



**THE UNIVERSITY OF TEXAS AT EL PASO**

---

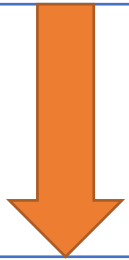
F.A.I.R. PRINCIPLES OF DATA SCIENCE  
Creating Open-Source Research for  
Collaboration and Publication

# What is Open Science?

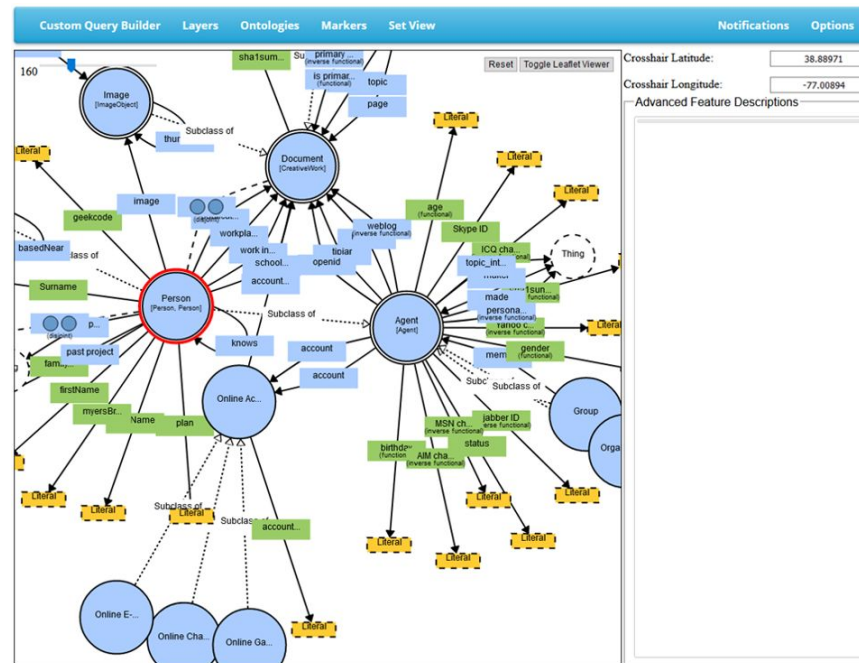
Open Science is the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods.

<https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>

DATA.XLSX



JOURNAL  
ARTICLE



(Credit: Tanner Fry. Public domain.)

<https://www.usgs.gov/media/images/varankaontologyviewsmall>

# F.A.I.R. PRINCIPLES

Findable

Accessible

Interoperable

Reuseable

<https://www.go-fair.org/fair-principles/>

# METADATA

The screenshot shows the AGU online library website. The article title is "Intensified dust storm activity and Valley fever infection in the southwestern United States" by Daniel Q. Tong, Julian X. L. Wang, Thomas E. Gill, Hang Lei, and Binyu Wang. The article is published in Geophysical Research Letters, Volume 44, Issue 9, on May 16, 2017. The abstract discusses the increasing frequency of dust storms in the southwestern United States and their correlation with Valley fever. The page also includes a citation statements section and a plain language summary.

```

<meta name="citation_journal_title" content="Geophysical Research Letters">
<meta name="citation_title" content="Intensified dust storm activity and Valley fever infection in the southwestern United States">
<meta name="citation_keywords" content="dust">
<meta name="citation_keywords" content="Valley fever">
<meta name="citation_keywords" content="climate">
<meta name="citation_keywords" content="air quality">
<meta name="citation_keywords" content="aerosol">
<meta name="citation_keywords" content="health">
<meta name="citation_volume" content="44">
<meta name="citation_issue" content="9">
<meta name="citation_firstpage" content="4304">
<meta name="citation_lastpage" content="4312">
<meta name="citation_publication_date" content="2017/05/16">
<meta name="citation_online_date" content="2017/05/06">
<meta name="citation_issn" content="1944-8007">
<meta name="citation_language" content="en">
<meta name="citation_author" content="Daniel Q. Tong">
<meta name="citation_author_institution" content="Center for Spatial Information Science and Systems George Mason University Fairfax Virginia USA">
<meta name="citation_author_institution" content="Air Resources Laboratory National Oceanic and Atmospheric Administration College Park Maryland USA">
<meta name="citation_author_institution" content="Cooperative Institute of Climate and Satellite University of Maryland College Park Maryland USA">
<meta name="citation_author" content="Julian X. L. Wang">
<meta name="citation_author_institution" content="Air Resources Laboratory National Oceanic and Atmospheric Administration College Park Maryland USA">
<meta name="citation_author" content="Thomas E. Gill">
<meta name="citation_author_institution" content="Department of Geological Sciences University of Texas at El Paso El Paso Texas USA">
<meta name="citation_author" content="Hang Lei">
<meta name="citation_author_institution" content="Center for Spatial Information Science and Systems George Mason University Fairfax Virginia USA">
<meta name="citation_author_institution" content="Air Resources Laboratory National Oceanic and Atmospheric Administration College Park Maryland USA">
<meta name="citation_author" content="Binyu Wang">
<meta name="citation_author_institution" content="Center for Spatial Information Science and Systems George Mason University Fairfax Virginia USA">
<meta name="citation_abstract_html_url" content="https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL073524">
<meta name="citation_fulltext_html_url" content="https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2017GL073524">
<meta name="citation_pdf_url" content="https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017GL073524">
<meta name="citation_publisher" content="John Wiley & Sons, Ltd"><meta name="citation_fulltext_world_readable" content=""/>
<meta name="article_references" content="Tong, D. Q., Wang, J. X. L., Gill, T. E., Lei, H., and Wang, B. (2017), Intensified dust storm activity and Valley fever infection in the southwestern United States, Geophys. Res. Lett., 44, 4304&#x2013;4312, doi:10.1002/2017GL073524.&#x26;">
<meta name="epdf_available" content="false">
<meta name="Description" content="Abstract Climate models have consistently projected a drying trend in the southwestern United States, aiding speculation of increasing dust storms in this region. Long&#x2010;term climatology is essential...">

```

<title>Intensified dust storm activity and Valley fever infection in the southwestern United States - Tong - 2017 - Geophysical Research Letters - Wiley Online Library</title>



# PROJECT WORKFLOW

1. OBTAIN ORCID iD
  - <https://orcid.org/>
2. OBTAIN A DATA MANAGEMENT CHECKLIST
  - <https://www.usgs.gov/products/data-and-tools/data-management/data-management-plans>
3. HAVE A MEETING
4. DISCUSS LICENCE ISSUES
5. EXECUTE

# F.A.I.R. DATA MANAGEMENT FOR OPEN SOURCE

What, where, and how are you going to put stuff?

## DATA FORMAT

### Geospacial

GML, GeoTIFF, DBF

### Text

geojson, csv

Sustainable Digital File Formats from  
the Library of Congress

<https://www.loc.gov/preservation/digital/formats/fdd/descriptions.shtml>

## USGS

Data dictionary

<https://pubs.usgs.gov/of/2003/of03-001/html/docs/datadict.htm>

Metadata Wizard

<https://www.usgs.gov/software/metadata-wizard-20>

## DATA REPOSITORIES

Zenodo

Figshare

Mendelay

OSF

<https://zenodo.org/record/3946720#.YFsu-q9KhjV>

<https://fairsharing.org/collection/GeneralRepositoryComparison>

# F.A.I.R. DATA MANAGEMENT FOR OPEN SOURCE

What, where, and how are you going to put stuff?

## AVAILABLE METADATA

### USGS

Data dictionary

<https://pubs.usgs.gov/of/2003/of03-001/htmldocs/datadict.htm>

Metadata Wizard

<https://www.usgs.gov/software/metadata-wizard-20>

# GeoJSON vs CSV

## Dataset

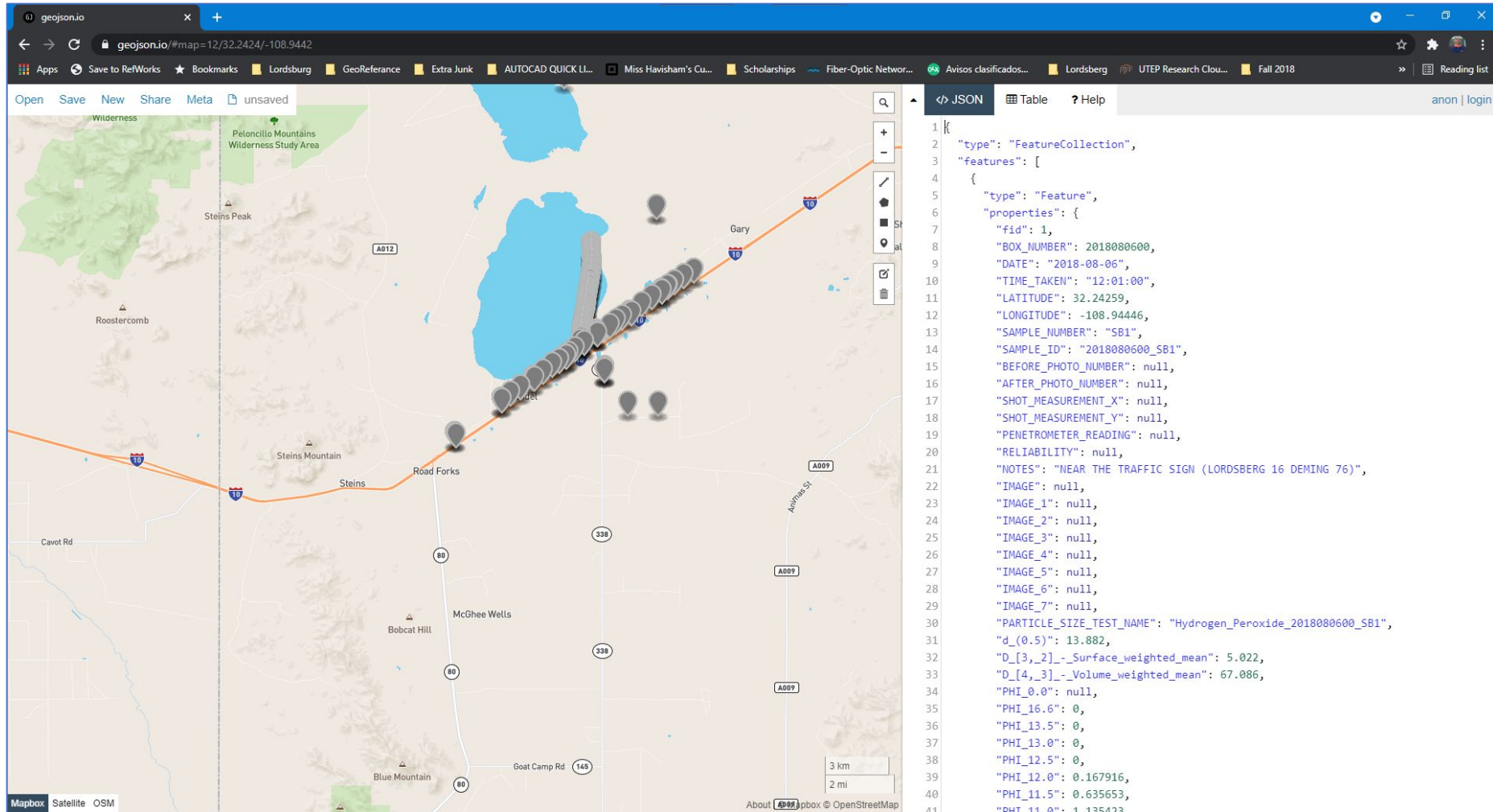
386 lines of sample data containing: Locations (Longitude, Latitude in WGS84 datum), Images, Penetrometer readings, Shot measurements, Notes, and Particle size measurements (Weighted averages)

|              | GeoJSON   | CSV                        |
|--------------|---|----------------------------|
| Form         | Java Script Object Notation   | Comma Separated Values     |
| Data Size    | 525KB   | 114 KB                     |
| Hierarchical | Data has a hierarchy that can use dictionaries and lists while using python or QGIS | Not able to show Hierarchy |
| Extensions   | .geojson or .geojson-ld   | .csv or .geocsv            |
| Scalability  | Very Scalable   | No support for Scalability |

Other formats can be found from the Open Geospatial Consortium (OGC)  
<https://www.ogc.org/docs/is>



# GeoJSON vs CSV



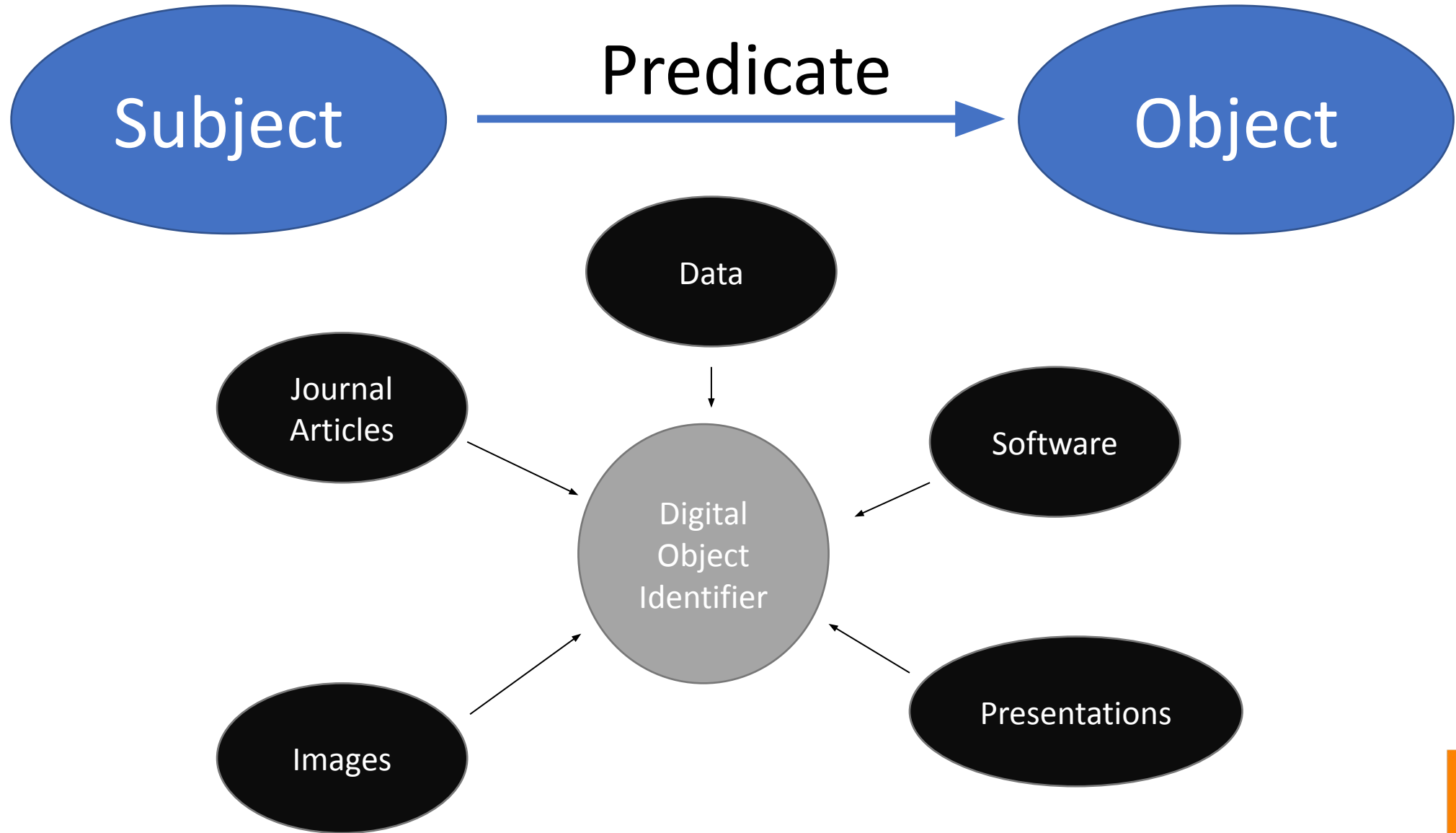
The screenshot displays the geojson.io web application. On the left, a map shows a road with a series of grey circular markers representing data points. The map includes labels for various locations such as Steins Peak, Steins Mountain, Road Forks, McGhee Wells, and Blue Mountain. On the right, a JSON data panel shows the GeoJSON structure for the selected feature. The JSON is a FeatureCollection containing a single Feature with the following properties:

```
1 [{"type": "FeatureCollection",
2   "features": [
3     {
4       "type": "Feature",
5       "properties": {
6         "fid": 1,
7         "BOX_NUMBER": 2018080600,
8         "DATE": "2018-08-06",
9         "TIME_TAKEN": "12:01:00",
10        "LATITUDE": 32.24259,
11        "LONGITUDE": -108.94446,
12        "SAMPLE_NUMBER": "SB1",
13        "SAMPLE_ID": "2018080600_SB1",
14        "BEFORE_PHOTO_NUMBER": null,
15        "AFTER_PHOTO_NUMBER": null,
16        "SHOT_MEASUREMENT_X": null,
17        "SHOT_MEASUREMENT_Y": null,
18        "PENETROMETER_READING": null,
19        "RELIABILITY": null,
20        "NOTES": "NEAR THE TRAFFIC SIGN (LORDSBURG 16 DEMING 76)",
21        "IMAGE": null,
22        "IMAGE_1": null,
23        "IMAGE_2": null,
24        "IMAGE_3": null,
25        "IMAGE_4": null,
26        "IMAGE_5": null,
27        "IMAGE_6": null,
28        "IMAGE_7": null,
29        "PARTICLE_SIZE_TEST_NAME": "Hydrogen Peroxide_2018080600_SB1",
30        "d_0.5": 13.882,
31        "D_3_2_-_Surface_weighted_mean": 5.022,
32        "D_4_3_-_Volume_weighted_mean": 67.086,
33        "PHI_0.0": null,
34        "PHI_16.6": 0,
35        "PHI_13.5": 0,
36        "PHI_13.0": 0,
37        "PHI_12.5": 0,
38        "PHI_12.0": 0.167916,
39        "PHI_11.5": 0.635653,
40        "PHI_11.0": 1.135423
41      }
42    }
43  ]
44 }
```

<https://geojson.io>



# LINKED DATA



## Works Cited

Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016).