Copyright:

CC-BY allows reusers to distribute, remix, adapt, and build upon the material in any medium or format so long as attribution is given to the creator. This license allows for commercial use.

- Perhaps the best copyright for myself would be CC-BY, credit and attribution, but open to all

Open Data is a cornerstone of OS. An argument for how is the ability to make more conclusions after an initial experiment is done. By providing your data, others can interpret or find new evidence from your collection.

GeneLab- Open data from NASA to study Bone loss, vertigo, anemia, muscle atrophy, increased risk for cancer of astronauts in space.

Using vast space-flown animal and human samples as well as astronaut biological data from GeneLab, a recent paper published in *Nature* found what may be the culprit behind many of the side effects from travel to space: mitochondrial stress. (Clyde, 2021)

Mitochondria are components within our cells that affect respiratory and energy function. This discovery could be crucial to overcoming human health-related problems in space. Understanding the source of this issue could help scientists develop countermeasures and therapies to keep people healthy in space for longer periods of time.

https://www.youtube.com/watch?v=c9moR-KQpDQ : Twins, one in space and one on the ground, a study that explores genetic changes from being in space and explores the idea of gene-unique space treatments.
DATA TYPES:

Primary (raw) data refers to data that is directly collected or created by researchers. Research questions guide the collection of the data. Typically, a researcher will formulate a question, develop a methodology and start collecting the data. Some examples of primary data include:

- Responses to interviews, questionnaires, and surveys.
- Data acquired from recorded measurements, including remote sensing data.
- Data acquired from physical samples and specimens form the base of many studies.
- Data generated from models and simulations.

Secondary (Processed) data typically refers to data that is used by someone different from who collected or generated the data. Often, this may include data that has been processed from its raw state to be more readily usable by others.

Metadata is a special type of data that describes other data or objects (i.e. samples). It is often used to provide a standard set of information about a dataset to enable easy use and interpretation of the data.

Published data is data shared to address a particular scientific study and/or for general use.

Good data is

- Sufficiently described with appropriate metadata, which greatly affects open data reusability. There is no one-size-fits-all for metadata, as its collection is guided by your data.
- Have the appropriate license, copyright, and citation information.
- Have appropriate access information.
- Be findable in an accredited or trustworthy resource.
- Be accompanied by a history of changes and versioning.
- Include details of all processing steps

Benefits described by the author: Doing open science not only lets other people understand and reproduce your results, but it lets you do the same. Implementing open science principles like good documentation and version control helps you, potential collaborators, and everyone else to understand your results. In 2 hours, 2 weeks, or 2 years, you will still be able to understand what you did.

Restrictions on Sharing Data (When not to share data)

- Some data should only be shared very carefully or not at all. Reasons not to share can include:

- Data includes a country's military secrets or violations of national interests.
- Data includes private medical information or an individual's personally identifiable data.
- Data includes intellectual property.

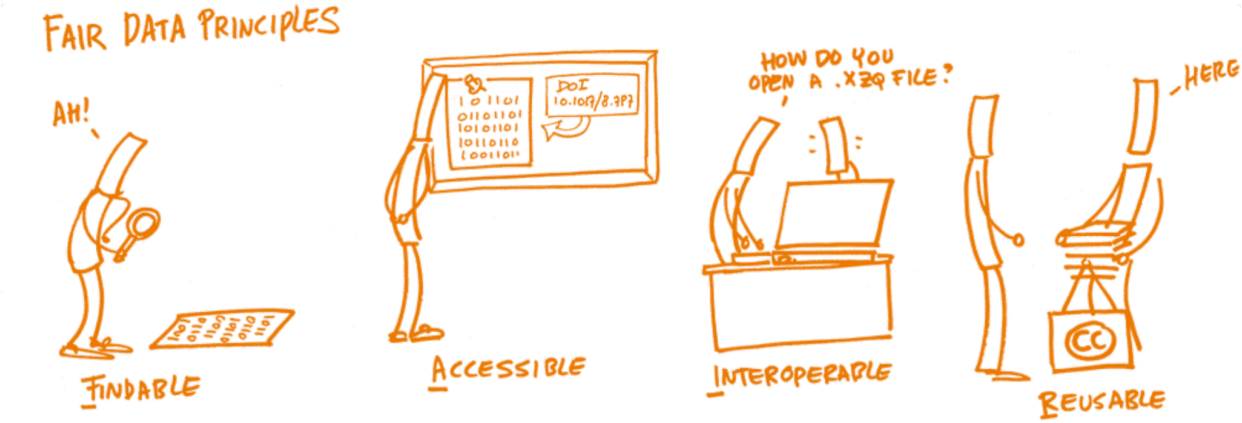

Image Credit: A cartoon describing FAIR data sharing by Patrick Hochstenbach, is licensed under CC0 1.0

From the A in fair and regarding restrictions, If the full content **cannot be made openly available** for any reason (e.g. data sensitivity, infrequent data access, file storage issues), the metadata can still be made openly available so that users can find out who they need to contact to request the data (if possible).

Using open data- and vetting it.
Open data isn't always simple to use in your research. Sometimes there are multiple versions of the same dataset, so learning how to discover, assess, and then use open data will help you save time.

What is the first and best way to find research data? Ask your community! This includes your research advisor, colleagues, team members, and people online. Knowing where to find reliable, good data is as much a skill as any lab technique

Questions to ask when assessing if data is good to use: (skill questions)

- Is the data well described?

- Is the reason the data is collected clear? Is the publisher's use for the data clear?
- Are any other existing uses of the data outlined?
- Is the data accessible?
- Is the data timestamped or up to date?
- Will the data be available for at least a year?
- Will the data be updated regularly?
- Is there a quality control process?
- Is the data available in a format appropriate for the content?
- Is the data available from a consistent location?
- Is the data well-structured and machine-readable?
- Are complex terms and acronyms in the data defined?
- Does the data use a schema or data standard?
- Is there an API available for accessing the data?
- What tools or software are needed to use this data?
- Is there an existing community of users of the data?
- Is the data already relied upon by large numbers of people?
- Is the data officially supported?
- Are service level agreements available for the data?
- It is clear who maintains and can be contacted about the data?

```
cff-version: 1.2.0
message: "If you use this software, please cite it as below."
authors:
  - family-names: Druskat
    given-names: Stephan
    orcid: https://orcid.org/1234-5678-9101-1121
title: "My Research Software"
version: 2.0.4
doi: 10.5281/zenodo.1234
date-released: 2021-08-11
```

Simple CITATION.cff file [Screenshot] GitHub (https://github.com/) (external link) LINK (EXTERNAL)

How to properly cite code (To add to Laika and Aether projects and continue in future)

- Create a DOI on Zenodo

Is general the information that is commonly found in a citation file includes

A. Authors and their institutions

B. Title

C. ORCiD

D. DOI

E. Version

F. URL

G. Creation date

Selecting Data Formats and Tools for Interoperability

NON-open formats:

Adobe Photoshop (.psd): The default proprietary file format for Adobe Photoshop, a popular image editing software.

AutoCAD Drawing (.dwg): A proprietary data format used for computer-aided design (CAD).

Microsoft Word (.doc/.docx):

Alternatives to these would be using a JPEG, a STEP, and a PDF or plaintext document. For cad on my cad library this should be a STP library. Often docs are already posted as pdf.

Metadata is structured, standardized, and machine-readable. Documentation is unstructured and can be any format (often a text file that accompanies the data).

To better understand documentation and metadata, let's take an example of an online recipe. Many online recipes start with a long description and history of the recipe, and perhaps cooking or baking tips for the dish, before listing ingredients and step-by-step cooking instructions.

- The ingredients and instructions are like metadata. They can be indexed and searched via Google and other search engines.
- The descriptive text that includes background and context for the recipe are like documentation. They are more free-form and not standardized.

NASA Metadata addendum:

SPD-41a requires metadata fields that validate the scientific conclusions of publications. Metadata should:

- Be robust, meet standards, and clearly and explicitly describe the data.

- Have replicability and/or can be combined in different settings.
- Include information on how the data were collected (e.g. which equipment/instruments were used).
- Include information on which variables/parameters were included in this dataset.

License	Type of use	Symbols
Public Domain CC0	The copyright holder has voluntarily waived their rights and the work can be freely used.	PUBLIC DOMAIN
Attribution (BY)	You must credit the creator, the title and the license the work is under.	BY
Share Alike (SA)	Any new material produced using the work must be made available under the same license as the original work.	BY SA
Non Commercial (NC)	The work cannot be used for commercial purposes.	BY NC
No Derivatives (ND)	The work can only be used exactly how it is. The work cannot be adapted or modified in any way.	BY ND

Credit: NASA

- Always add documentation that enables other researchers to assess the relevance and reusability of your product. This includes metadata, README files, and version control details.
- It is important to assign an open license to your data to enable reuse. There are several different types of open licenses that can fit your needs.

Data sharing, things to consider:

Data sharing should typically be done through **a long-term data center or repository** which will be responsible for ingesting, curating, and distributing/publishing your open data. You are responsible for providing information/metadata to help make your data readily discoverable, accessible, and citable. **The cost of archiving and publishing data should also be considered. Some repos like github are free (Where I store code) publication may be upwards of 4k!**

Open Data Sharing Process

In general, sharing open data requires the following steps:

- ☐ Make sure your data can be shared.
- ☐ Select or identify a repository to host your data.
- ☐ Work with your repository to follow their process and meet their requirements.
- ☐ Make sure your data is findable and accessible through the repository and that it can be maintained and archived.
- ☐ Request a DOI for your data set so that it is easily citable.
- ☐ Choose a data license.

NASA FAR Supplement 1852.227 outlines patent and data rights for government contracts.

Making it Easy to Cite Your Data

The goal is to make it easy to cite your data. Best practices include:

Include a citation statement that lists your DOI.

Different repositories and journals have different standards for how to cite data. If your repository encourages it, include a .CFF file with your data that explains how to cite your data.

Clearly identify the data creators and/or their institution in your citation.

This allows users to follow up with the creators if they have questions or discover issues.

Include ORCID iDs of data authors where possible in the citation.

Good practice is to assign a leader in a team who will handle dealing with repositories and ensuring meta data and documentation is up to true standard. This spearhead also is the one to interact with the responsible repo reps and ensure documentation is correct.

Someone(s) should HAVE THESE ROLES:

- Clearly communicate the open protocols needed for the data/metadata.
- Develop the metadata
- Develop the documentation (e.g. README file or report)
- Be responsible for privacy concerns and approval processes for release - is the data appropriately anonymized?
- Be responsible for engagement with communities that data may be about.
- Be responsible for how data can be correctly interpreted by the community.

Tips and sites to write a good DMP (data management plan)

- https://dmptool.org/ (how to)
- http://science.nasa.gov/oss-guidance
- http://opendatahandbook.org/guide/en/

OSDMP Example ThEBES