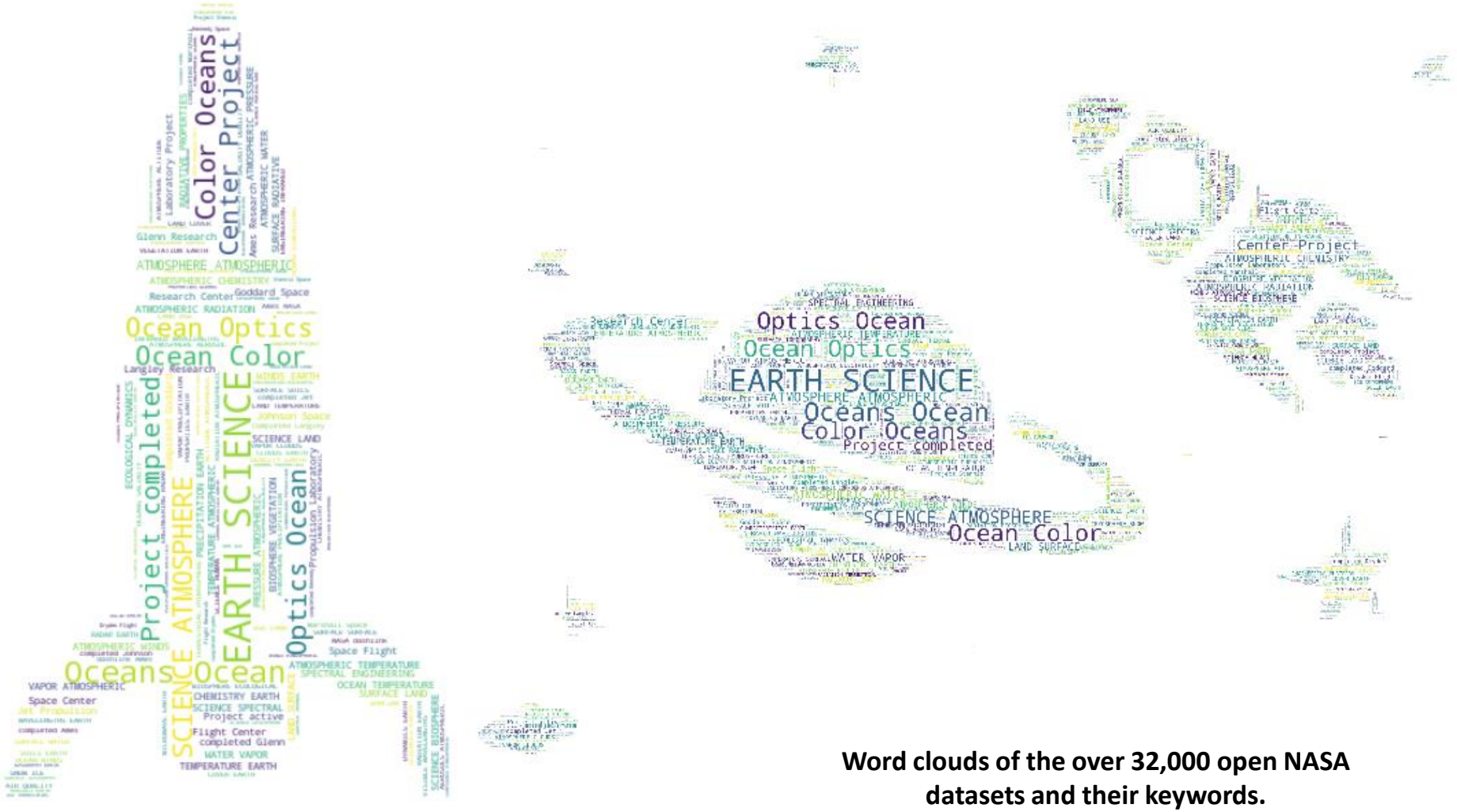
Word Cloud with Python: Keywords





Word clouds of the over 32,000 open NASA datasets and their keywords.

Word Cloud with Python: Keywords

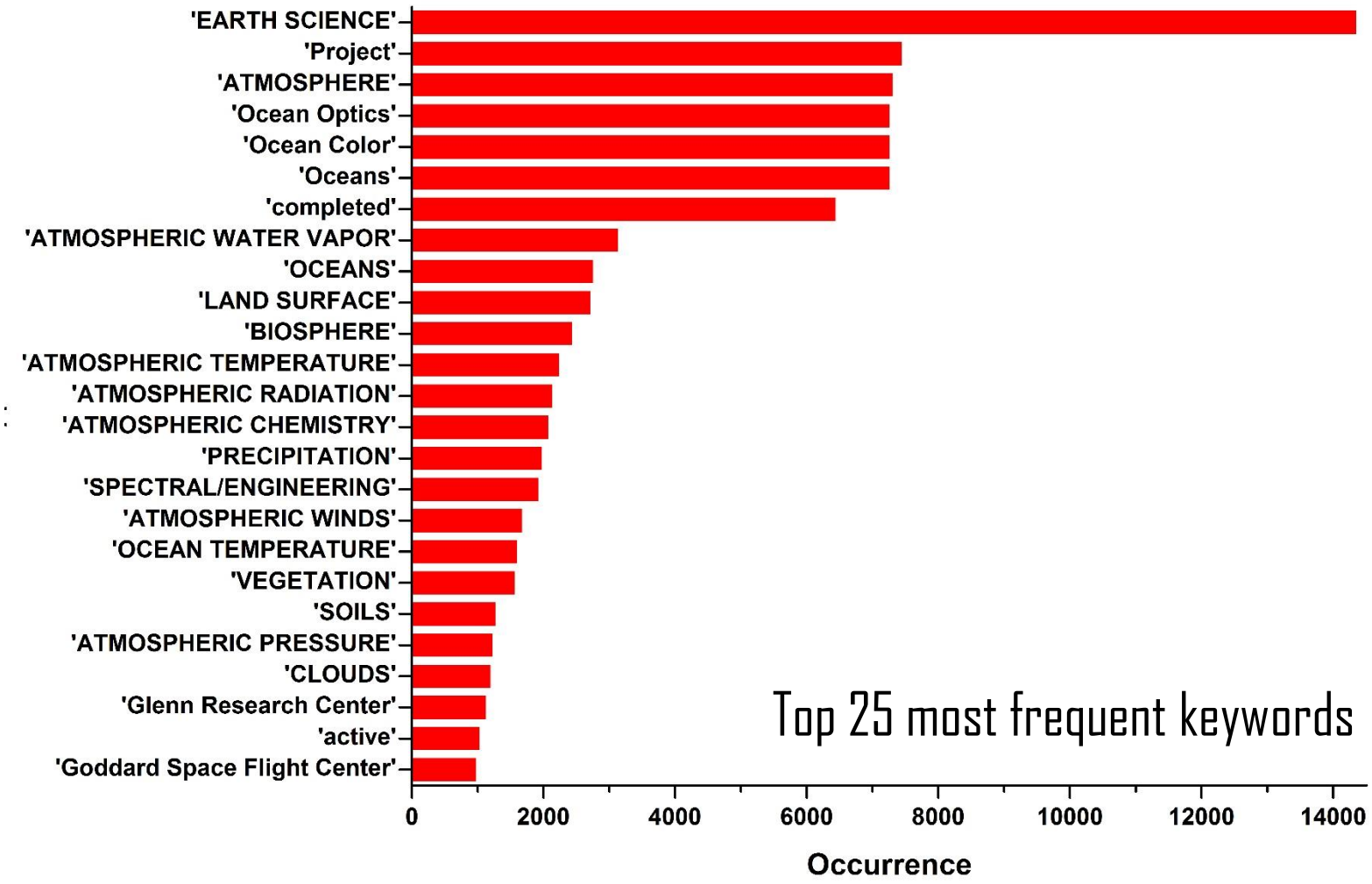


Word clouds of the over 32,000 open NASA datasets and their keywords.

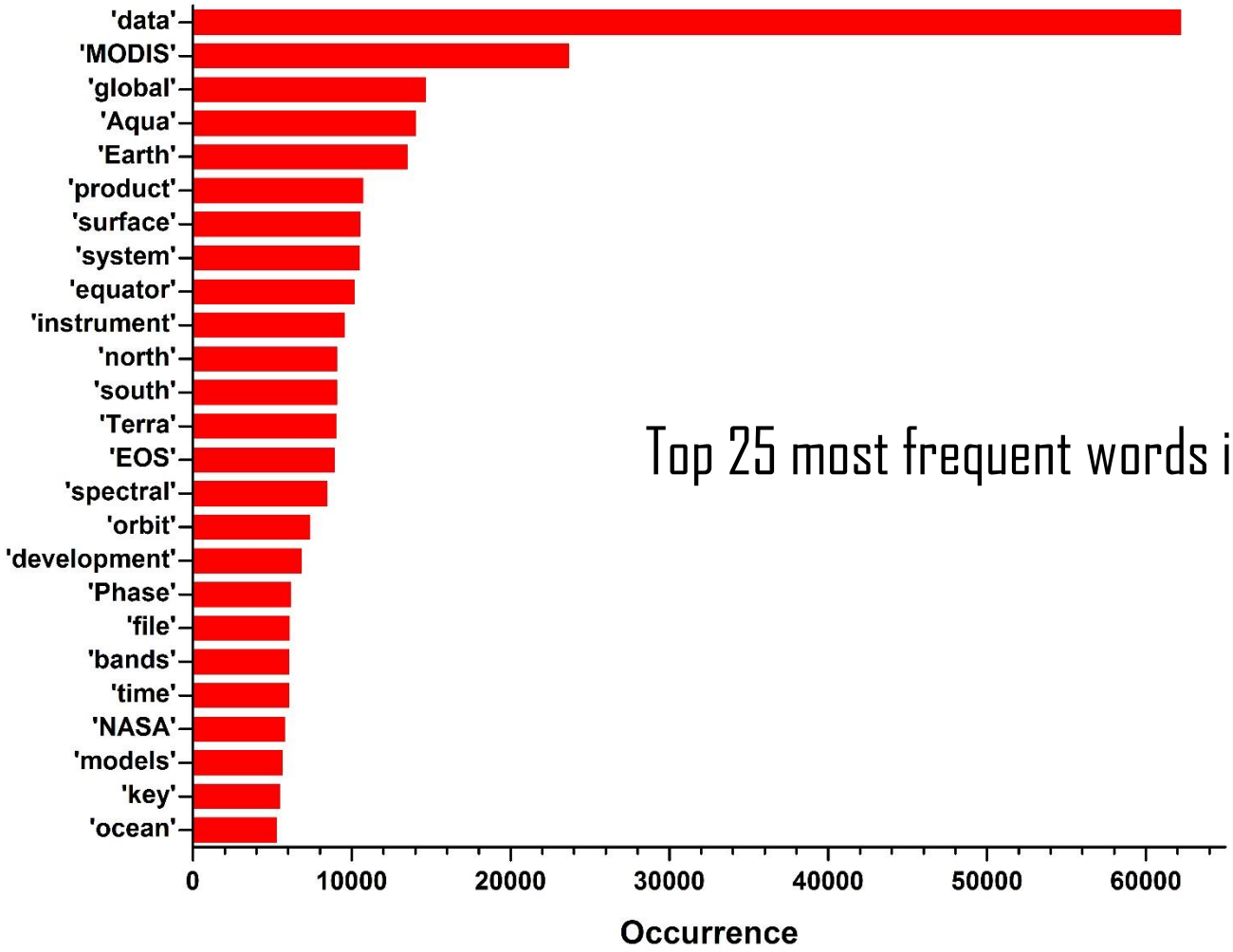
How to Generate Your Own Wordcloud in Python

1. Get text
2. Get stencil (in whatever shape you want)
3. Use code: https://github.com/amueller/word_cloud
4. Modify code for your data

Term Frequency - Keywords

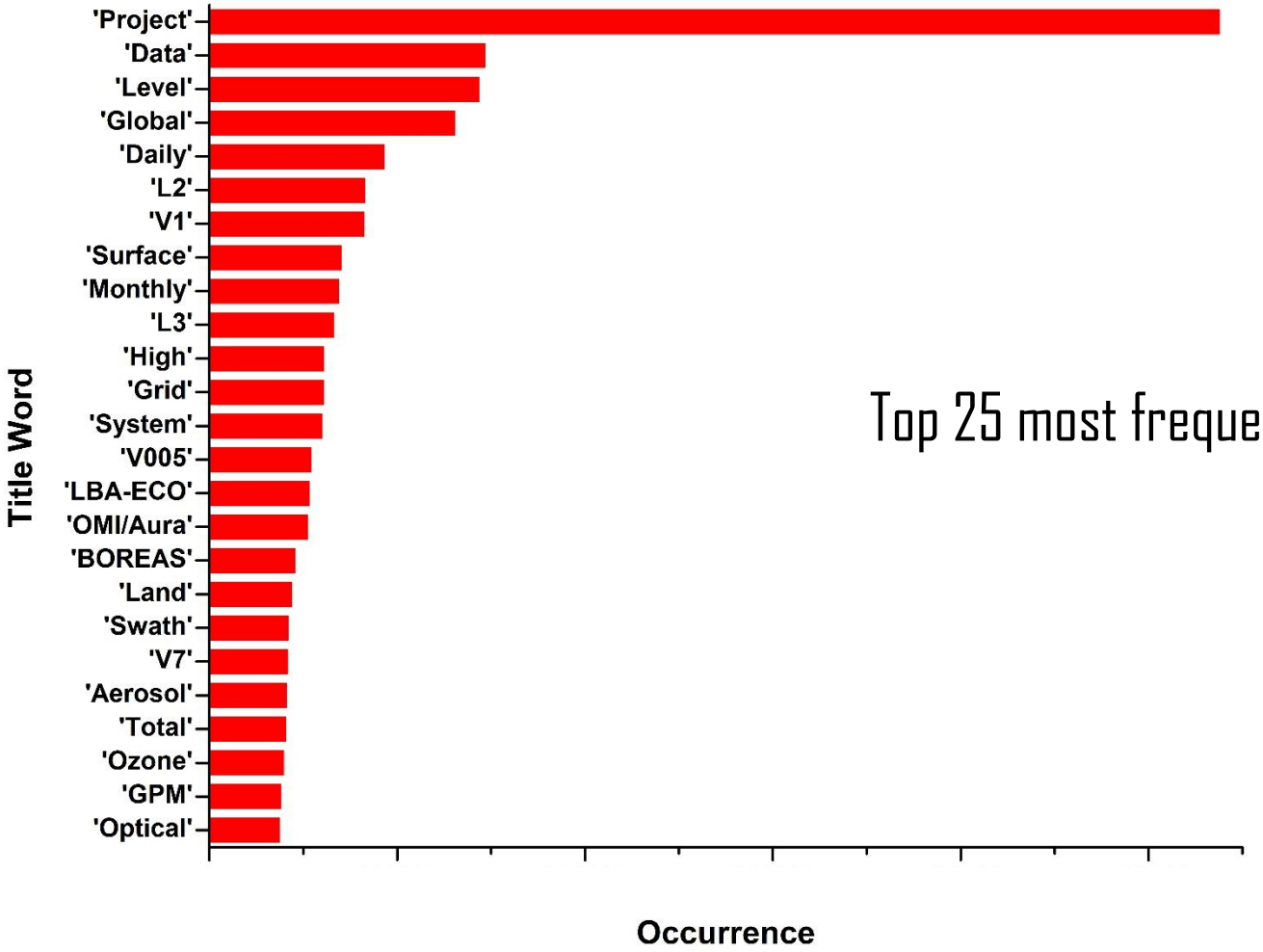


Term Frequency - Description



Top 25 most frequent words in Description

Term Frequency - Title



Top 25 most frequent words in Title

For more on NLP/Topic Analysis…

JSC Data Science Day 2017 - USTREAM

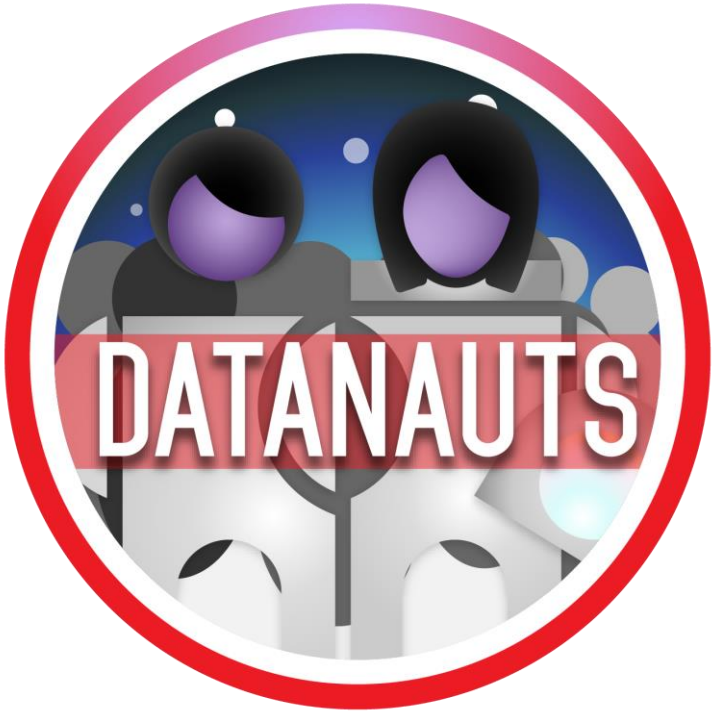
Event Timing: April 26 - 27
Event Venue: USTREAM WEB
A link will be emailed to you prior to the event.
Contact us at david.meza-1@nasa.gov or julie.a.barnes@nasa.gov
Follow the event on twitter: @JSCDataScience and #JSCDataScienceDay

Register here to receive the link for the USTREAM feed of the sessions held in the main ballroom of this event.
The link will be on the confirmation box that pops up after you submit the form.

The event is designed to share knowledge and provide training on how data analysis and visualization are used in our every day lives. Speakers from NASA, academia, and industry will be on hand to share examples on how data science impacts their field. This event is appropriate for anyone wanting to learn and share ideas, from the novices to experts, all are welcomed.Come help us make this a great knowledge sharing event.

Confirmed Speakers:
Jeff Williams - Astronaut
Lisa Vaughn - Meteorologist Fox 26 News and Data Scientist
Google
Amazon
RStudio
Microsoft
Harvard Business
Washington University in St. Louis
Topcoder

Website: https://fal.jsc.nasa.gov/DSD/index.htm



Thank you!