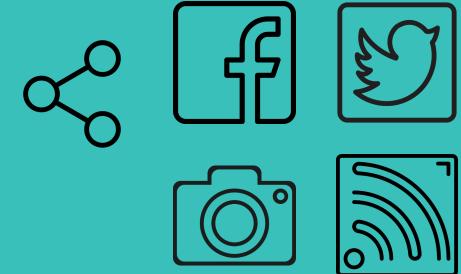


- You are free to
- share, adapt or re-mix
 - photograph, video, or broadcast
 - blog, live-blog or post-video
- this presentation



Provided that

you attribute the work to its author and respect
the rights and licenses associated with its
components



University of
Zurich^{UZH}



Dealing with FAIR data

Open Science Summer School, UZH
14th July 2022

my twitter handle: [pcmasuzzo](#)

my email address: paola.masuzzo@gmail.com

Hi, I am Paola! I am a data scientist, Open Science advocate and independent researcher at [IGDORE](#)

follow this presentation here

<https://tinyurl.com/summerschoolfair>

the HTML document can be accessed here

<https://tinyurl.com/summerschoolfairhtml>

- disclaimers

Thank you, Twitter folks!

The web is full of resources, ping me after the workshop if you cannot find something!

Some of my favourites are:

1. [How To FAIR](#)
2. [FAIR Cookbook](#)
3. [Top 10 FAIR Data & Software Things](#)
4. [FAIR training resources | FAIR Data 101](#)
5. [PARTHENOS Guidelines to FAIRify data management and make data reusable](#)

Pinned Tweet



Paola Chiara Masuzzo 
@pcmasuzzo

Twitter folks! I am giving a workshop on FAIR data at the **#OpenScience** Summer School **@UZH_en**. I obviously have ideas & material I am preparing, but would very much like to ask you: what would you definitely want to see in a workshop on **#FairData**? Which type of exercises? >

We'll be using a collaborative etherpad, where I have already pasted some links and content to make the whole workshop go a bit smoother :) You can find it here, I will share it again later:
https://pad.riseup.net/p/fair_data_workshop



this icon will indicate that it's time for an exercise together!

- What we will be doing



09:15 - 09:45 Theoretical background

09:45 - 10:45 Session I: The FAIR principles in action

10:45 - 11:15 Coffee break

11:15 - 12:00 Session II: FAIRify your research data

12:00 - 12:15 Bonus Track: FAIR in the era of social machines

Theoretical background

A close-up photograph of a row of colored pencils arranged diagonally from top-left to bottom-right. The pencils are sharpened at the top, showing a vibrant rainbow of colors: purple, blue, teal, green, yellow, orange, red, and pink. The wooden bodies of the pencils are visible between the sharpened tips.

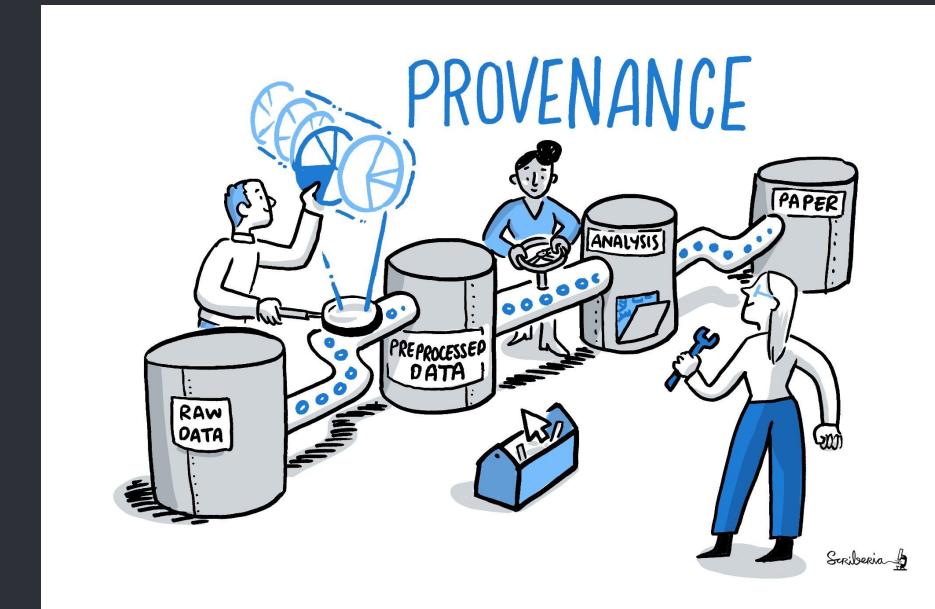
the Open Science
movement encourages
researchers to share
research output beyond
the contents of a published
academic article

A close-up photograph of a row of colored pencils arranged diagonally from top-left to bottom-right. The colors transition through a full spectrum: purple, blue, teal, green, yellow, orange, red, and pink. The pencils have light brown wooden bodies and sharp, pointed tips.

the Open Science
movement encourages
researchers to share
research data beyond
the contents of a published
academic article

- what is research data?

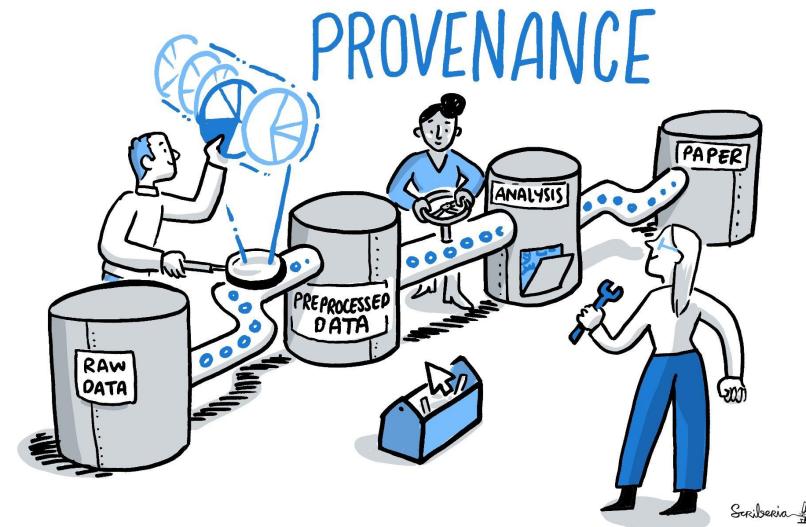
Any type of information that is collected, observed, or created, in the context of research:



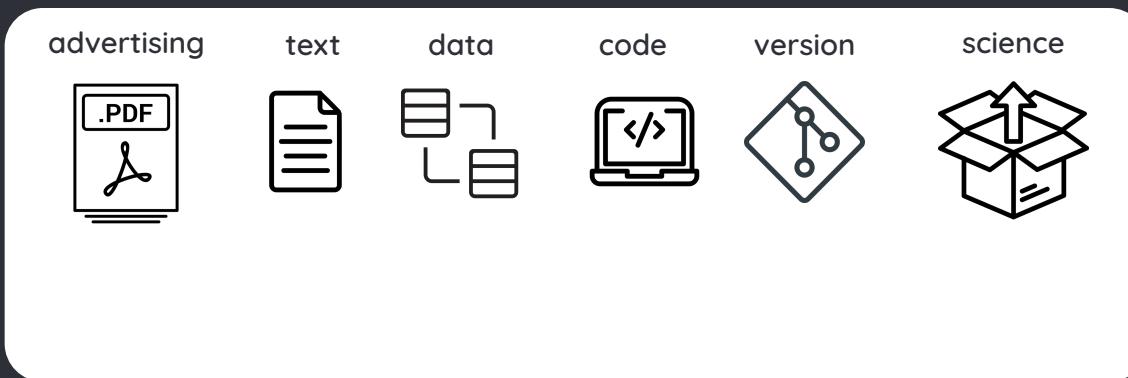
- what is research data?

Any type of information that is collected, observed, or created, in the context of research:

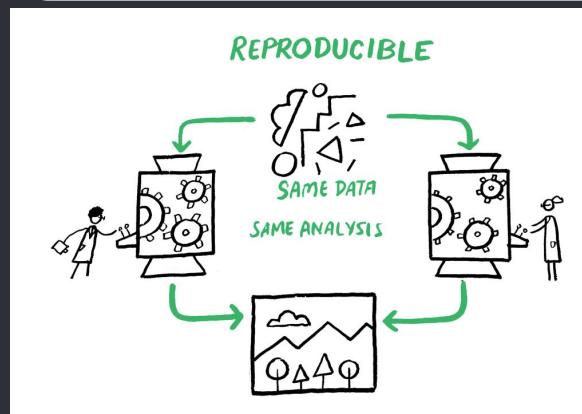
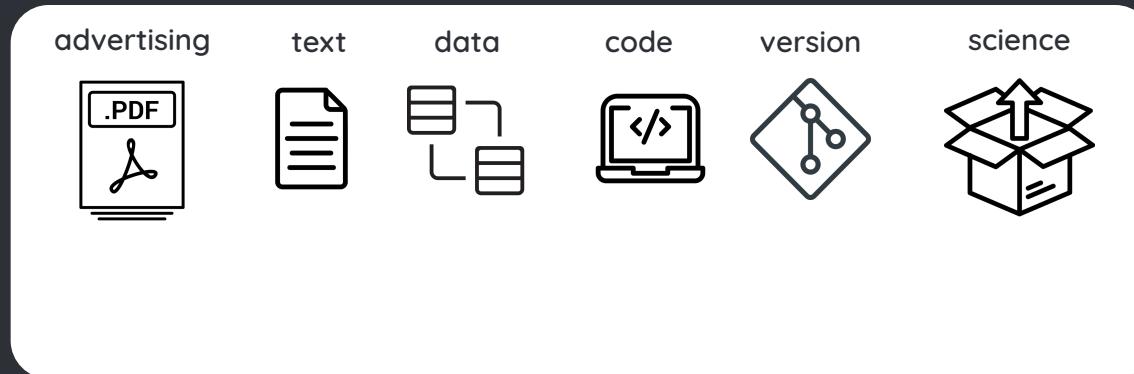
- Primary: raw data from measurements or instruments
- Secondary: processed from secondary analysis and interpretations
- Published: final format available for use and reuse
- Metadata: data about your data



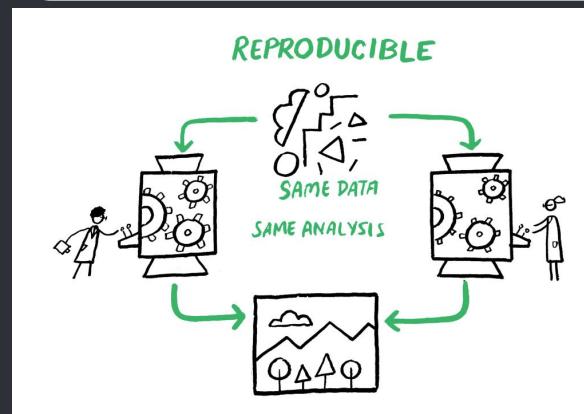
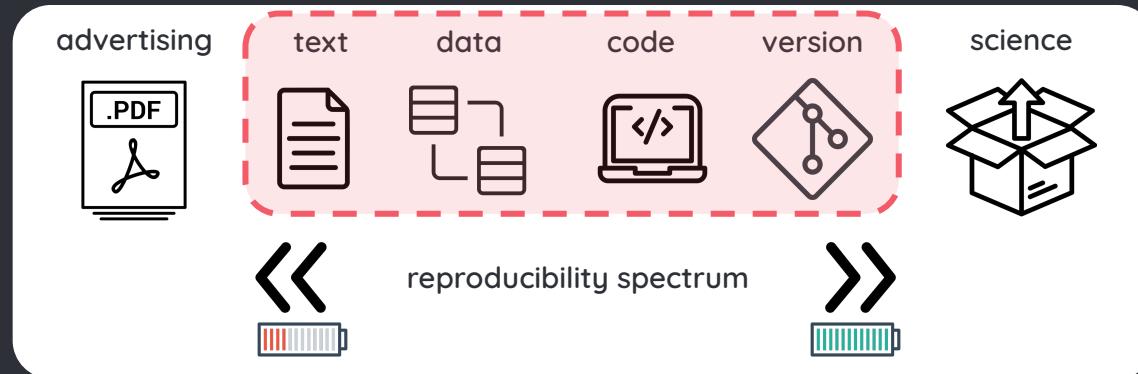
- research: much more than a PDF



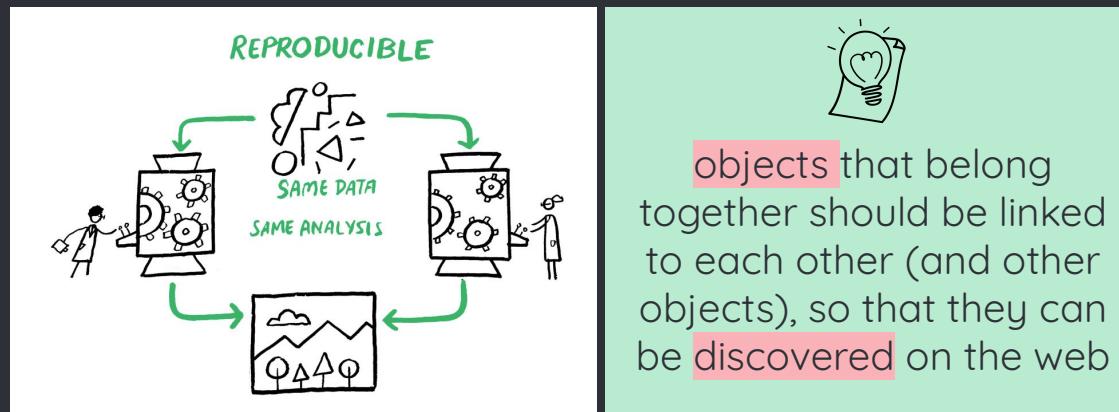
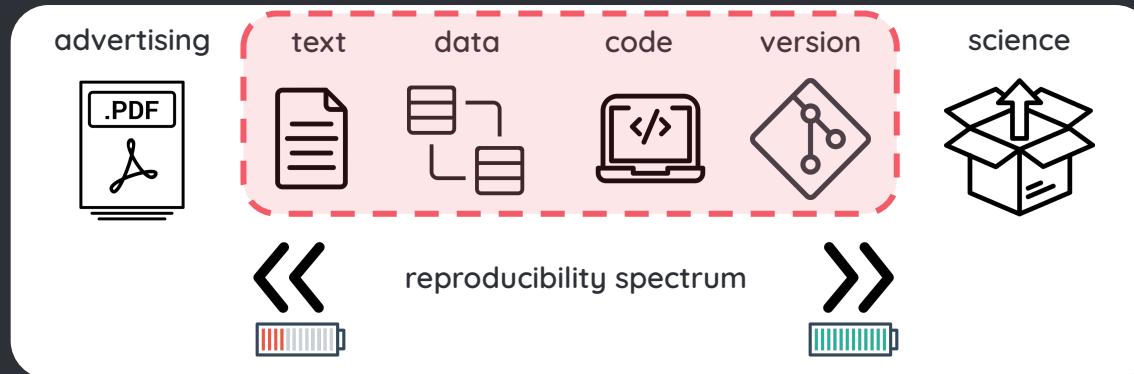
- reproducibility: minimum standard for research validity



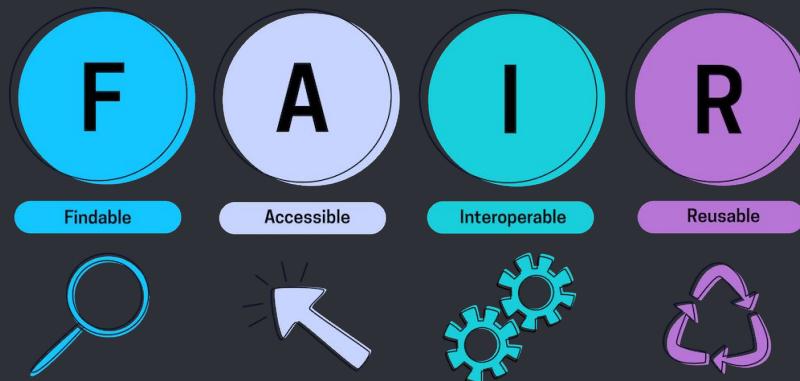
- publishing research objects enables reproducibility



- publishing research objects on the web: how?

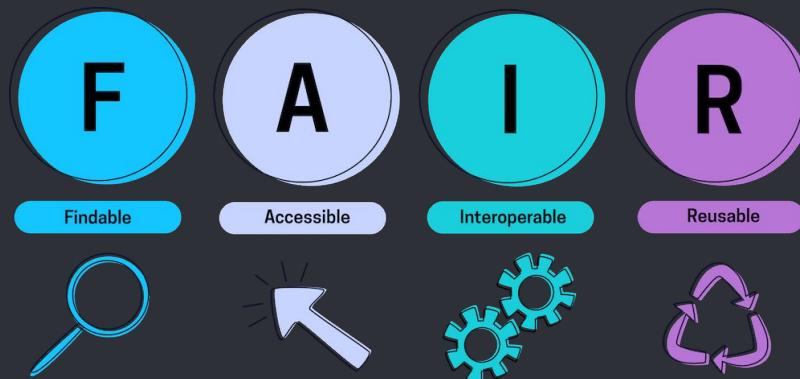


- the FAIR principles: guidance for data stewardship



the FAIR principles have been designed to assist *discovery* and *reuse* of research objects through the web

- the FAIR principles: guidance for data stewardship



the FAIR principles have been designed to assist **discovery** and **reuse** of research objects through the web

IMPORTANT:

FAIR comes in degrees
FAIR = agnostic of technical implementations
FAIR requires work / labour

FAIR is not the same as OPEN

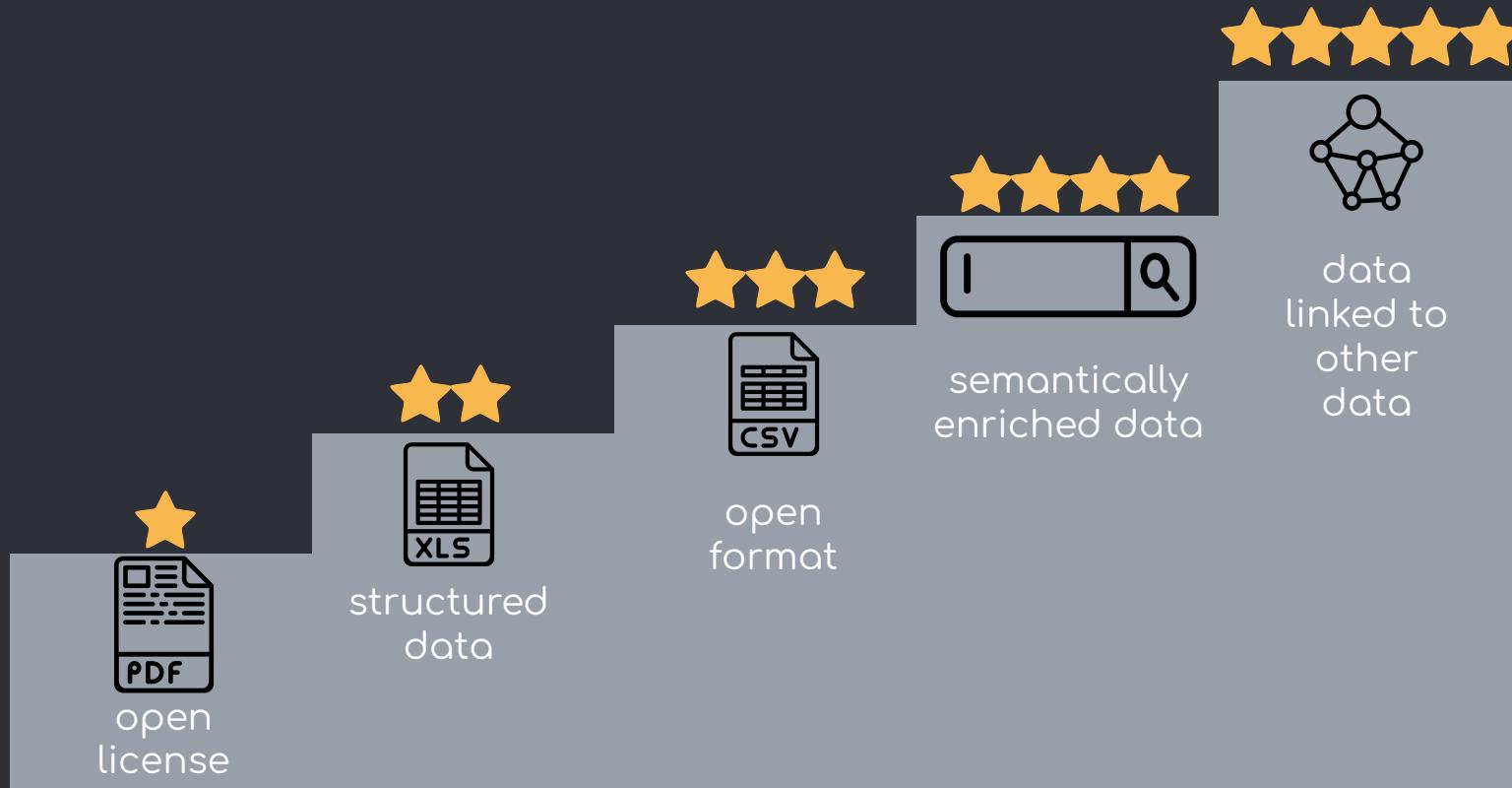


- definition of open data

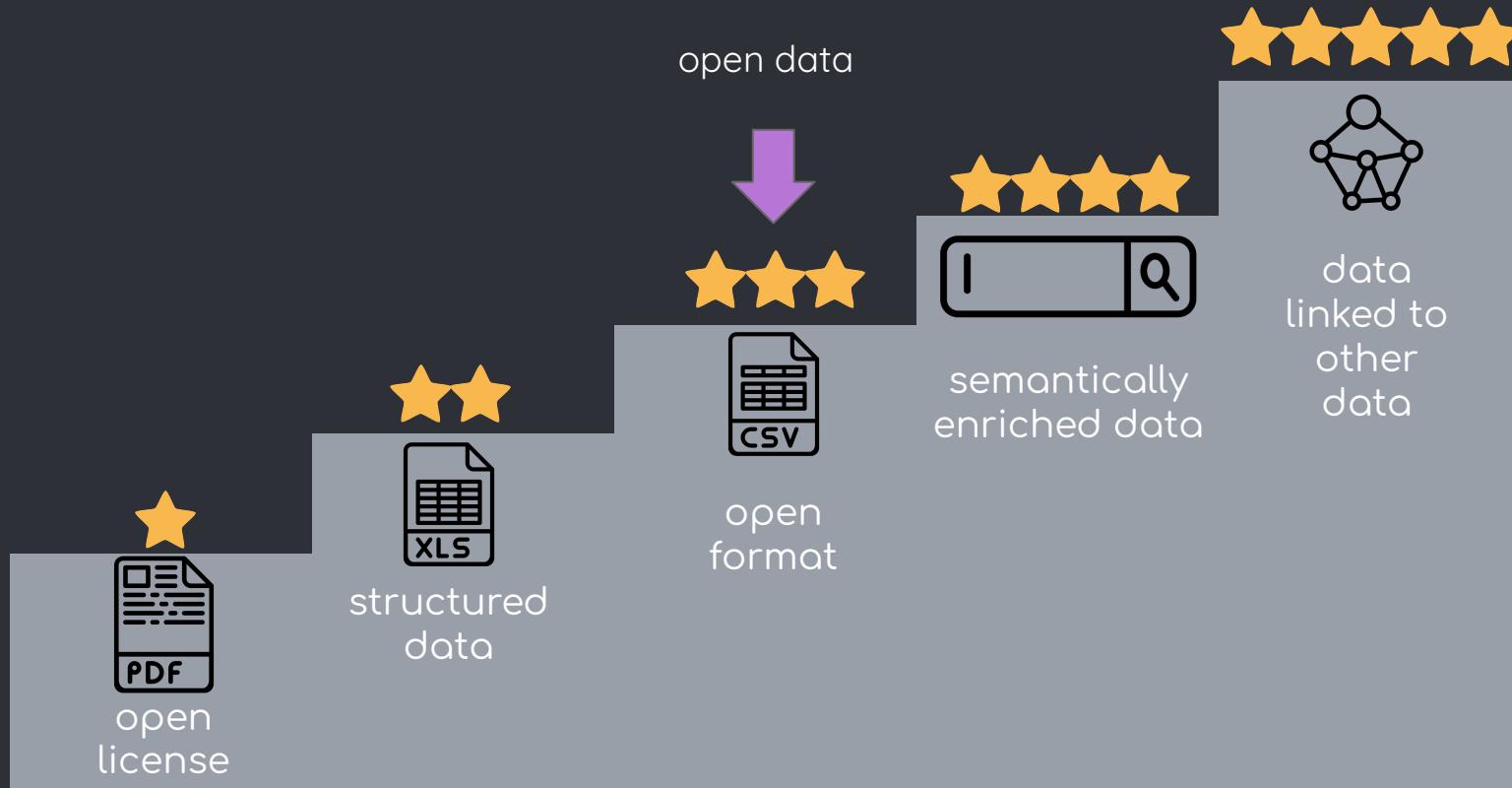
Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike.



- open data and the five-star model



- open data and the five-star model



- open data is not FAIR data, and vice versa

FAIR is not equivalent of OPEN, but OPEN data needs to be FAIR to be useful

Making your data freely available on the web doesn't translate to it being reusable

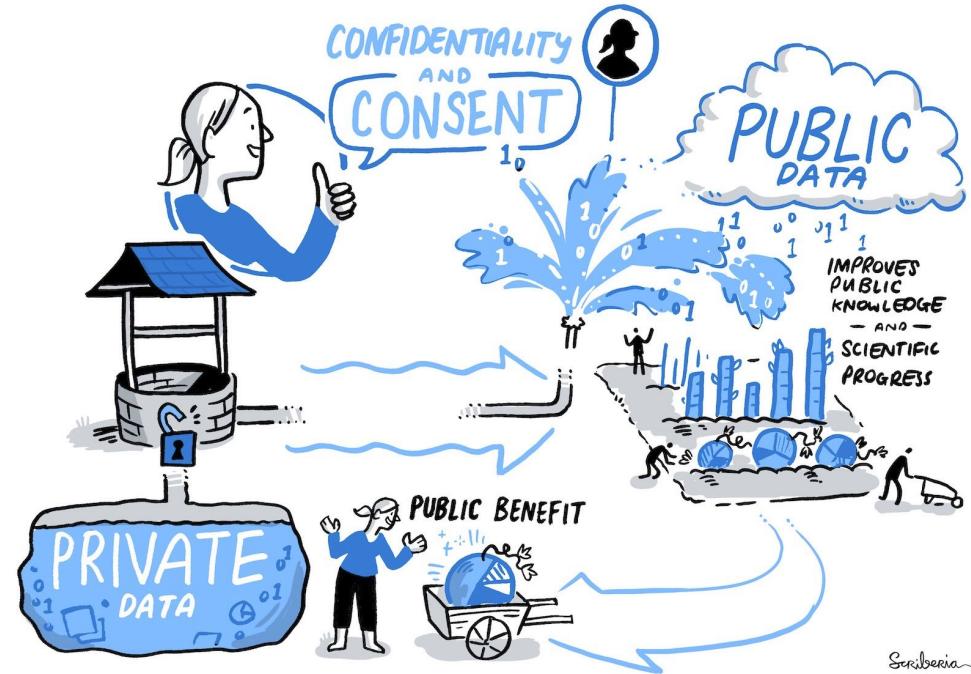


FAIR is not the same as OPEN

- open data is not FAIR data, and vice versa

Even confidential
and highly
protected
datasets can be
FAIR

⇒ as open as
possible, as closed
as necessary



- open data is not FAIR data, and vice versa

FAIR is not equivalent of OPEN, but OPEN data needs to be FAIR to be useful

Making your data freely available on the web doesn't translate to it being reusable

Even confidential and highly protected datasets can be FAIR \Rightarrow as open as possible, as closed as necessary

Ideally, you want FAIR data shared openly!

- research data life cycle



- research data life cycle & some terminology

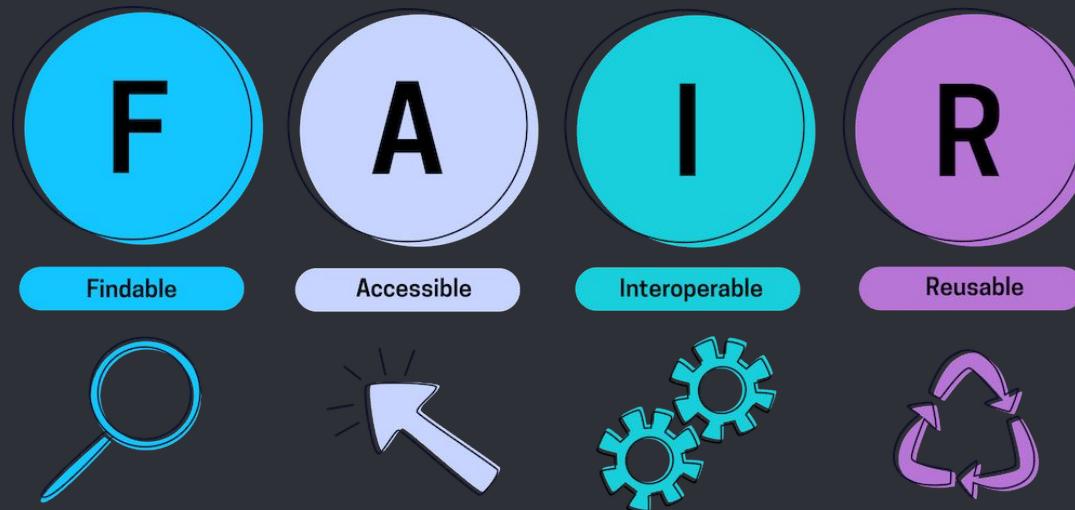


data management → activities during a project to collect, annotate, and archive data

data stewardship → making data reusable for the long-term, also after the project has ended (data preservation)

data curation → creating, organizing and maintaining data sets, so that these can be accessed and used by people looking for information (part of the data management process)

- the FAIR principles: guidance for data stewardship



time for questions



the link to the collaborative pad is:
https://pad.riseup.net/p/fair_data_workshop

Session I: the FAIR principles in action

- F = Findable

data & metadata should be
easy to find for both humans
and computers

machine-readable metadata
are essential for automatic
discovery of datasets and
services

sufficiently rich metadata &
unique and persistent
identifiers need to be used



Findable

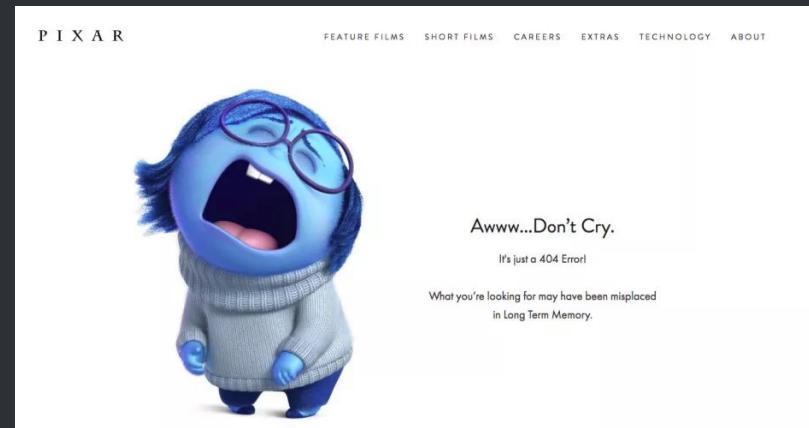


- F = Findable

data & metadata should be **easy to find** for both humans and computers

machine-readable metadata are essential for automatic discovery of datasets and services

sufficiently rich metadata & **unique and persistent identifiers** need to be used



- persistent identifiers

A persistent identifier (PID) is a long-lasting reference to a digital resource and provides the information required to reliably identify, verify and locate your research data eliminating many misunderstandings.

PIDs are sometimes described as a social security number for a research object. Another analogy which might be helpful when thinking about PIDs is with a statue, unique, long-lasting and robust.

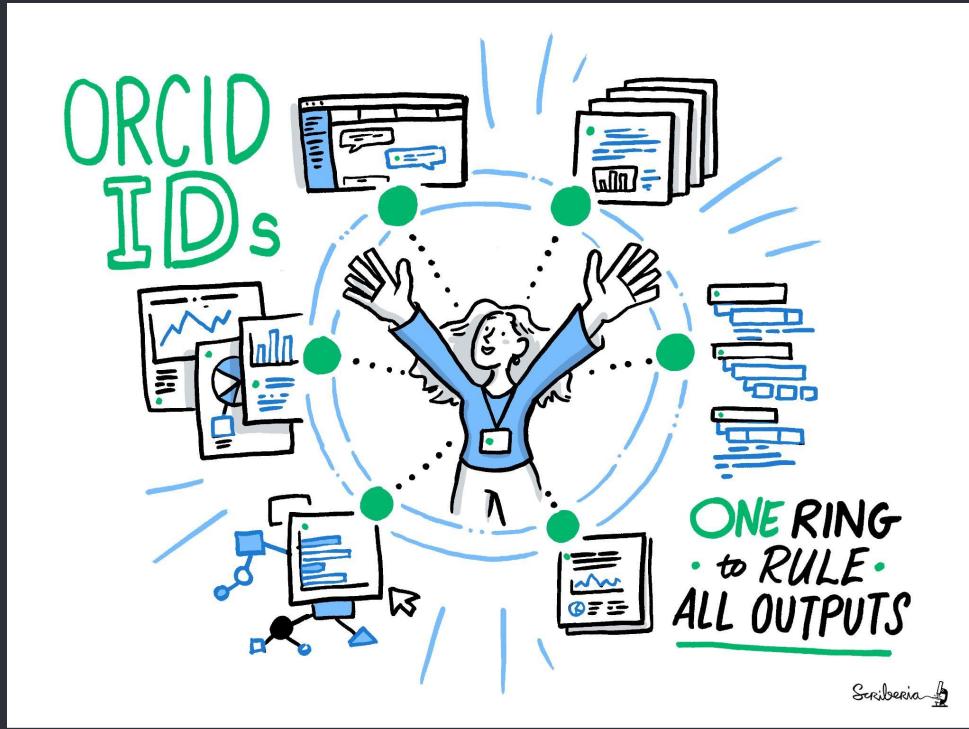
- persistent identifiers

A persistent identifier (PID) is a long-lasting reference to a digital resource and provides the information required to reliably identify, verify and locate your research data eliminating many misunderstandings.

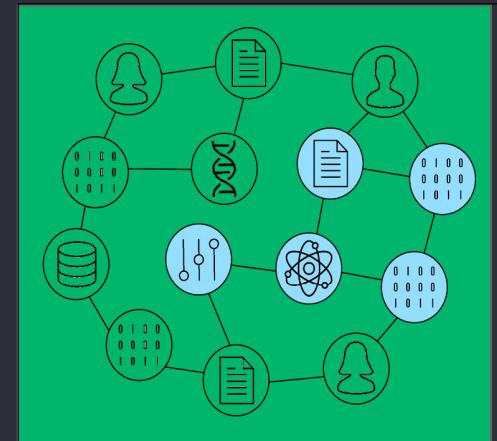
PIDs are sometimes described as a social security number for a research object. Another analogy which might be helpful when thinking about PIDs is with a statue, unique, long-lasting and robust.

Common PID are the [Digital Object Identifier](#) (DOI) and the [Handle System](#) which can both be assigned to data to identify them uniquely. While [DOIs](#) are mainly assigned to resources ready for public dissemination, [Handles](#) are in general used to persistently identify other categories of digital resources (e.g. those created in the labs) to make them referable by software, workflows etc.

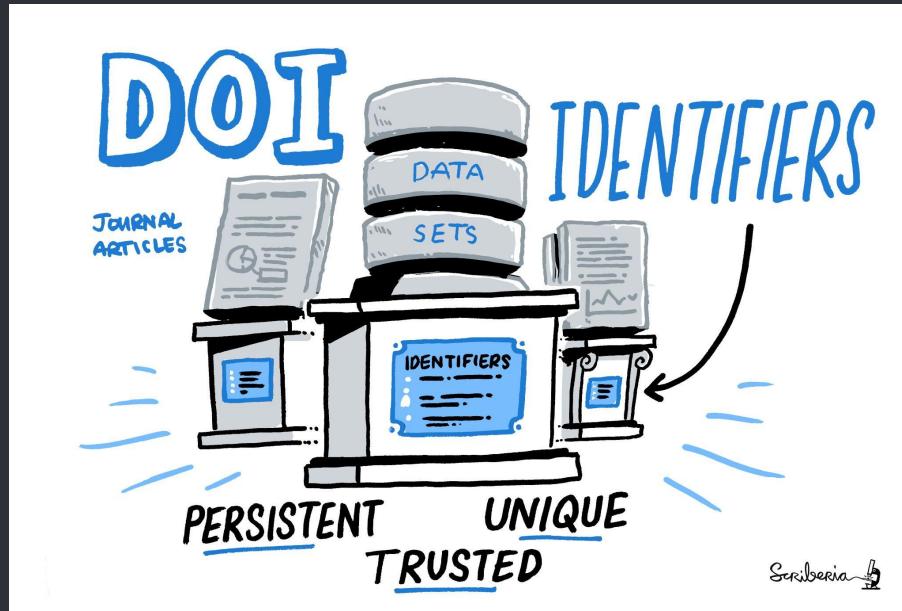
- a PID for researchers: the Open Researcher and Contributor ID



<https://orcid.org/0000-0003-3699-1195>



- anatomy of a DOI



<https://doi.org/10.5281/zenodo.3679141>

resolver
service



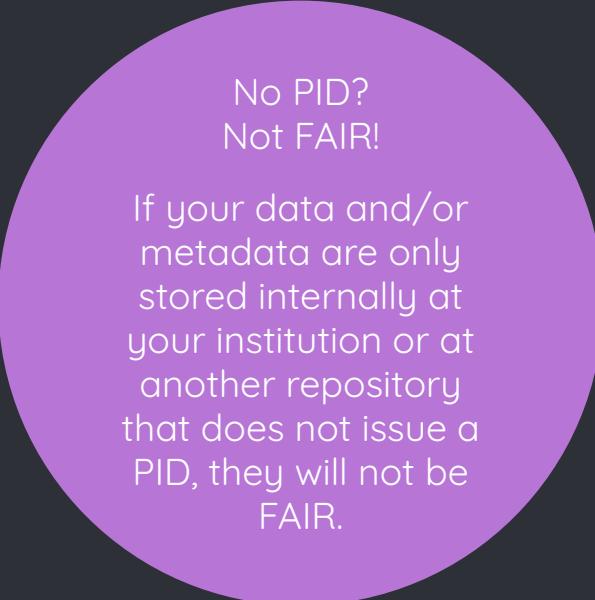
directory
indicator
+prefix
(assigning
body)

suffix
resource



Go to <https://www.doi.org/> and resolve a DOI name, you can pick some examples from <https://www.doi.org/demos.html>, or you can use this DOI: 10.5281/zenodo.3679141

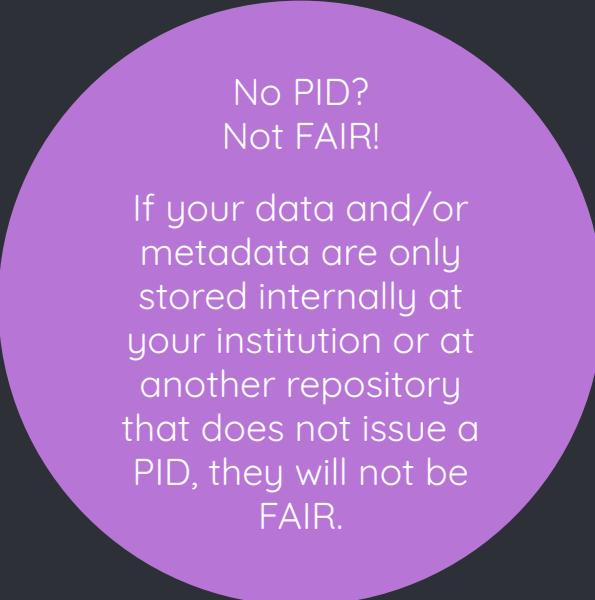
- no PID? Not FAIR!



No PID?
Not FAIR!

If your data and/or metadata are only stored internally at your institution or at another repository that does not issue a PID, they will not be FAIR.

- no PID? Not FAIR!

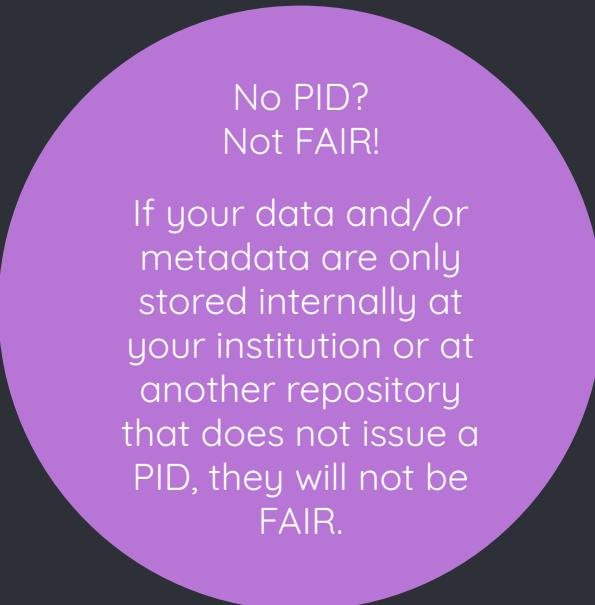


No PID?
Not FAIR!

If your data and/or metadata are only stored internally at your institution or at another repository that does not issue a PID, they will not be FAIR.

OK, so how do you get a PID?

- no PID? Not FAIR!



No PID?
Not FAIR!

If your data and/or metadata are only stored internally at your institution or at another repository that does not issue a PID, they will not be FAIR.

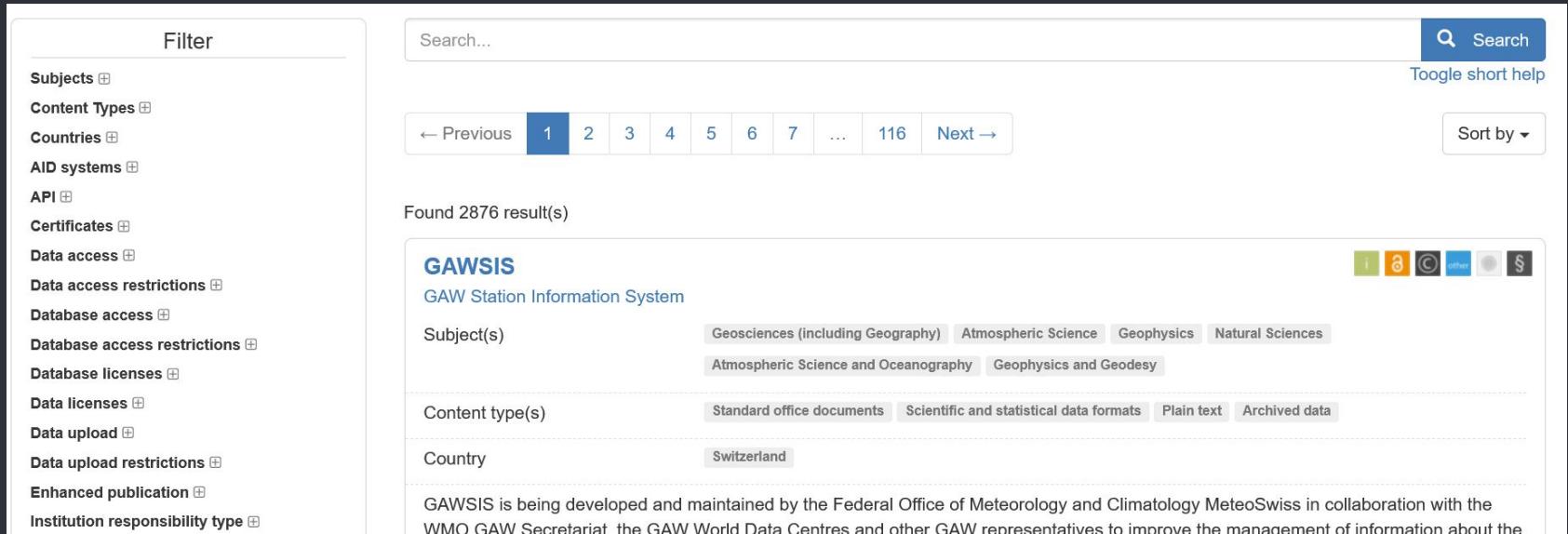
OK, so how do you get a PID?

deposit your data in a trusted repository that issues a PID

- institutional repository
- domain specific repository
- general-purpose repository



re3data listed repositories



The screenshot shows the re3data search interface. On the left, there is a sidebar titled "Filter" with various categories like Subjects, Content Types, Countries, AID systems, API, Certificates, Data access, Data access restrictions, Database access, Database access restrictions, Database licenses, Data licenses, Data upload, Data upload restrictions, Enhanced publication, and Institution responsibility type. In the center, there is a search bar with a placeholder "Search..." and a "Search" button. Below the search bar is a navigation bar with buttons for "← Previous", page numbers (1, 2, 3, 4, 5, 6, 7, ..., 116), and "Next →". To the right of the search bar is a "Sort by" dropdown. The main content area displays the results for "GAWSIS". It includes the title "GAWSIS", a subtitle "GAW Station Information System", and a row of icons. Below this, there are three sections: "Subject(s)" with categories like Geosciences (including Geography), Atmospheric Science, Geophysics, Natural Sciences, Atmospheric Science and Oceanography, and Geophysics and Geodesy; "Content type(s)" with categories like Standard office documents, Scientific and statistical data formats, Plain text, and Archived data; and "Country" with Switzerland. At the bottom, there is a note about GAWSIS being developed by MeteoSwiss and others.

Found 2876 result(s)

GAWSIS

GAW Station Information System

Subject(s)

- Geosciences (including Geography)
- Atmospheric Science
- Geophysics
- Natural Sciences
- Atmospheric Science and Oceanography
- Geophysics and Geodesy

Content type(s)

- Standard office documents
- Scientific and statistical data formats
- Plain text
- Archived data

Country

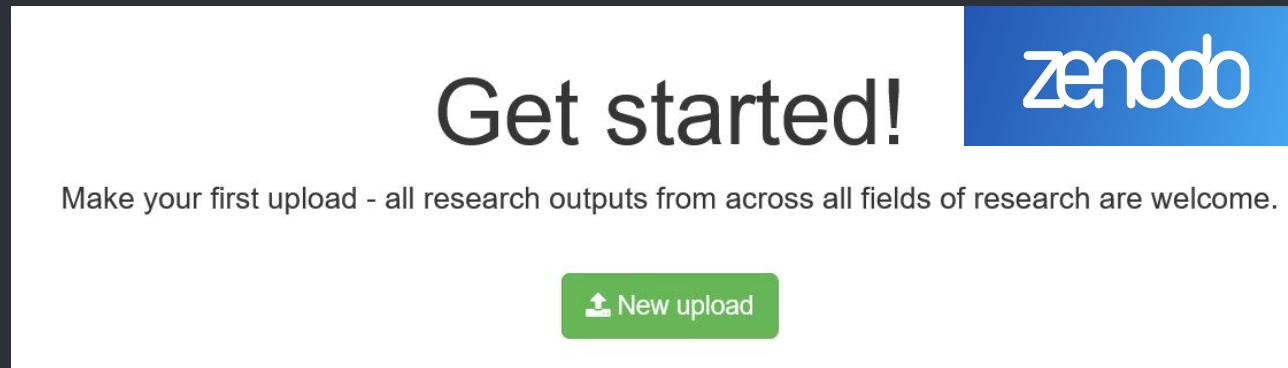
Switzerland

GAWSIS is being developed and maintained by the Federal Office of Meteorology and Climatology MeteoSwiss in collaboration with the WMO GAW Secretariat, the GAW World Data Centres and other GAW representatives to improve the management of information about the



Go to <https://www.re3data.org/search> and look for a repository that could host the type of data of your research; check the PID & the data access fields

- Zenodo: a general-purpose repository



The image shows the 'Get started!' page of the Zenodo website. It features a large 'Get started!' heading, a subtext 'Make your first upload - all research outputs from across all fields of research are welcome.', and a green 'New upload' button with an upward arrow icon. To the right of the main content area is a blue box containing the Zenodo logo.



note: Zenodo sandbox is a test environment:
your data will not
be permanent,
will be deleted
after some time!!!



Go to <https://sandbox.zenodo.org/>, register or login, we'll need the sandbox later today.



Accessible



- A = Accessible

accessible does not imply
open

data & metadata need to be
retrievable by their identifier
using a standardized
communications protocol

research repositories often
use the OAI-PMH or REST API
protocols to interface with
data in the repository

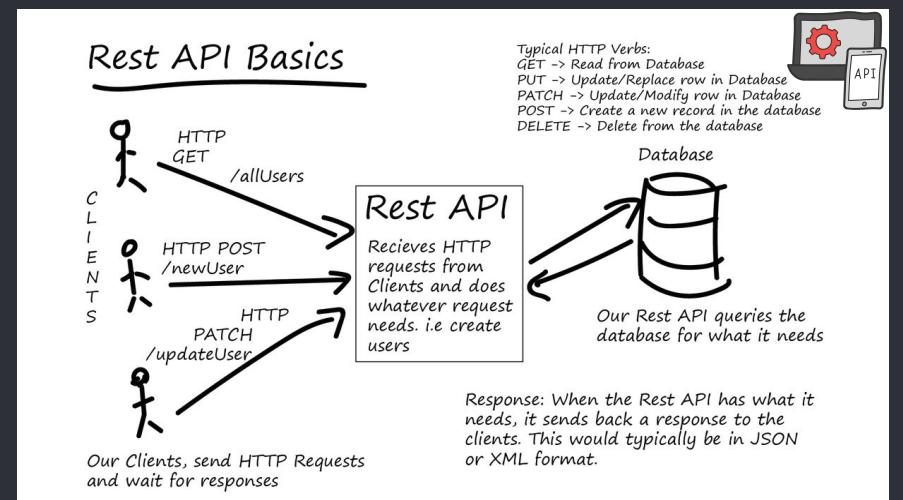


- A = Accessible

accessible does not imply open

data & metadata need to be retrievable by their identifier using a standardized communications protocol

research repositories often use the OAI-PMH or REST API protocols to interface with data in the repository

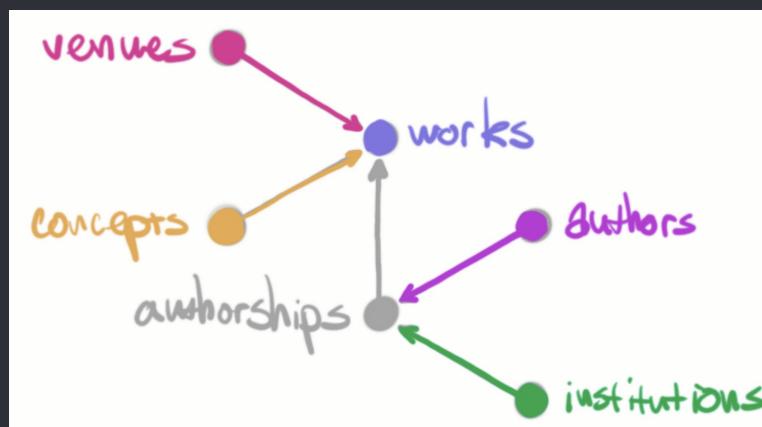


- OpenAlex: an open catalog of scholarly output



OpenAlex
documentation

OpenAlex is a [fully open catalog](#) of the global research system. It's named after the [ancient Library of Alexandria](#) and made by the nonprofit [OurResearch](#).

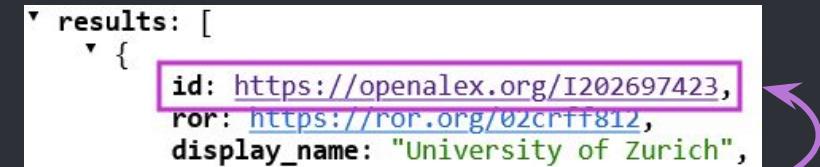
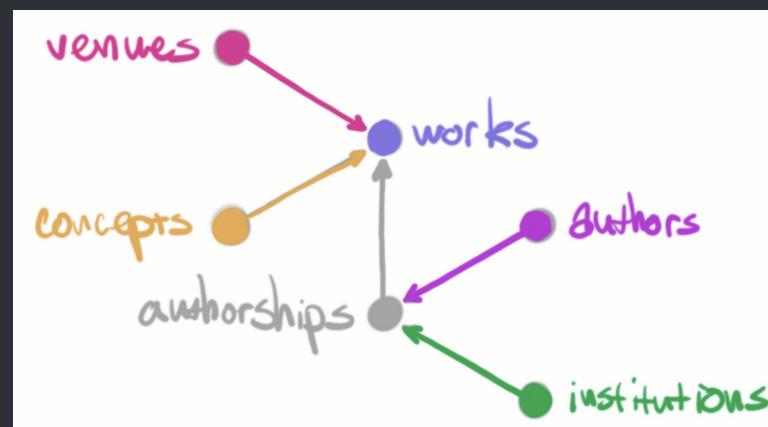


- OpenAlex: an open catalog of scholarly output



OpenAlex
documentation

OpenAlex is a [fully open catalog](#) of the global research system. It's named after the [ancient Library of Alexandria](#) and made by the nonprofit [OurResearch](#).



- https://api.openalex.org/institutions?filter=display_name.search:university%20of%20zurich
- [https://api.openalex.org/works?filter=institutions.id:https://openalex.org/I202697423,is_paratext:false,type:journal-article,from_publication_date:2012-07-14&group_by=is_oa](https://api.openalex.org/works?filter=institutions.id:https://openalex.org/I202697423,is_paratext:false,type:journal-article,from_publication_date:2012-07-14)
- https://api.openalex.org/works?filter=institutions.id:https://openalex.org/I202697423,is_paratext:false,type:journal-article,from_publication_date:2012-07-14&group_by=is_oa



- I = Interoperable

data & metadata need to be interoperable: they need to be combined with other data & tools



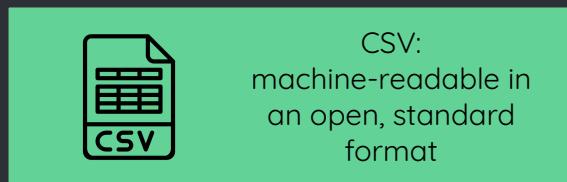
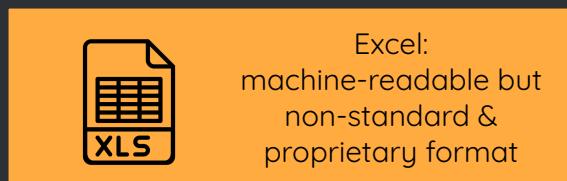
data & metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation



- I = Interoperable

data & metadata need to be interoperable: they need to be **combined with other data & tools**

data & metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation



Interoperable



- human-readable data

Data in a format that can be conveniently read by a human. Some human-readable formats, such as **PDF**, are not machine-readable as they are not structured data, i.e., the representation of the data on disk does not represent the actual relationships present in the data.



- machine-readable data

Data in a data format
that can be
automatically read and
processed by a
computer, such as [CSV](#),
[RDF](#), [JSON](#), [XML](#), etc.

Machine-readable data
must be structured
data.

- machine-readable data

Data in a data format
that can be
automatically read and
processed by a
computer, such as **CSV**,
RDF, **JSON**, **XML**, etc.

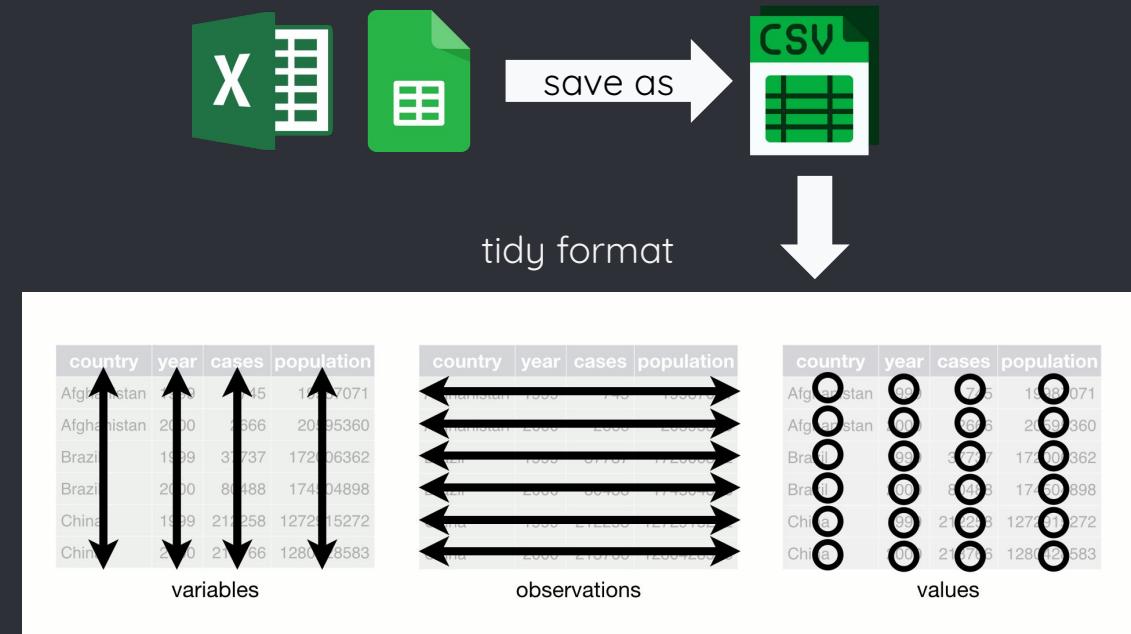
Machine-readable data
must be structured
data.



- machine-readable data

Data in a data format that can be automatically read and processed by a computer, such as **CSV**, **RDF**, **JSON**, **XML**, etc.

Machine-readable data must be structured data.

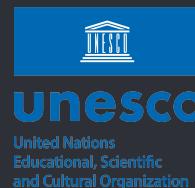


- controlled vocabularies for FAIR metadata

A controlled vocabulary reflects agreement on terminology used to label concepts. When research communities agree to use common language for the concepts in datasets, then the discovery, linking, understanding and reuse of research (data) are improved.

- controlled vocabularies for FAIR metadata

A controlled vocabulary reflects agreement on terminology used to label concepts. When research communities agree to use common language for the concepts in datasets, then the discovery, linking, understanding and reuse of research (data) are improved.



United Nations
Educational, Scientific
and Cultural Organization

TITLE	UNESCO Thesaurus
DESCRIPTION	The UNESCO Thesaurus is a controlled and structured list of terms used in subject analysis and retrieval of documents and publications in the fields of education, culture, natural sciences, social and human sciences, communication and information. Continuously enriched and updated, its multidisciplinary terminology reflects the evolution of UNESCO's programmes and activities.
IDENTIFIER	http://vocabularies.unesco.org/thesaurus



NASA Thesaurus

[Cite the NASA Thesaurus](#) | [Access the NASA Thesaurus](#)

The NASA Thesaurus contains the authorized NASA subject terms used to index and retrieve materials in the [STI Repository](#). The scope of this controlled vocabulary includes not only aerospace engineering, but all supporting areas of engineering and physics, the natural space sciences (astronomy, astrophysics, and planetary science), Earth sciences, and the biological sciences. The NASA Thesaurus contains over 18,400 subject terms, 4,300 definitions, and more than 4,500 USE cross references.

FAIRsharing.org
standards, databases, policies

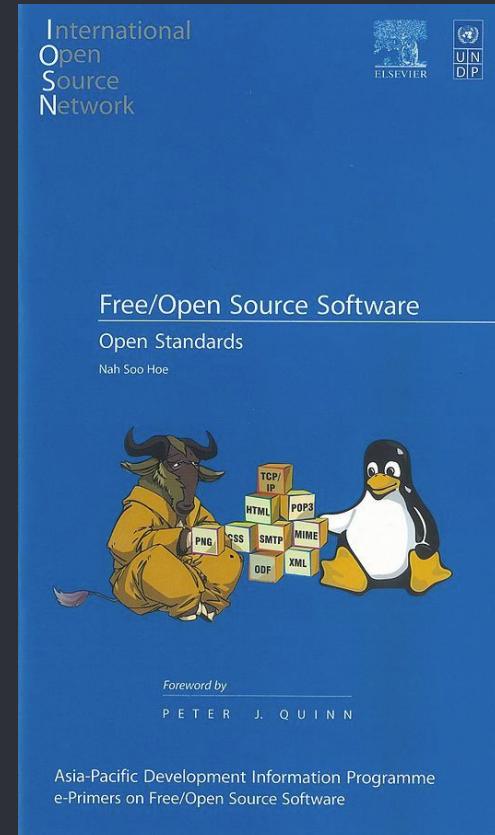
- open file formats

Go to the Wikipedia page of the [List of open file formats](#); is there a format you don't know?



Browse the wikibook [FOSS Open Standards/Comparison of File Formats](#), choose a category and have a look at the differences between the formats reported.

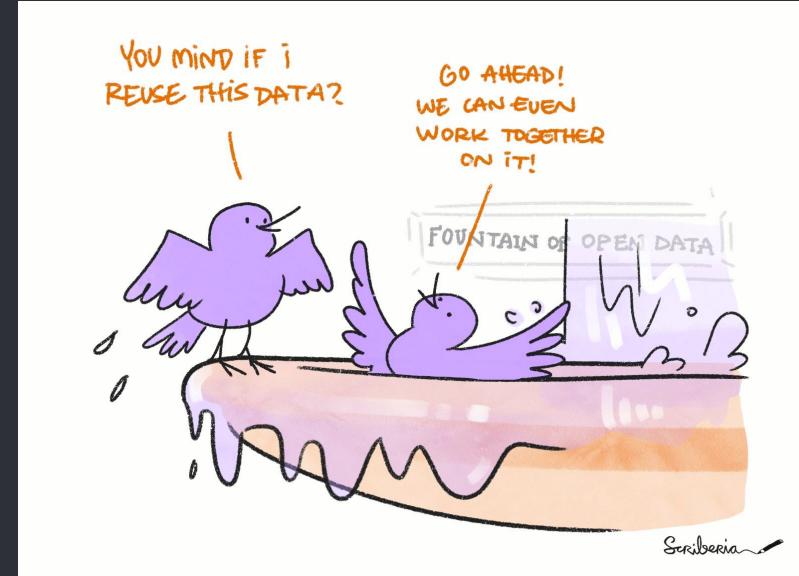
Can you pick up one format that you are willing to try from now on to share your research data/results?



- R = Reusable

data & metadata need to be well-described so that they can be replicated and/or combined in different settings

data & metadata need to be accompanied by clear, open, understandable licenses



- licenses conformant to the open data definition



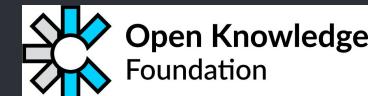
<u>License (SPDX IDs)</u>	Domain	By	SA	Comments
Creative Commons CCZero (CC0-1.0)	Content, Data	N	N	Dedicate to the Public Domain (all rights waived)
Open Data Commons Public Domain Dedication and Licence (PDDL-1.0)	Data	N	N	Dedicate to the Public Domain (all rights waived)
Creative Commons Attribution 4.0 (CC-BY-4.0)	Content, Data	Y	N	
Open Data Commons Attribution License (ODC-BY-1.0)	Data	Y	N	Attribution for data(bases)
Creative Commons Attribution Share-Alike 4.0 (CC-BY-SA-4.0)	Content, Data	Y	Y	
Open Data Commons Open Database License (ODbL-1.0)	Data	Y	Y	Attribution-ShareAlike for data(bases)

- licenses conformant to the open data definition



<u>License (SPDX IDs)</u>	Domain	By	SA	Comments
Creative Commons CCZero (CC0-1.0)	Content, Data	N	N	Dedicate to the Public Domain (all rights waived)
Open Data Commons Public Domain Dedication and Licence (PDDL-1.0)	Data	N	N	Dedicate to the Public Domain (all rights waived)
Creative Commons Attribution 4.0 (CC-BY-4.0)	Content, Data	Y	N	Creator must be credited
Open Data Commons Attribution License (ODC-BY-1.0)	Data	Y	N	Attribution for data(bases)
Creative Commons Attribution Share-Alike 4.0 (CC-BY-SA-4.0)	Content, Data	Y	Y	Derivatives or redistributions must have identical license
Open Data Commons Open Database License (ODbL-1.0)	Data	Y	Y	Attribution-ShareAlike for data(bases)

- licenses conformant to the open data definition



License (SPDX IDs)	Domain	By	SA	Comments
Creative Commons CCZero (CC0-1.0)	Content, Data	N	N	Dedicate to the Public Domain (all rights waived)
Open Data Commons Public Domain Dedication and Licence (PDDL-1.0)	Data	N	N	Dedicate to the Public Domain (all rights waived)
Creative Commons Attribution 4.0 (CC-BY-4.0)	Content, Data	Y	N	Creator must be credited
Open Data Commons Attribution License (ODC-BY-1.0)	Data	Y	N	Attribution for data(bases)
Creative Commons Attribution Share-Alike 4.0 (CC-BY-SA-4.0)	Content, Data	Y	Y	Derivatives or redistributions must have identical license
Open Data Commons Open Database License (ODbL-1.0)	Data	Y	Y	Attribution-ShareAlike for data(bases)



- licenses conformant to the open data definition

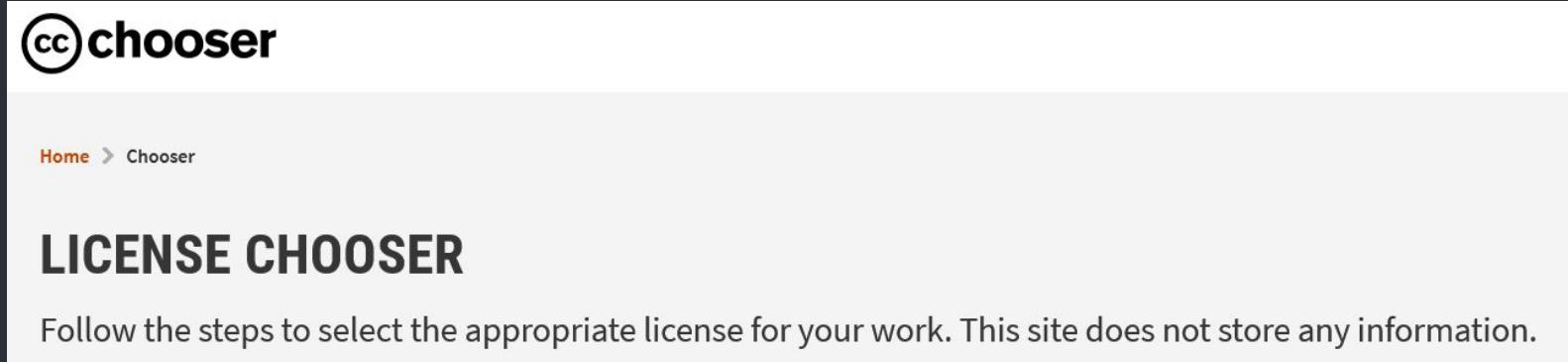


License (SPDX IDs)	Domain	By	SA	Comments
Creative Commons CCZero (CC0-1.0)	Content, Data	N	N	Dedicate to the Public Domain (all rights waived)
Open Data Commons Public Domain Dedication and Licence (PDDL-1.0)	Data	N	N	Dedicate to the Public Domain (all rights waived)
Creative Commons Attribution 4.0 (CC-BY-4.0)	Content, Data	Y	N	Creator must be credited
Open Data Commons Attribution License (ODC-BY-1.0)	Data	Y	N	Attribution for data(bases)
Creative Commons Attribution Share-Alike 4.0 (CC-BY-SA-4.0)	Content, Data	Y	Y	Derivatives or redistributions must have identical license
Open Data Commons Open Database License (ODbL-1.0)	Data	Y	Y	Attribution-ShareAlike for data(bases)



Open Source Initiative

- choose a license



The screenshot shows the homepage of the CC License Chooser. At the top left is the logo 'cc chooser'. Below it is a navigation bar with 'Home > Chooser'. The main title 'LICENSE CHOOSEN' is in large, bold, dark letters. Below the title is a subtitle: 'Follow the steps to select the appropriate license for your work. This site does not store any information.'



Go to the new [License Chooser beta, Creative Commons](#) and go through the steps to select an appropriate license for your work.

- another hunt on re3data

 **Paola Chiara Masuzzo**  @pcmasuzzo · Jun 26

Twitter folks! I am giving a workshop on FAIR data at the #OpenScience Summer School [@UZH_en](#). I obviously have ideas & material I am preparing, but would very much like to ask you: what would you definitely want to see in a workshop on #FairData? Which type of exercises? >

11 31 40

sciformation @sciformation

Replying to [@pcmasuzzo](#) and [@UZH_en](#)

UZH is using Sciformation ELN, which allows to publish data (experiments & spectra) at [sciflection.com](#) under #FAIR conditions. People can even import interesting experiments they find there into their own lab books, [sciflection.com/b27fe8f6-a68d-...](#) is called in background

2:31 PM · Jun 29, 2022 · Twitter Web App



Visit [re3data](#) and search for Sciflection:
<https://www.re3data.org/search?query=sciflection>; check the use terms and the data license associated with the repository.

Do you use this service?

Let's have a look at some other repositories you might find interesting! Feel free to annotate them in the collaborative etherpad!

time for questions



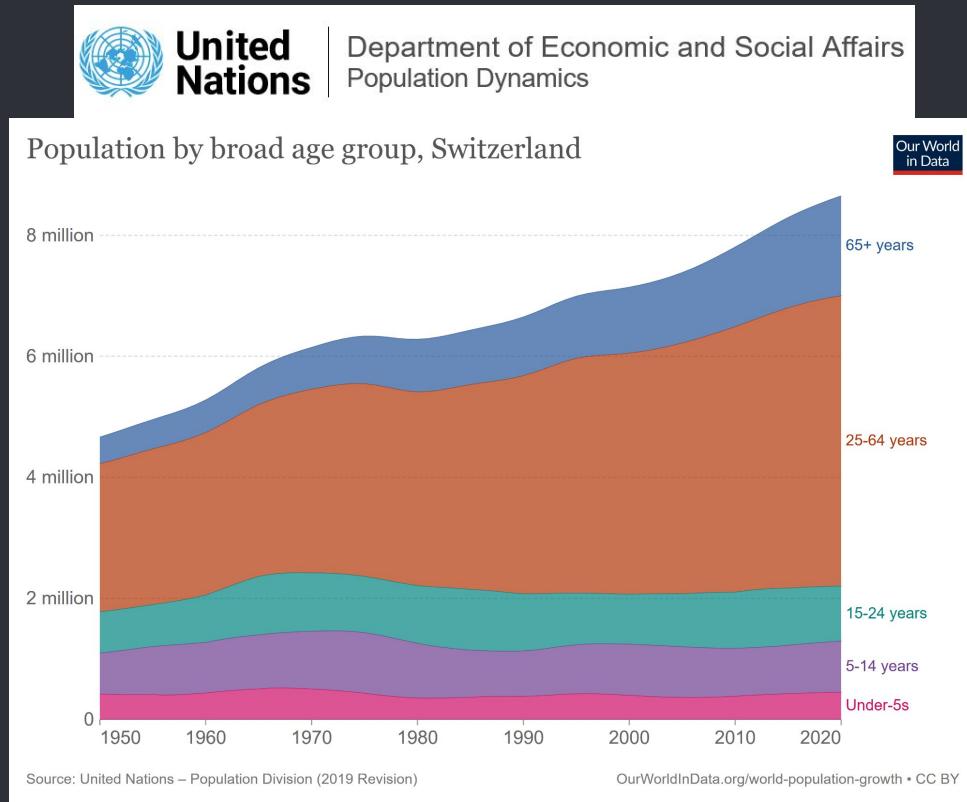
time for a coffee break



the link to the collaborative pad is:
https://pad.riseup.net/p/fair_data_workshop

Session II: FAIRify your (research) data

- the dataset we'll be working with



[Link of the graph reported in this slide: population by broad age group in Switzerland](#)

Download the [Population by Age Groups - Both Sexes \(XLSX, 10.32 MB\)](#) dataset.



What do you think about the way this data is represented/stored?

Let's use the collaborative pad to share some thoughts.



FAIR comes in many different shades





- OpenRefine: working with messy data



Launch OpenRefine (<http://127.0.0.1:3333>) and import the Excel file you downloaded. Make sure to skip the first 16 rows of the file, and to select only the first sheet of data.

Run a *customized facet by blank* (null or empty string) on the ‘Reference date’ column: do you see missing values? How are *missing values* treated in the other (numerical) columns?

choice	count
false	3825



● OpenRefine: working with messy data

Filter on the ‘Region, subregion, country or area *’ column and retain only Switzerland.

Then, convert the wide format into a long one using the Transpose function. Why is this a better format for data representation?



Finally, click on Export, use the Custom tabular exporter and export the dataset as a comma-separated values file (choose the columns you want to export).

Transpose Cells Across Columns into Rows

From Column	To Column	Transpose into
Index	10-14	<input checked="" type="radio"/> Two new columns
Variant	15-19	Key Column <input type="text" value="Age Group"/> (containing original columns' names)
Region, subregion, country or area *	20-24	Value Column <input type="text" value="Population"/> (containing original cells' values)
Notes	25-29	
Country code	30-34	
Type	35-39	
Parent code	40-44	
Reference date (as of 1 July)	45-49	
0-4	50-54	
5-9	55-59	
10-14	60-64	
15-19	65-69	
20-24	70-74	
25-29	75-79	
30-34	80-84	
35-39	85-89	
40-44	90-94	
45-49	95-99	
50-54	100+	
55-59	(last column)	

prepend the original column's name to each cell followed by before the cell's value

Ignore blank cells

Fill down in other columns

Transpose **Cancel**

- let's build a data package



Create a folder on your computer to store the data file we just created.



Tip: make a folder for this project, with a dedicated data folder within.

- let's build a data package



Create a folder on your computer to store the data file we just created.



Tip: make a folder for this project, with a dedicated data folder within.

A) Organized by File type

Dataset.A

```
| - Code  
|   | - Step.1  
|   | - Step.2  
| - Data  
|   | - Processed  
|   | - Raw  
| - Results/  
|   | - Figure.1  
|   | - Figure.2  
|   | - Models  
| readme.txt
```

B) Organized by Analysis

Dataset.B

```
| - Figure.1  
|   | - Code  
|   | - Data  
|   | - Results  
| - Figure.2  
|   | - Code  
|   | - Data  
|   | - Results  
| - Table.1  
|   | - Code  
|   | - Data  
|   | - Results  
| readme.txt
```

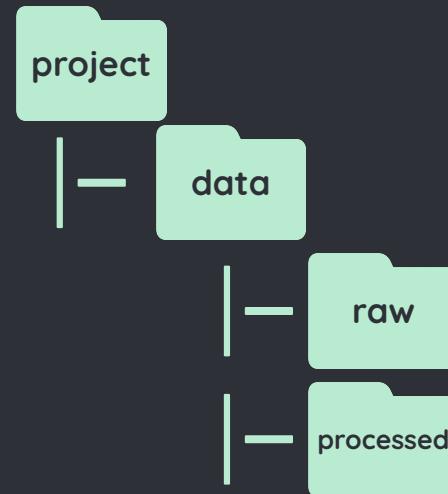
- let's build a data package



Create a folder on your computer to store the data file we just created.



Tip: make a folder for this project, with a dedicated data folder within.

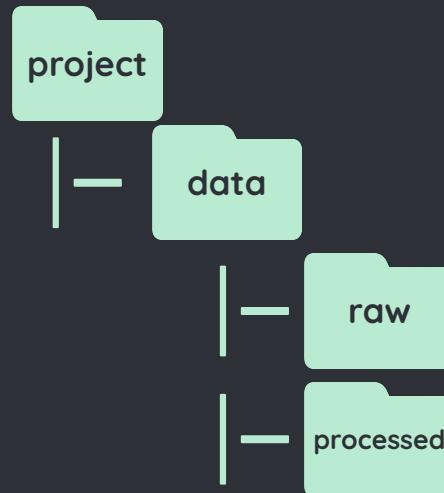


- let's build a data package



Create a folder on your computer to store the data file we just created.

Tip: make a folder for this project, with a dedicated data folder within.



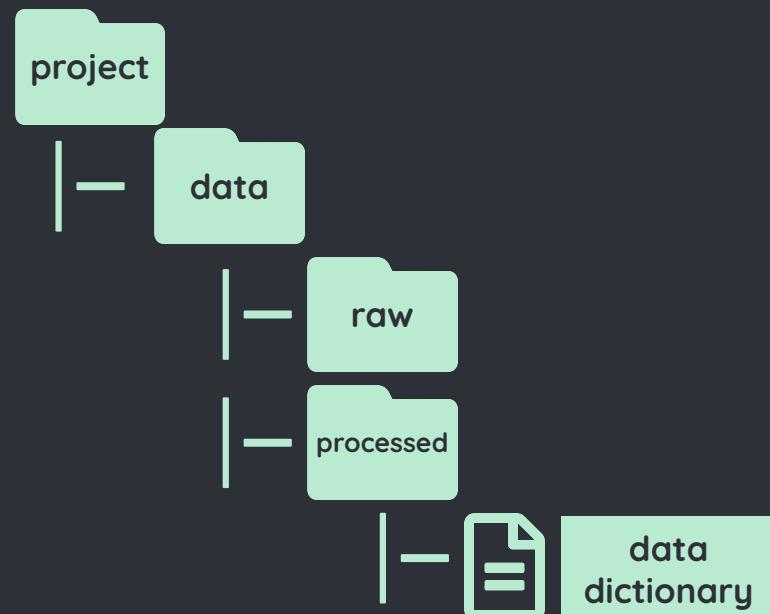
What else do we add to the project folder for people to be able to understand what this dataset is about and use it in different contexts?

- let's write a data dictionary



A **data dictionary** is a text file that explains what all the variable names and values in your csv/spreadsheet file really mean.

Ideally, we write a data dictionary for each data file in our dataset. Let's write one for the processed data.



- let's write a data dictionary



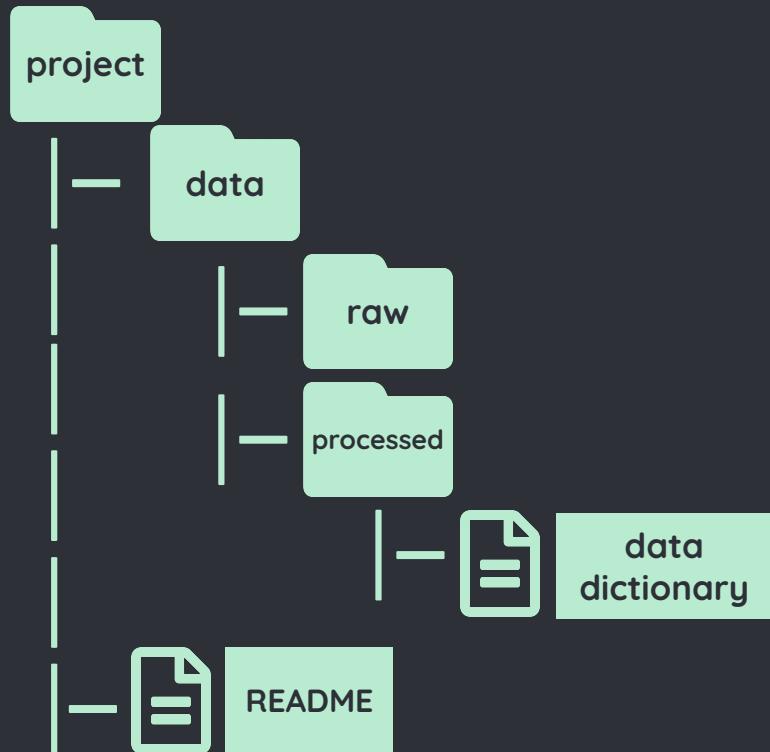
```
"Region, subregion, country or area *",Country code,Reference date (as of 1 July),Age Group,Population  
Switzerland,756,1950,0-4,421.699  
Switzerland,756,1950,5-9,368.699  
Switzerland,756,1950,10-14,307.301  
Switzerland,756,1950,15-19,333.501  
Switzerland,756,1950,20-24,348.5
```

Variable, Variable Name, Measurement Unit, Allowed Values, Description
"Region, subregion, country or area *", Geographical Area, String, Switzerland, country selected
Country code, ISO code for country, Numeric, https://en.wikipedia.org/wiki/List_of_ISO_3166_country_codes, ISO 3166-1 numeric standard
Reference date (as of 1 July), Year, Numeric, 1950-2020, Time period in 5 years bucket
Age Group, Age Group, String, from 0-5 to 100+, Broad group of age in buckets of 5 years
Population, Population, Numeric (figures in thousands), -, "De facto population in a country, area or region as of 1 July of the year indicated. Figures are presented in thousands"

- let's add a README file

A **README** file is a bit like the identity card of your dataset, and as such should contain:

- name/institution/address of the person responsible for collecting the data
- contact person for questions
- date and geographic location of data collection
- language information
- Information about funding sources that supported the collection of the data
- anything else that might help to understand the data and its context



- let's add a README file



Go to

<https://www.makeareadme.com/>

and write a simple README file for your project in markdown (save it as README.md).

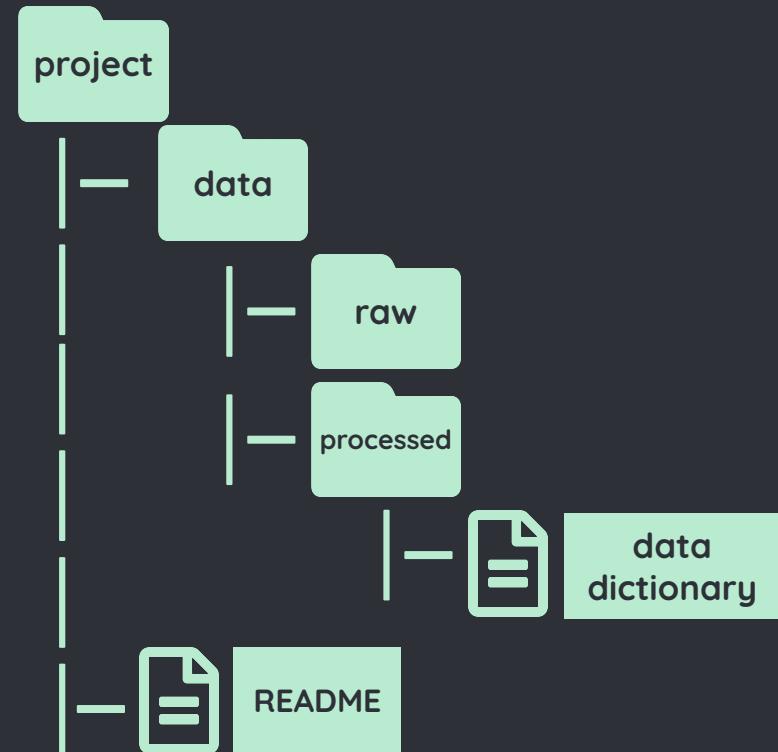
This tool <https://stackedit.io/app#> can help you with the formatting.

This project contains the following directories:

1. data: directory with raw and processed data
2. README file: this file.

To-do:

Add a license file



- let's add a license file



Go to <https://population.un.org/wpp/Download/Standard/Population/> and click on the license information, let's see what it says.

Copyright © 2019 by United Nations, made available under a Creative Commons license CC BY 3.0 IGO: <http://creativecommons.org/licenses/by/3.0/igo/>
Suggested citation: United Nations, Department of Economic and Social Affairs, Population Division (2019). *World Population Prospects 2019, Online Edition*. Rev. 1.

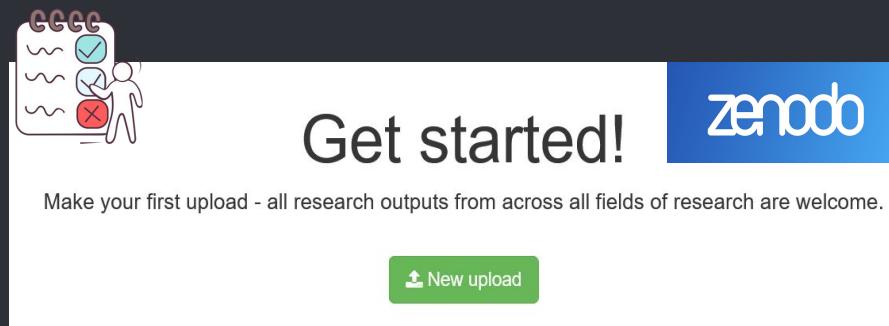


Attribution 3.0 IGO (CC BY 3.0 IGO)



Create an empty file, call it LICENSE, copy the legal code from the license content: <https://creativecommons.org/licenses/by/3.0/igo/legalcode> and paste it into this file.

- your dataset is ready to get a PID



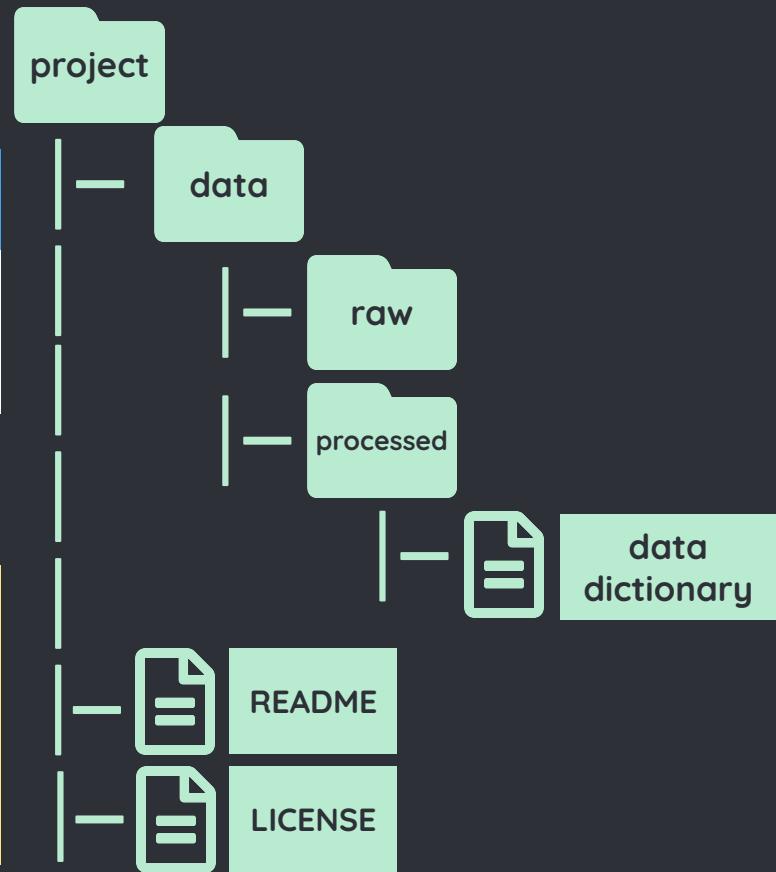
Publication date:
July 3, 2022

DOI:
DOI [10.5072/zenodo.1078378](https://doi.org/10.5072/zenodo.1078378)

License (for files):
[Creative Commons Attribution 4.0 International](#)

Cite as
Masuzzo, Paola. (2022). fairify your data [Data set].
<https://doi.org/10.5072/zenodo.1078378>

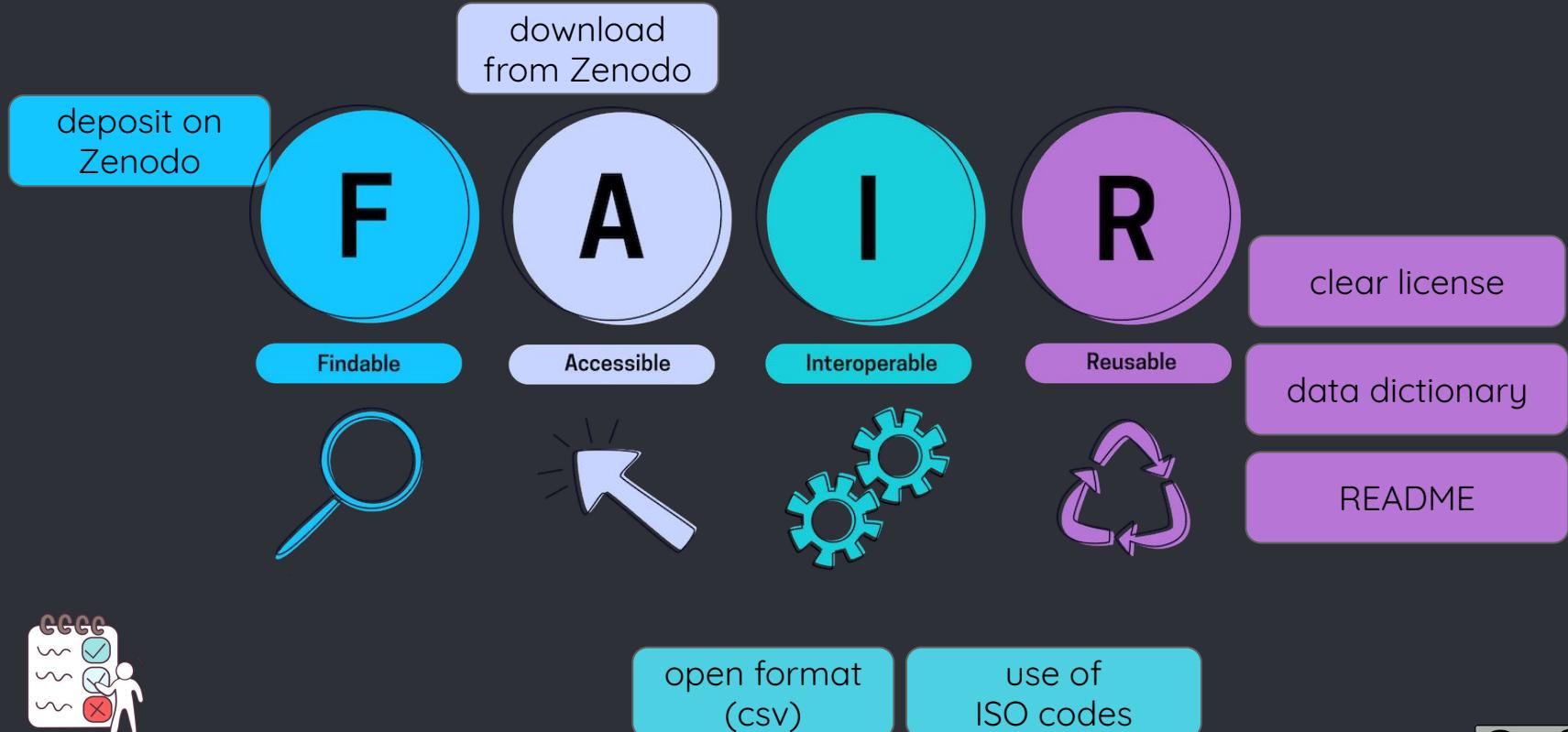
note: Zenodo sandbox is a test environment: your data will not be permanent, will be deleted after some time!!!



- linking back our steps to the FAIR principles



- linking back our steps to the FAIR principles



Bonus track: FAIR in the era of social machines

- data sharing or data visitation?

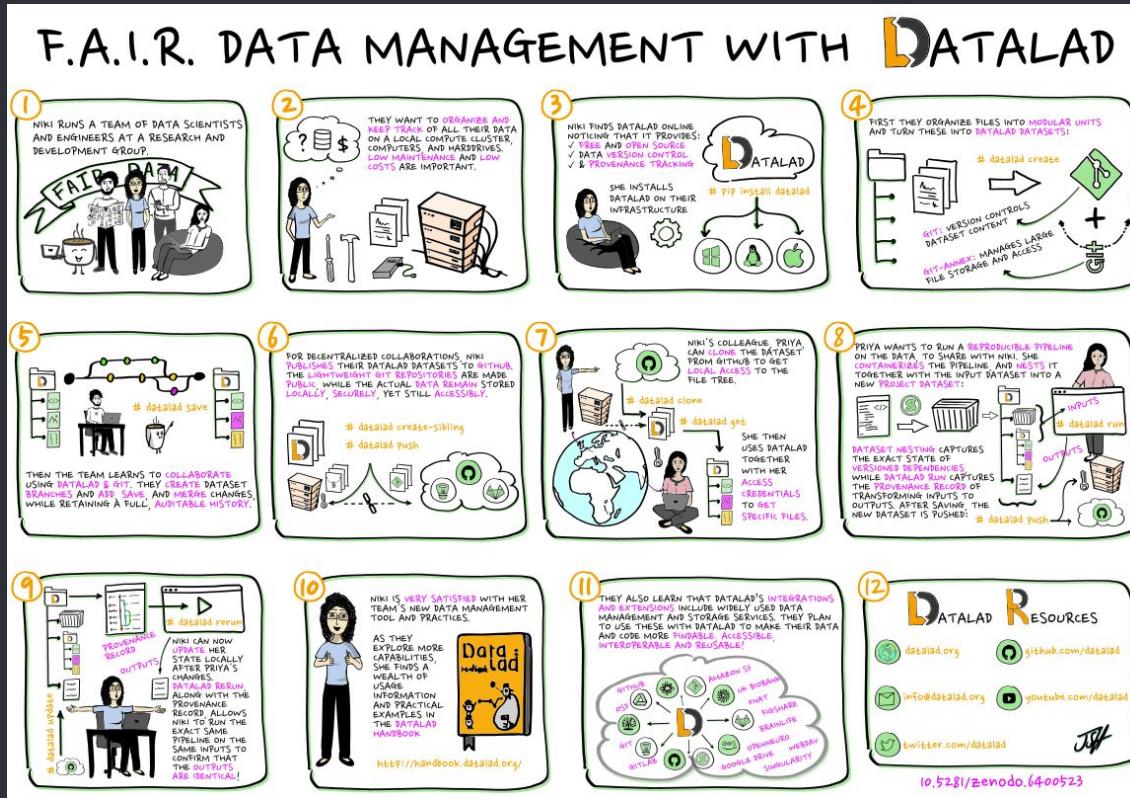
March 01 2019

FAIR Science for Social Machines: Let's Share Metadata Knowlets in the Internet of FAIR Data and Services

Barend Mons  



- DataLad for FAIR data management



- DataLad and the Open NEURO datasets



DataLad

OpenNeuro Runs on DataLad

Want to access OpenNeuro datasets with DataLad? Visit the [dataset collection on GitHub](#).

A data management solution built on [Git](#) and [Git-annex](#). Read more about [DataLad](#)

ds004187 Public

OpenNeuro dataset available at <https://openneuro.org/datasets/ds004187>

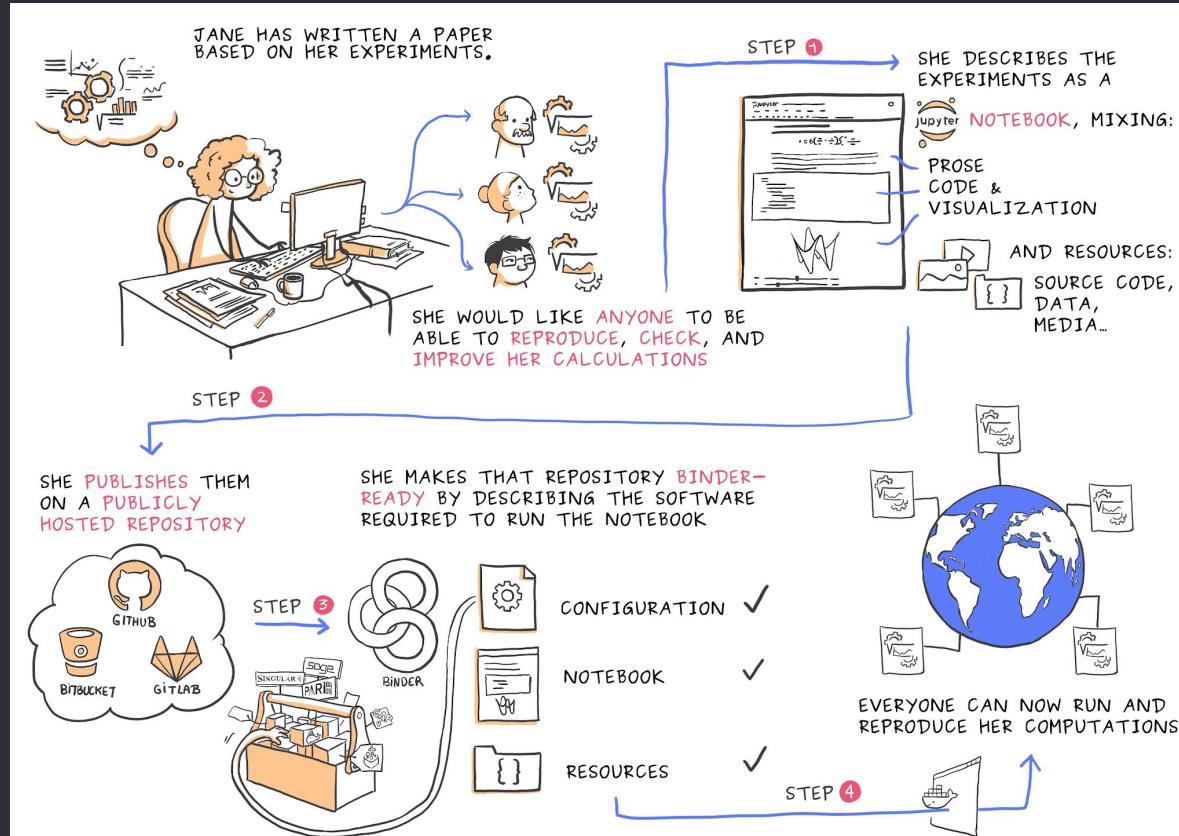
0 stars 0 forks 0 updated 13 hours ago

OpenNeuro Accession Number: ds004187 **Files:** 409 **Size:** 36.91GB

LGN Layers Data

- CHANGES
- dataset_description.json
- README
- derivatives
- sub-01
- sub-02
- sub-03

- literate programming: human-readable text & machine-readable code



- NeuroLibre: interactive and reproducible neuroscience



NeuroLibre

NeuroLibre is a preprint server for **reproducible** data analyses.

Committed to publishing curated Jupyter notebooks with zero article processing charges or subscription fees.

 Submit a preprint to NeuroLibre ›



How can we really transition to a fully open,
transparent, FAIR research culture?



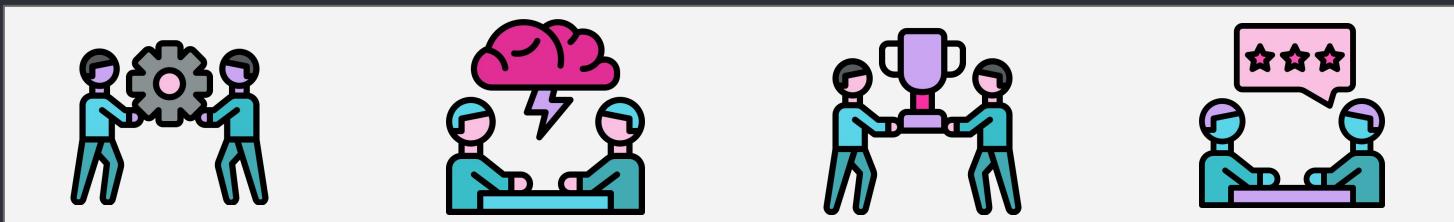
- invest in data stewardship through people

address data management needs across disciplines

assist with coding tasks and issues

teach, support, and inspire other researchers

head of research data services



data stewards and managers

software developers

(open) champions

people who believe in people

Thank you!

You can find me on Twitter with the handle [@pcmasuzzo](https://twitter.com/@pcmasuzzo) and you can always email me at:
paola.masuzzo@gmail.com

