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

(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
Abstract
Climate models have consistently projected a drying trend in the southwestern United States, aiding speculation of increasing dust storms in this region. Long-term climatology is essential to understanding the climatic conditions that lead to observed climate variability. We have reconstructed a long-term dust climatology in the western United States, based on a comprehensive dust identification method and validated against observational datasets from the Intercontinental Monitoring of Aerosol-Cloud-Climate (IMACC) network. We report here a direct evidence of rapid intensification of dust storm activity over apparent decreases in long-term trends, driven by large-scale variations in soil moisture anomalies and dust transport from the southwestern United States to the Pacific Ocean. This dust trend has increased from 1980s to 2010s, in agreement with observations of increased dust storms in the southwestern United States.

Plain Language Summary
Computer models predict that the Earth’s climate will become drier in the dry areas of the Earth, such as the southwestern United States. This increase in dryness has led to an increase in dust storms in the region. Understanding the long-term climatology of dust storms is essential to understanding the factors that lead to dust storms. We have reconstructed a long-term dust climatology for the western United States, based on a comprehensive dust identification method and validated against observational datasets from the Intercontinental Monitoring of Aerosol-Cloud-Climate (IMACC) network. We report here a direct evidence of rapid intensification of dust storm activity over apparent decreases in long-term trends, driven by large-scale variations in soil moisture anomalies and dust transport from the southwestern United States to the Pacific Ocean. This dust trend has increased from 1980s to 2010s, in agreement with observations of increased dust storms in the southwestern United States.
PROJECT WORKFLOW

1. OBTAIN ORCID iD
   • https://orcid.org/

2. OBTAIN A DATA MANAGEMENT CHECKLIST

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?

<table>
<thead>
<tr>
<th>DATA FORMAT</th>
<th>DATA REPOSITORIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geospatial</td>
<td>Zenodo</td>
</tr>
<tr>
<td>GML, GeoTIFF, DBF</td>
<td>Figshare</td>
</tr>
<tr>
<td>Text</td>
<td>Mendelay</td>
</tr>
<tr>
<td>geojson, csv</td>
<td>OSF</td>
</tr>
</tbody>
</table>

Sustainable Digital File Formats from the Library of Congress
https://www.loc.gov/preservation/digital/formats/fdd/descriptions.shtml

USGS
Data dictionary
Metadata Wizard
https://www.usgs.gov/software/metadata-wizard-20

https://zenodo.org/record/3946720#.YFsu-q9KhiV
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
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)

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

Other formats can be found from the Open Geospatial Consortium (OGC) [https://www.ogc.org/docs/is](https://www.ogc.org/docs/is)
GeoJSON vs CSV

https://geojson.io
The term “Linked Data” refers to a set of best practices for publishing and connecting structured data on the Web. These best practices have been adopted by an increasing number of data providers over the last three years, leading to the creation of a global data space containing billions of assertions— the Web of Data.

LINKED DATA

Subject

Predicate

Object

Data

Journal Articles

Digital Object Identifier

Images

Software

Presentations

https://www.w3.org/wiki/GeoRDF