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

You can follow the HTML version of this presentation here: http://tiny.cc/fairdatamaribor

Hi, I am Paola!
I am a data scientist, Open Science advocate and independent researcher at IGDORE

IGDORE
Institute for Globally Distributed Open Research and Education

Dealing with FAIR Data

University of Maribor
Open Science Summer School
12th September 2023

Picture

some disclaimers and practicalities

- The web is full of resources, and by no means this workshop covers them all (not even close!) - I have prepared some reading material for you at the end of this presentation

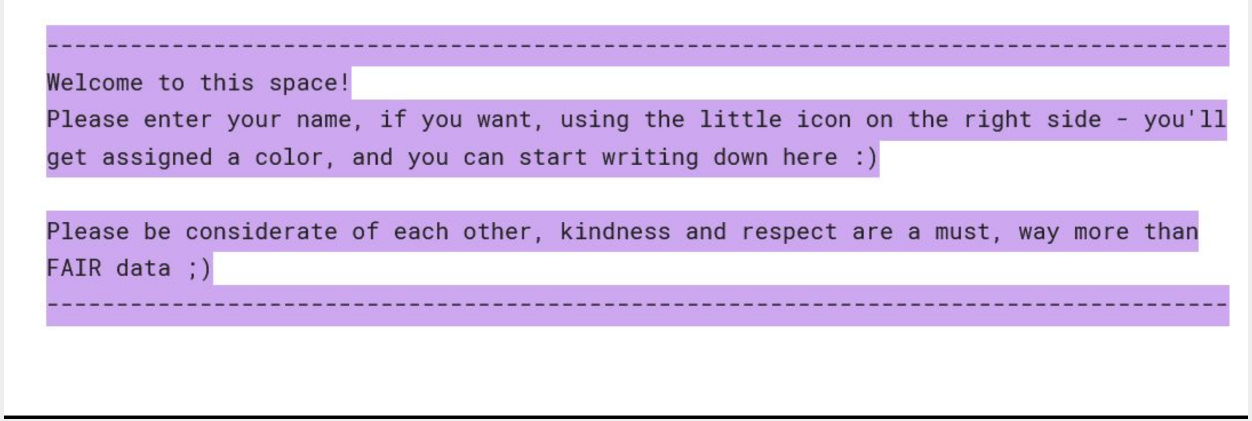
- You think you don't produce data / you don't have anything to make FAIR? You'll soon change your mind ;)

- A marker will indicate when it's time to get our hands dirty (looking up stuff online, browsing a database, using software, etc.)

we’ll use a collaborative pad

http://tiny.cc/maribor

Please use it to share comments, questions, etc.
I will copy-paste useful stuff in there, and you can then copy-paste these on your laptop :)



```
-------------------------------------------------------------------------------
Welcome to this space!
Please enter your name, if you want, using the little icon on the right side - you'll
get assigned a color, and you can start writing down here :)

Please be considerate of each other, kindness and respect are a must, way more than
FAIR data ;)
-------------------------------------------------------------------------------
```

the (rough) agenda

Introduction to research data and FAIR - 12h15-12:45

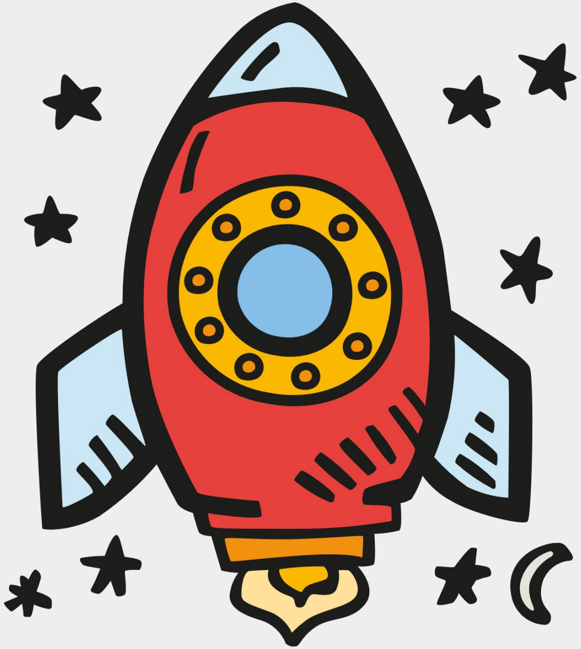
The FAIR principles in action - 13h00-14h00

F for Findable
A for Accessible
I for Interoperable
R for Reusable

Lunch 14h00-15h00

FAIRify (your) data - 15h00-16h15

LET'S GO!



the (rough) agenda

Introduction to research data and FAIR - 12h15-12:45

The FAIR principles in action - 13h00-14h00

F for Findable
A for Accessible
I for Interoperable
R for Reusable

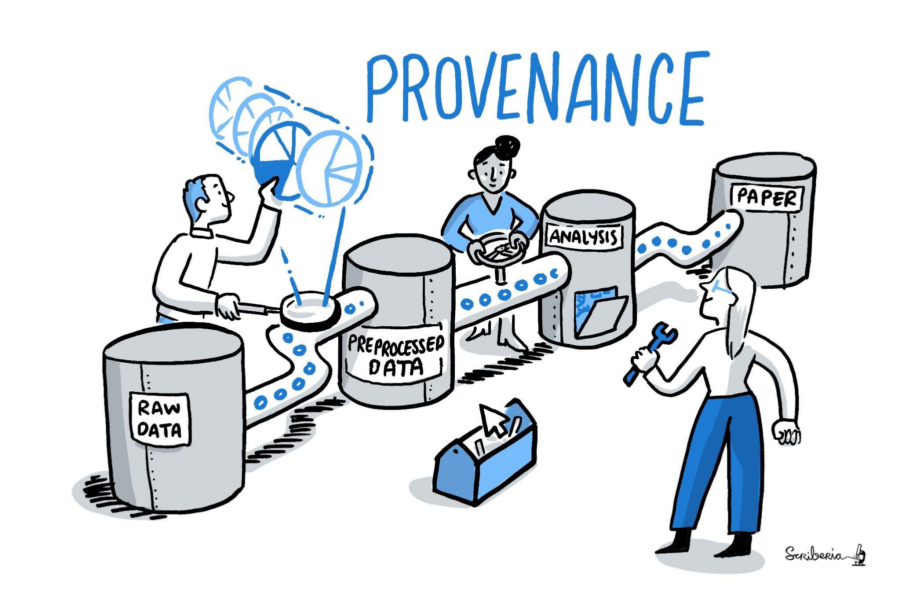
Lunch 14h00-15h00

FAIRify (your) data - 15h00-16h15

what is research data?

Research data: any type of information created, collected, observed, in the context of research

- Primary: raw data from measurements or instruments
- Secondary: processed from second-order analysis and interpretation
- Published: final format available for use and reuse
- Metadata: data about the data



https://zenodo.org/record/3695300

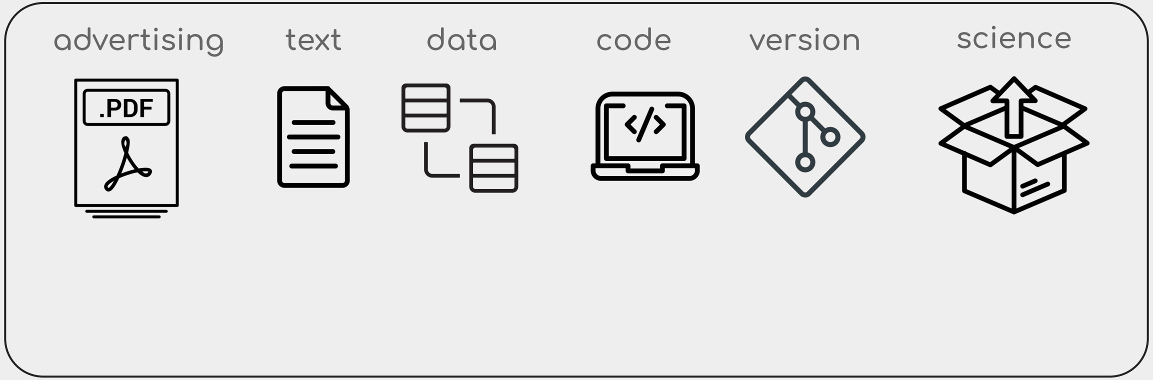
examples of research data

```
//fires the appear event when appropriate
var check = function() {

  //is the element hidden?
  if (!t.is(':visible')) {

    //it became hidden
    t.appeared = false;
    return;
  }

  //is the element inside the visible window?
  var a = w.scrollLeft();
  var b = w.scrollTop();
  var o = t.offset();
  var x = o.left;
  var y = o.top;

  var ax = settings.accX;
  var ay = settings.accY;
  var th = t.height();
  var wh = w.height();
  var tw = t.width();
  var ww = w.width();

  if (y + th + ay >= b &&
      y <= b + wh + ay &&
      x + tw + ax >= a &&
      x <= a + ww + ax) {

        //trigger the custom event
        if (!t.appeared) t.trigger('appear', settings.data);

  } else {

        //it scrolled out of view
        t.appeared = false;
  }
};

//create a modified fn with some additional logic
var modifiedFn = function() {

  //mark the element as visible
  t.appeared = true;

  //is this supposed to happen only once?
  if (settings.one) {

    //remove the check
    w.unbind('scroll', check);
    var i = $.inArray(check, $.fn.appear.checks);
    if (i >= 0) $.fn.appear.checks.splice(i, 1);
  }

  //trigger the original fn
  fn.apply(this, arguments);
};

//bind the modified fn to the element
one) t.one('appear', settings.data, modifiedFn);
settings.data, modifiedFn);
```

Accurate historical weather data
Cause of June 2015 black church fires
Publicly available gun trace data
Amount of phosphorous in
Number of mosques surveilled by FBI
Employment statistics that include
Location-specific data on
Accurate birth registration
Financial information for all

https://mimionuoha.com/the-library-of-missing-datasets; https://unsplash.com/photos/mifdr7-s4Us; https://unsplash.com/photos/pNEmlb1CMZM; https://unsplash.com/photos/1LLh8k2_YFk; https://unsplash.com/photos/Oaqk7qqNh_c the Open Science movement encourages researchers to share research output beyond the contents of a published academic article the Open Science movement encourages researchers to share research data beyond the contents of a published academic article

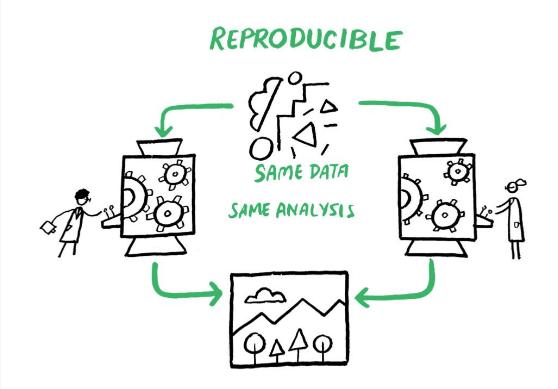
research is much more than a PDF



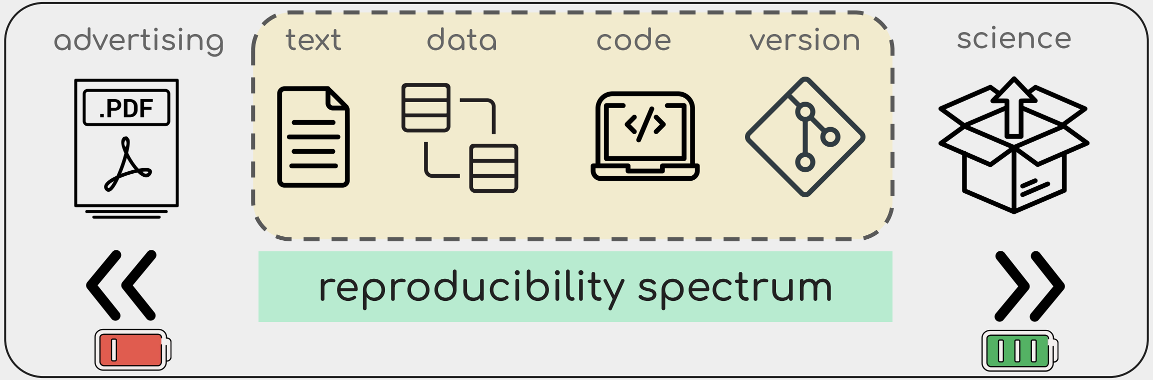
advertising text data code version science

reproducibility: minimum standard for research validity

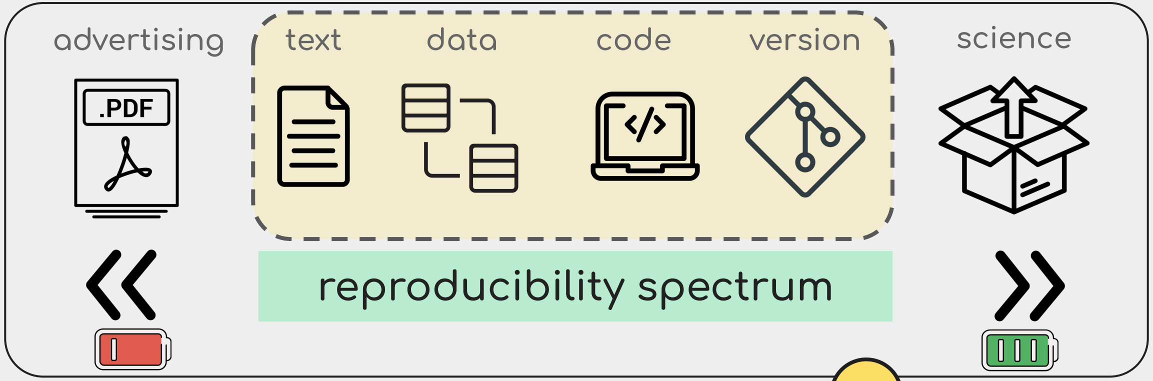
advertising text data code version science



reproducibility: minimum standard for research validity



data as substrate for knowledge discovery



advertising

.PDF

text

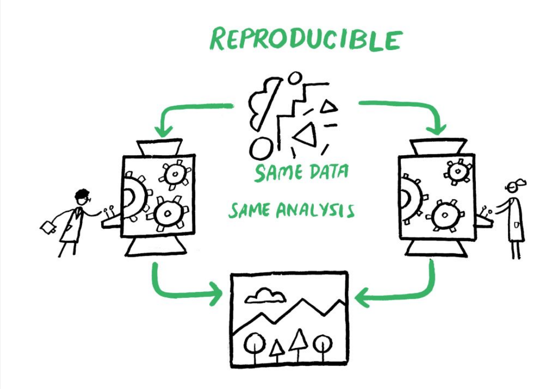
data

code

version

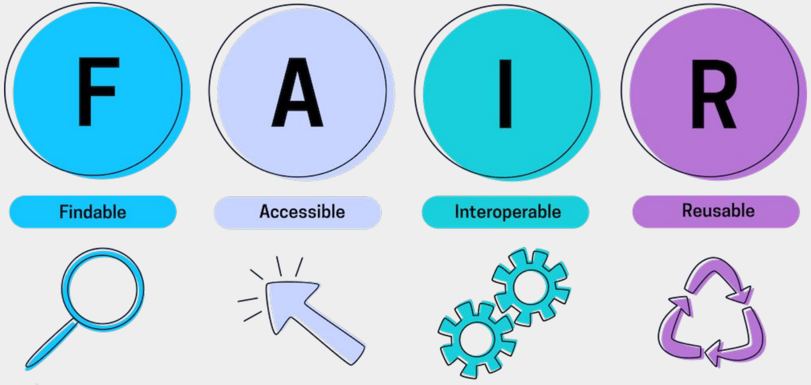
science

reproducibility spectrum



objects that belong together should be linked to each other (and other objects), so that they can be discovered on the web

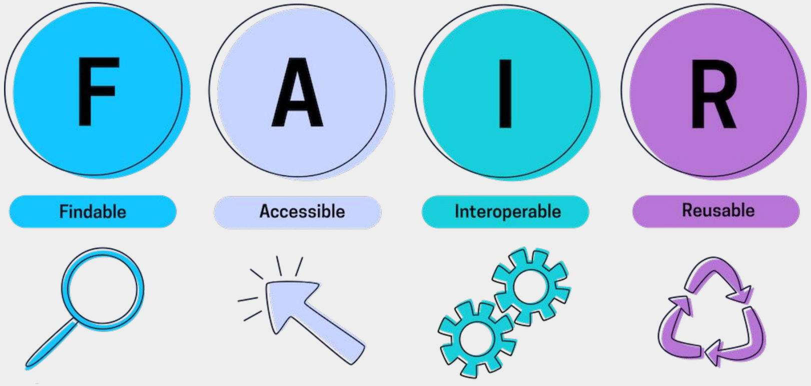
the FAIR principles as guidance for data stewardship



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

https://www.nature.com/articles/sdata201618; https://medium.com/fluree/making-data-f-a-i-r-93629e82c459

the FAIR principles as guidance for data stewardship



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

FAIR comes in degrees

FAIR is agnostic of technical implementations

FAIR requires work!

FAIR is not the same as OPEN

https://www.nature.com/articles/sdata201618; https://medium.com/fluree/making-data-f-a-i-r-93629e82c459

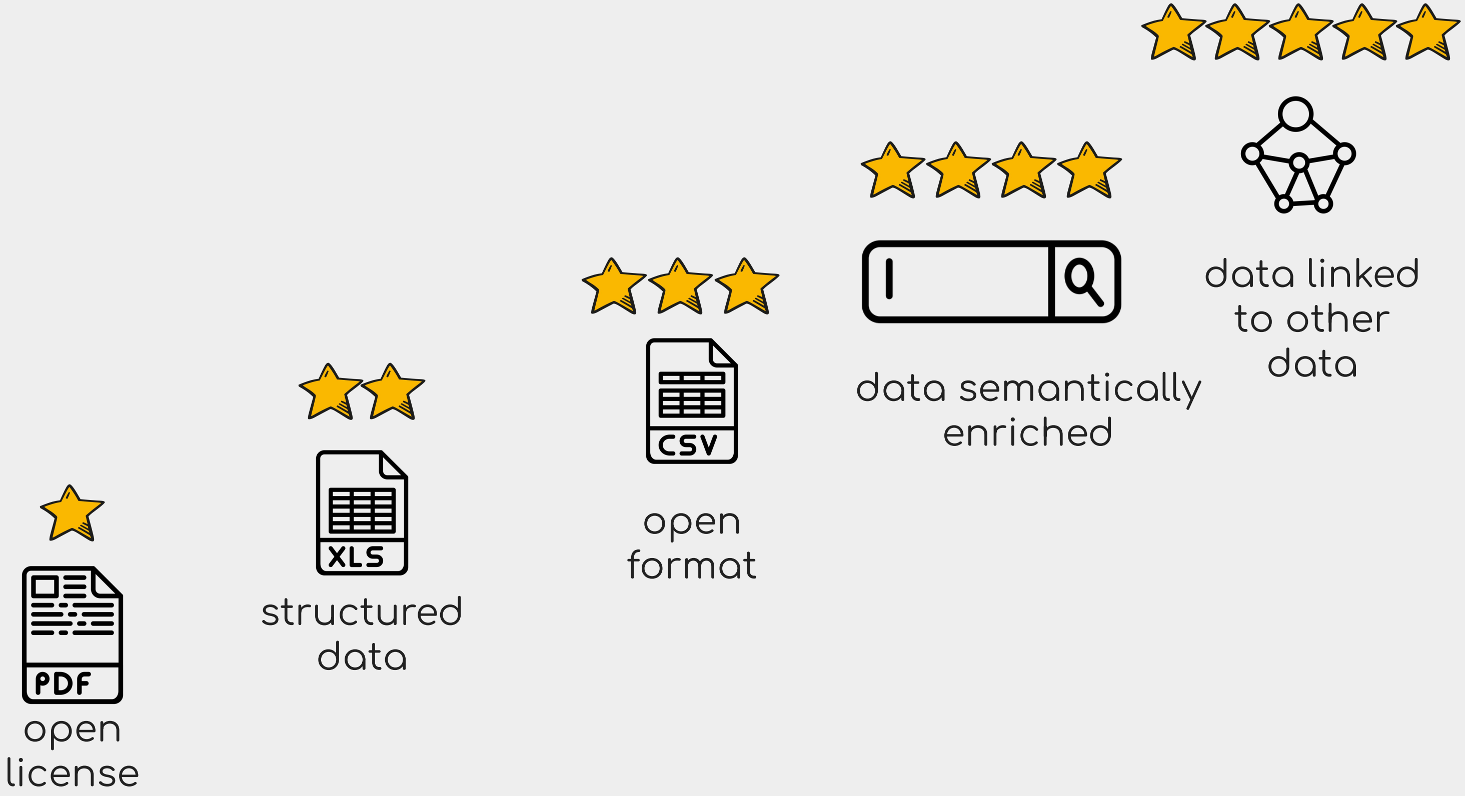
open data: a definition

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 Knowledge Foundation

https://opendatahandbook.org/guide/en/what-is-open-data/

the 5 star open data model



open license

structured data

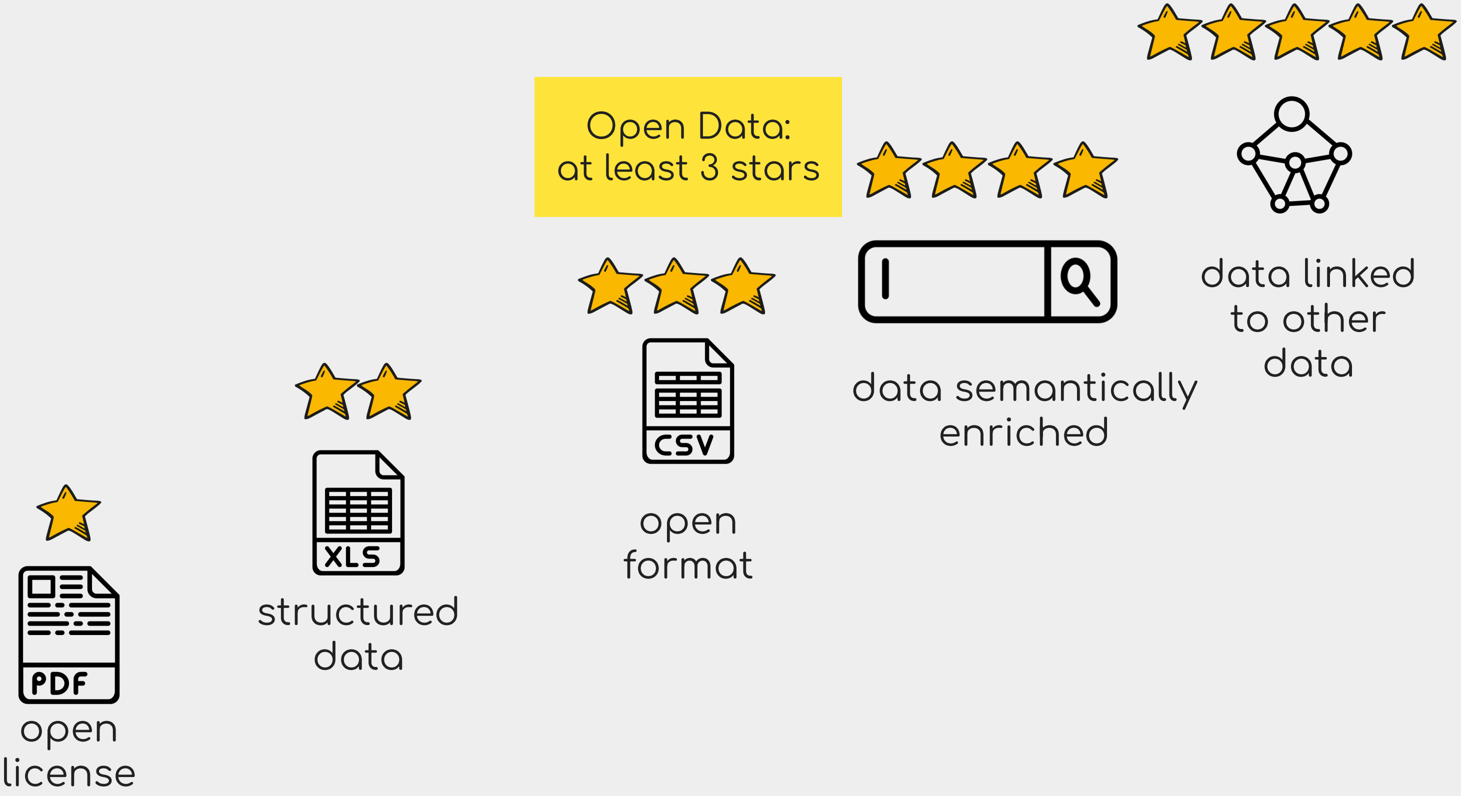
open format

data semantically enriched

data linked to other data

https://5stardata.info/en/

the 5 star open data model



Open Data:
at least 3 stars

open
license

structured
data

open
format

data semantically
enriched

data linked
to other
data

https://5stardata.info/en/

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

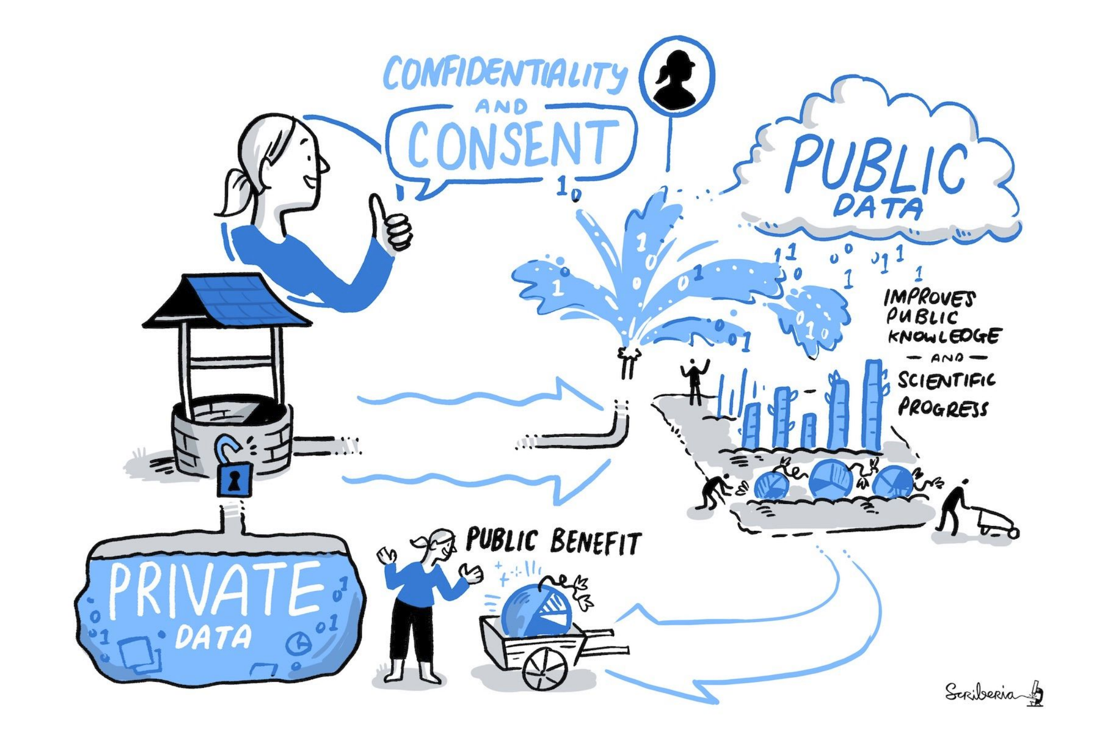
Making your data freely available on the web doesn’t translate to it being reusable

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



http://doi.org/10.5281/zenodo.3695300

open data is not FAIR data, and vice versa

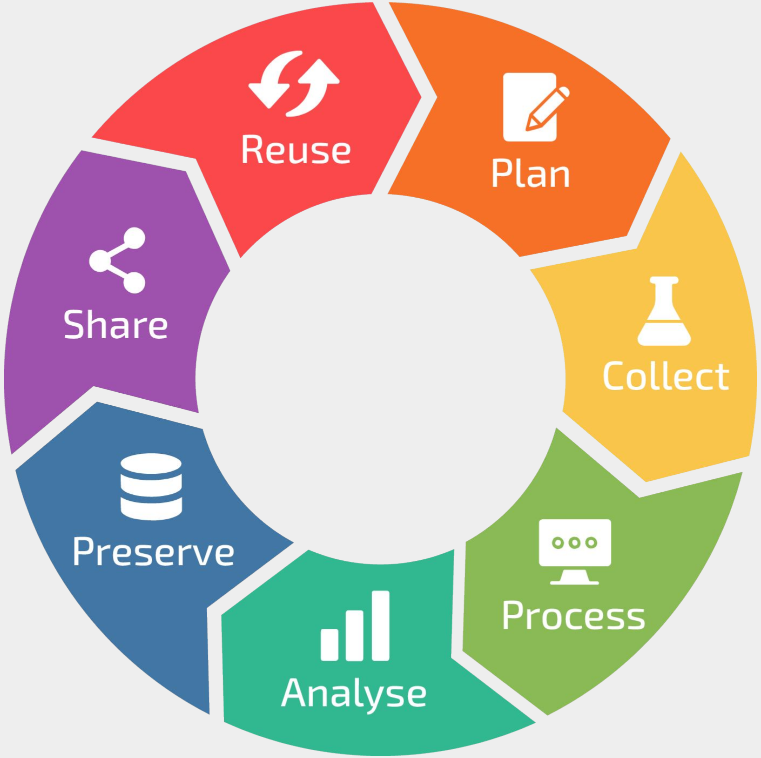
FAIR is not equivalent of OPEN, but OPEN data needs to be FAIR to be useful

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

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

Ideally, you want FAIR data shared openly!

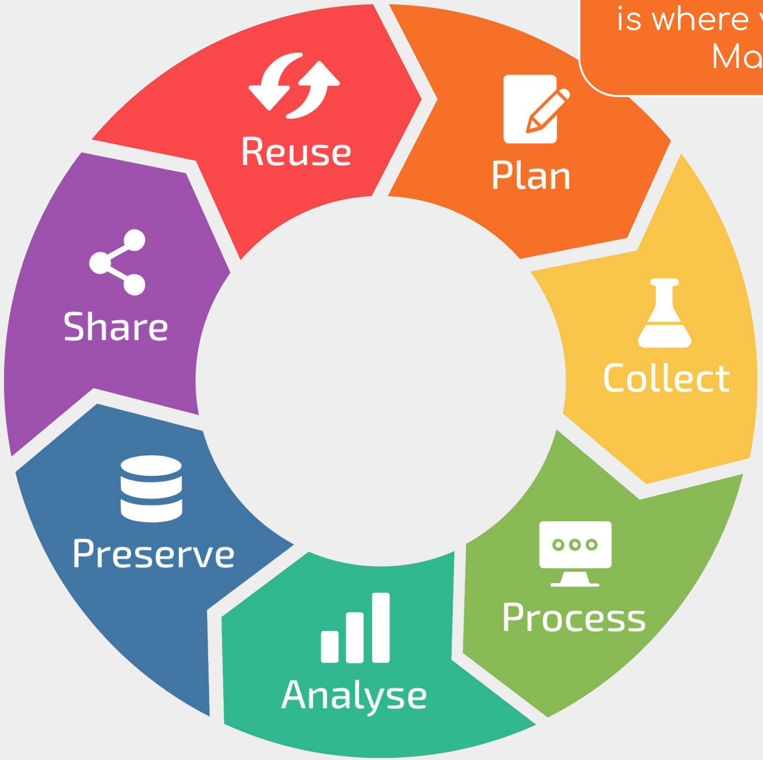
the research data life cycle



Picture

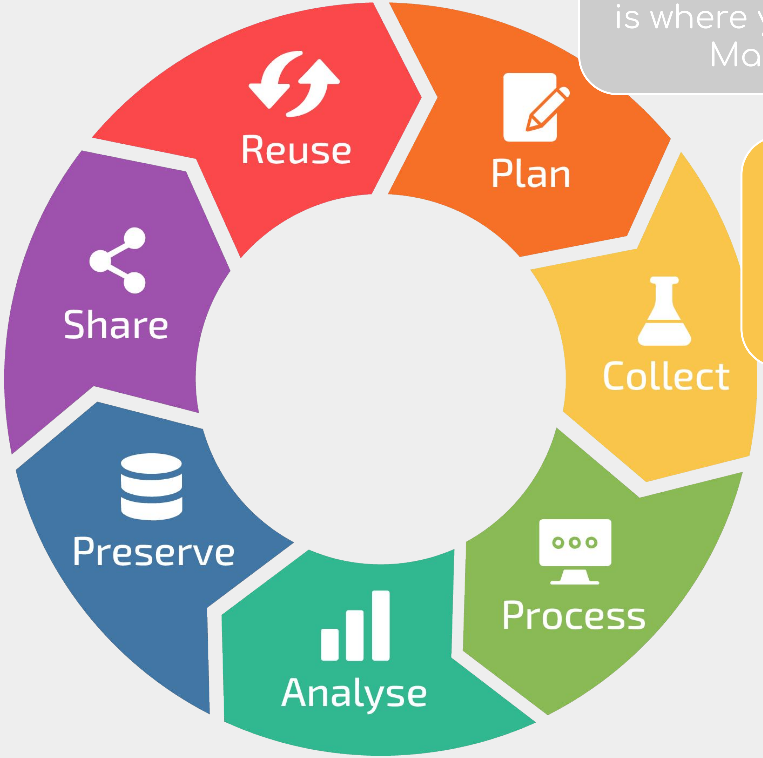
the research data life cycle

Ideally, you start thinking about FAIR data even before you start collecting them! This is where you design your Data Management Plan.



Picture

the research data life cycle

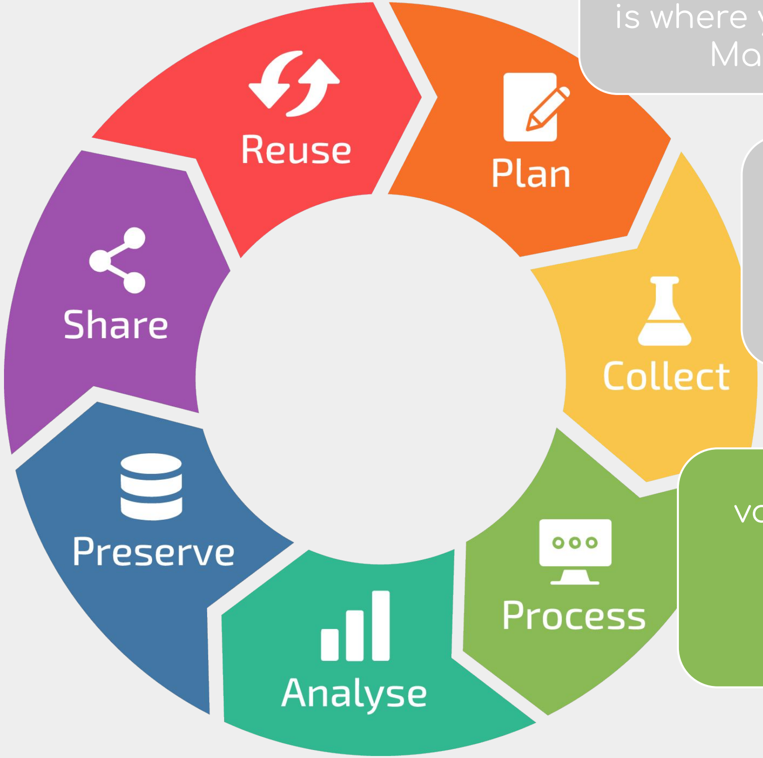


Ideally, you start thinking about FAIR data even before you start collecting them! This is where you design your Data Management Plan.

Data collection is where the data is created, either by collecting new data, or locating existing data, for example reusing someone other's data.

Picture

the research data life cycle



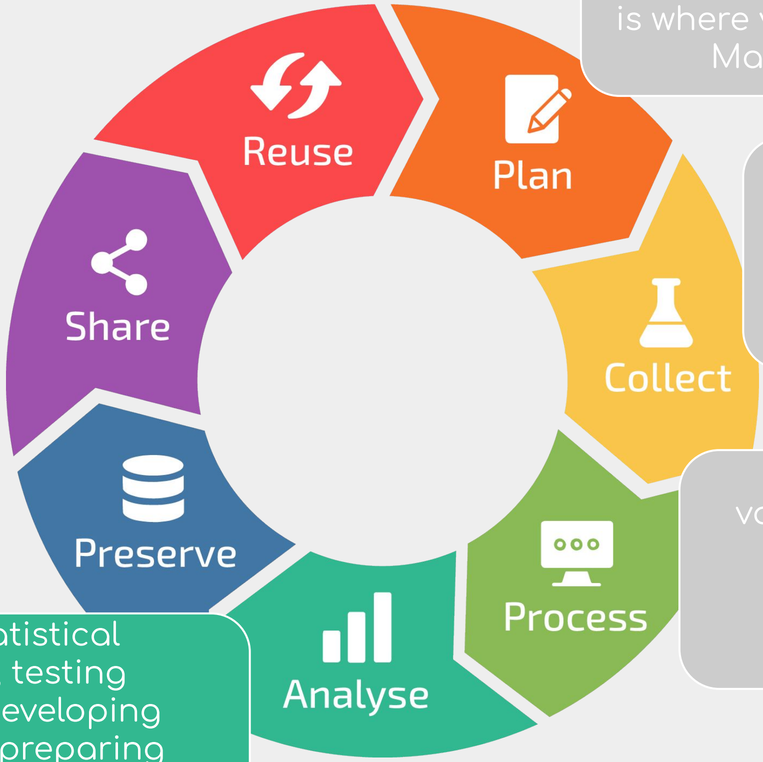
Ideally, you start thinking about FAIR data even before you start collecting them! This is where you design your Data Management Plan.

Data collection is where the data is created, either by collecting new data, or locating existing data, for example reusing someone other's data.

Data processing means validating and cleaning the data, organizing and transforming the data, anonymizing them, if necessary.

Picture

the research data life cycle



Ideally, you start thinking about FAIR data even before you start collecting them! This is where you design your Data Management Plan.

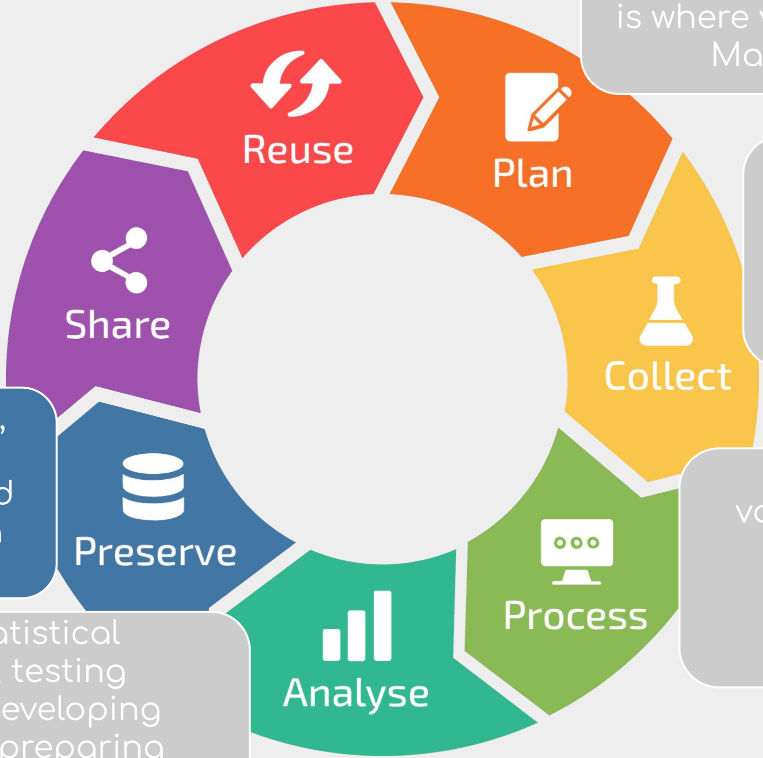
Data collection is where the data is created, either by collecting new data, or locating existing data, for example reusing someone other's data.

Data processing means validating and cleaning the data, organizing and transforming the data, anonymizing them, if necessary.

Running statistical procedures, testing hypotheses, developing explanations, preparing insights, reports, thinking about authorship.

Picture

the research data life cycle



Ideally, you start thinking about FAIR data even before you start collecting them! This is where you design your Data Management Plan.

Data collection is where the data is created, either by collecting new data, or locating existing data, for example reusing someone other's data.

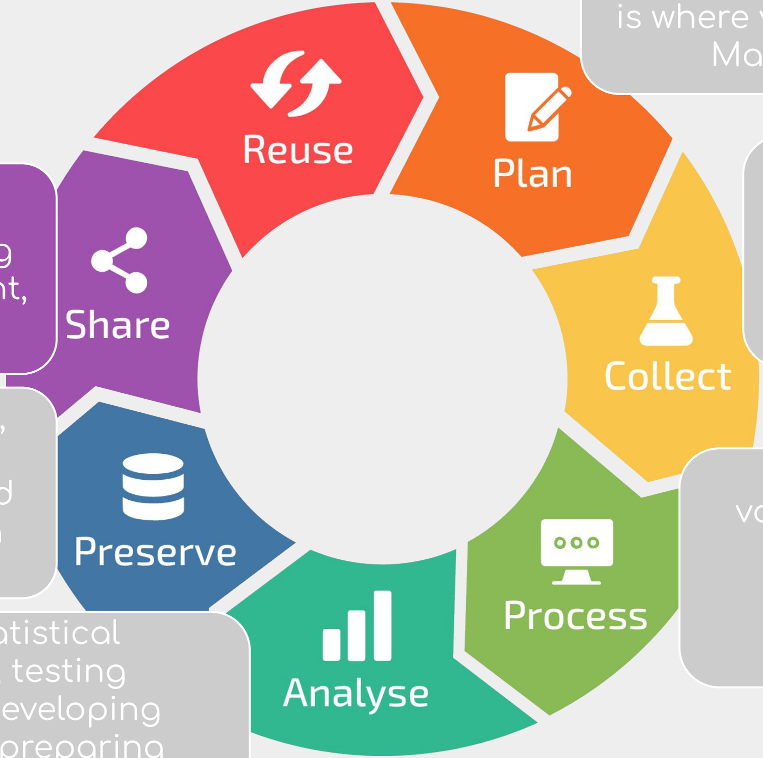
Data processing means validating and cleaning the data, organizing and transforming the data, anonymizing them, if necessary.

Running statistical procedures, testing hypotheses, developing explanations, preparing insights, reports, thinking about authorship.

Taking care of data storage, back-up and archiving, migrating to best format and medium, creating metadata and documentation.

Picture

the research data life cycle



Ideally, you start thinking about FAIR data even before you start collecting them! This is where you design your Data Management Plan.

Data collection is where the data is created, either by collecting new data, or locating existing data, for example reusing someone other's data.

Data processing means validating and cleaning the data, organizing and transforming the data, anonymizing them, if necessary.

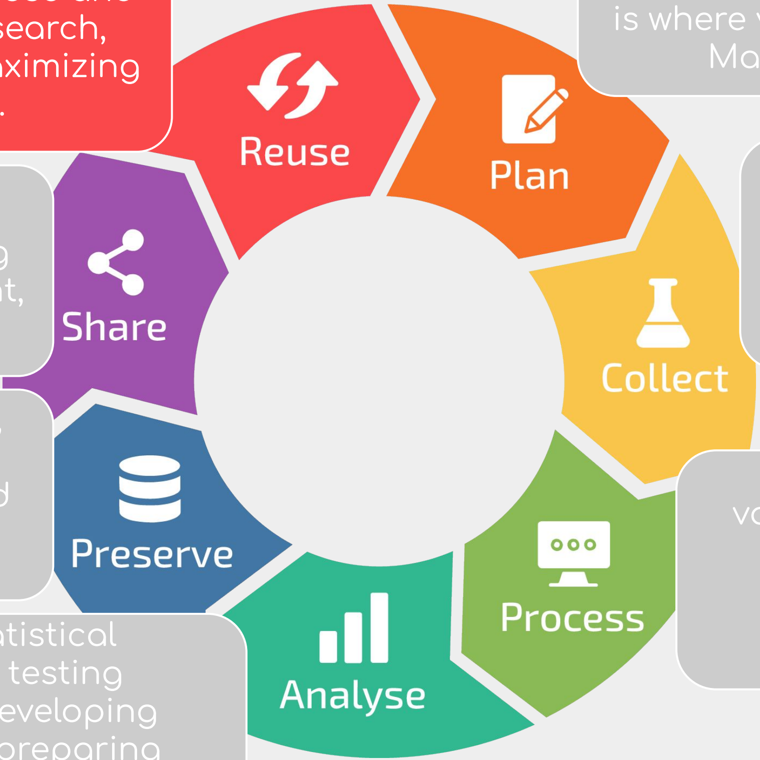
Running statistical procedures, testing hypotheses, developing explanations, preparing insights, reports, thinking about authorship.

Taking care of data storage, back-up and archiving, migrating to best format and medium, creating metadata and documentation.

Making data available: distributing data, controlling access, establishing copyright, promoting data.

Picture

the research data life cycle



Ideally, you start thinking about FAIR data even before you start collecting them! This is where you design your Data Management Plan.

Data collection is where the data is created, either by collecting new data, or locating existing data, for example reusing someone other's data.

Data processing means validating and cleaning the data, organizing and transforming the data, anonymizing them, if necessary.

Running statistical procedures, testing hypotheses, developing explanations, preparing insights, reports, thinking about authorship.

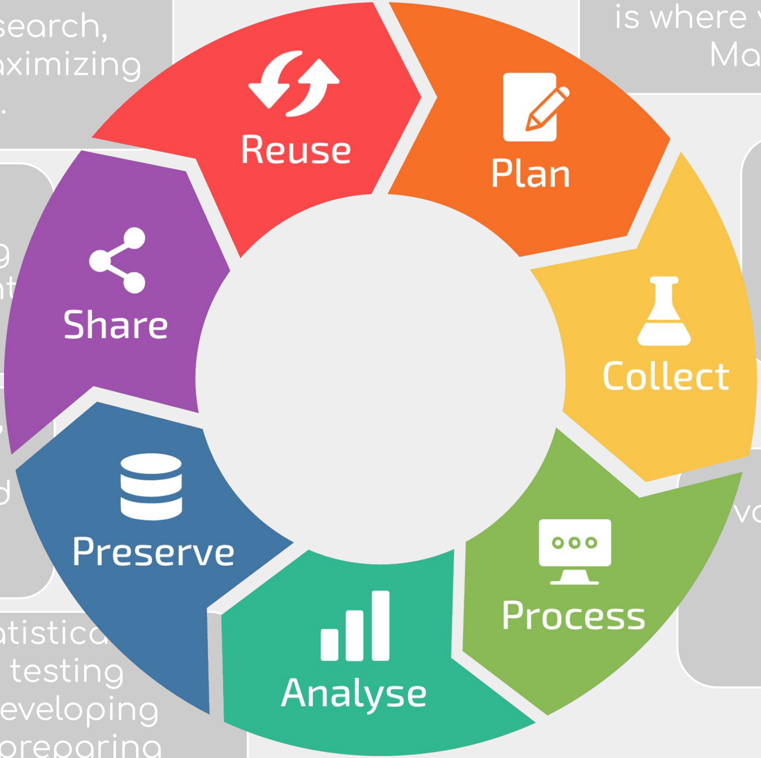
Taking care of data storage, back-up and archiving, migrating to best format and medium, creating metadata and documentation.

Making data available: distributing data, controlling access, establishing copyright, promoting data.

Enabling data repurpose and reuse, follow-ups research, teaching, learning, maximizing data life cycle.

Picture

the research data life cycle



Ideally, you start thinking about FAIR data even before you start collecting them! This is where you design your Data Management Plan.

Data collection is where the data is created, either by collecting new data, or locating existing data, for example reusing someone other's data.

Data processing means validating and cleaning the data, organizing and transforming the data, anonymizing them, if necessary.

Running statistica procedures, testing hypotheses, developing explanations, preparing insights, reports, thinking about authorship.

Taking care of data storage, back-up and archiving, migrating to best format and medium, creating metadata and documentation.

Making data available: distributing data, controlling access, establishing copyright promoting data.

Enabling data repurpose and reuse, follow-ups research, teaching, learning, maximizing data life cycle.

Picture

data terminologies (1)

deposit data: upload a digital object on a platform that allows to correctly describe it through its metadata, and that implements long-term preservation

give access to data: once the object has been deposited somewhere, it’s up to you to choose which type of access you want to grant (through licenses, which we will see later)

data terminologies (2)

data sharing: any way of sharing information; you ask me some data, I email you back with an attachment

data publishing: depositing data so that it becomes a citable artifact, discoverable

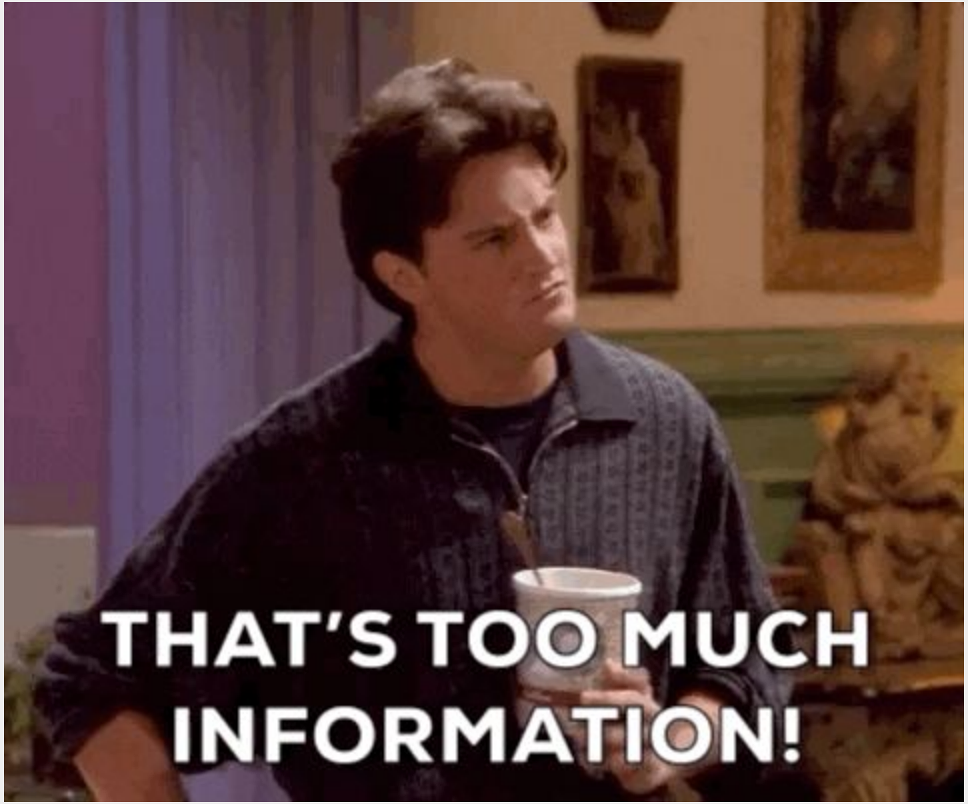
data archiving: thinking about data storage in the long-term, preserving your data

data terminologies (3)

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

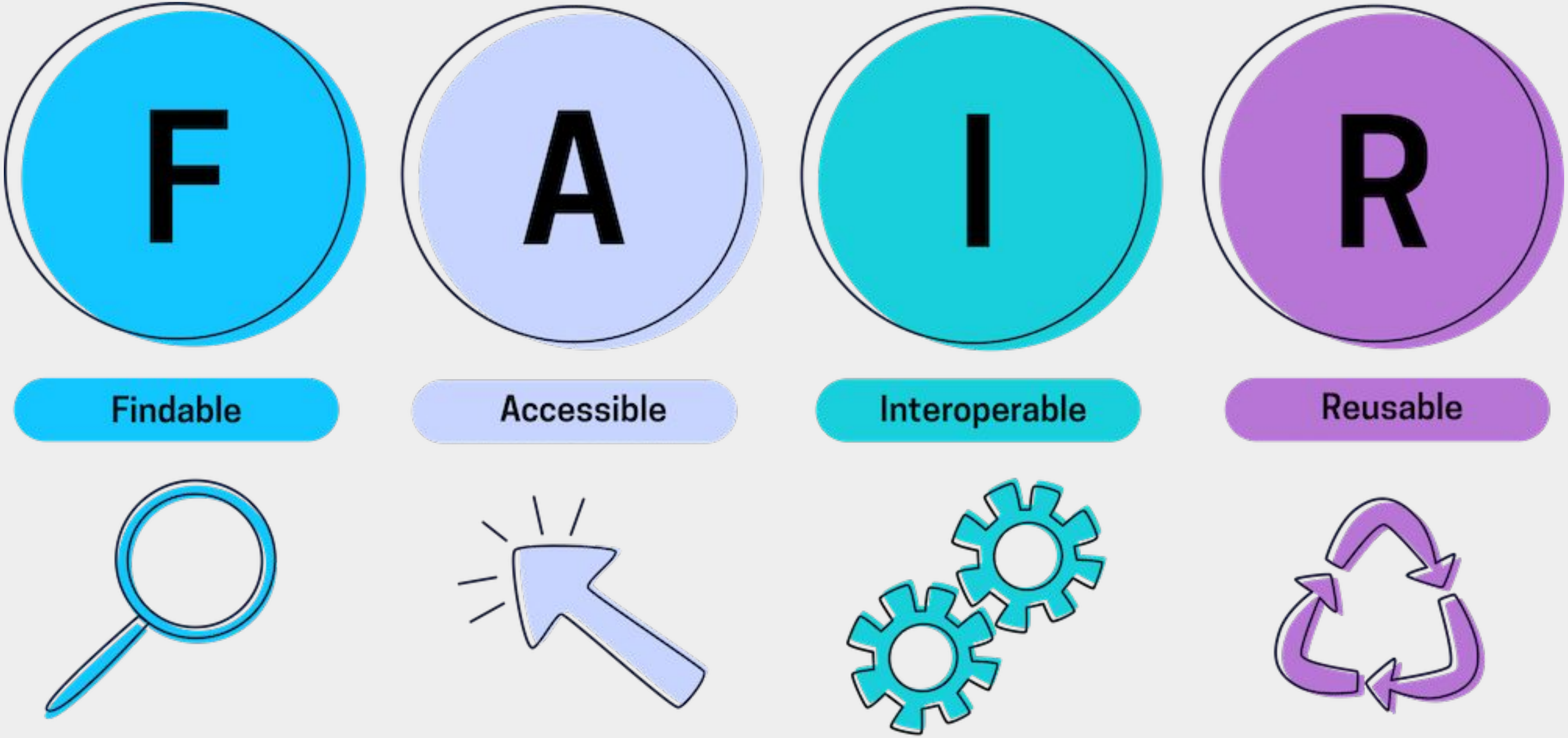
data stewardship: making data reusable for the long-term, also after the project has ended (data preservation)

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)

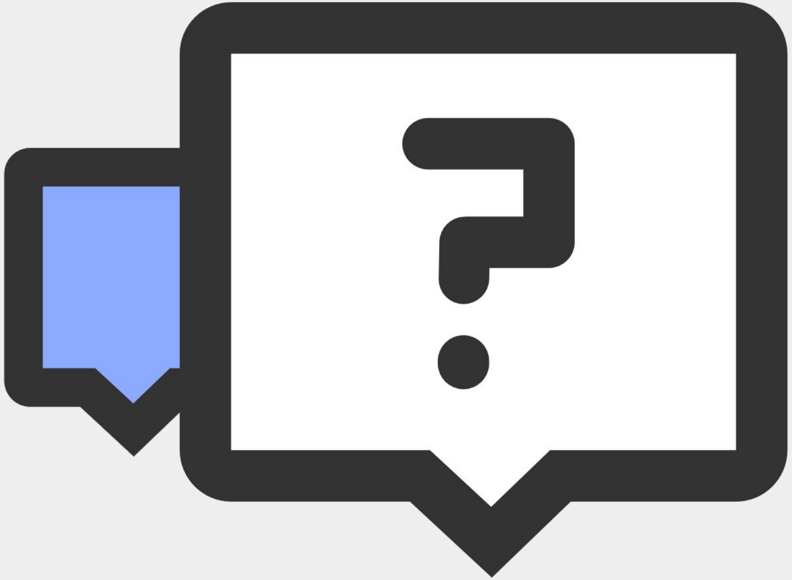


https://media1.giphy.com/media/lPF1CyJXXcTZmUrP2J/giphy.gif

the FAIR principles as guidance for data stewardship



Time for Questions



the (rough) agenda

Introduction to research data and FAIR - 12h15-12:45

The FAIR principles in action - 13h00-14h00

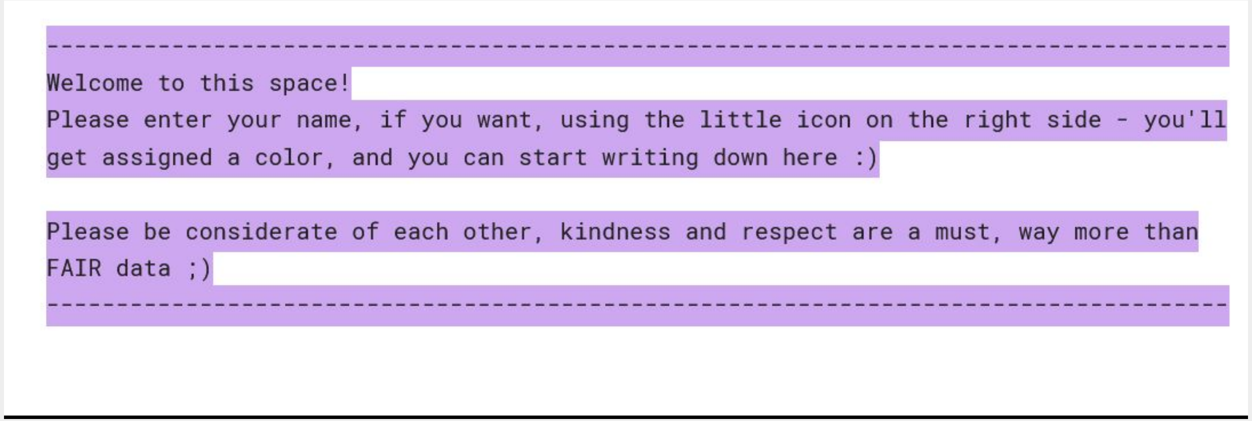
F for Findable
A for Accessible
I for Interoperable
R for Reusable

Lunch 14h00-15h00

FAIRify (your) data - 15h00-16h15

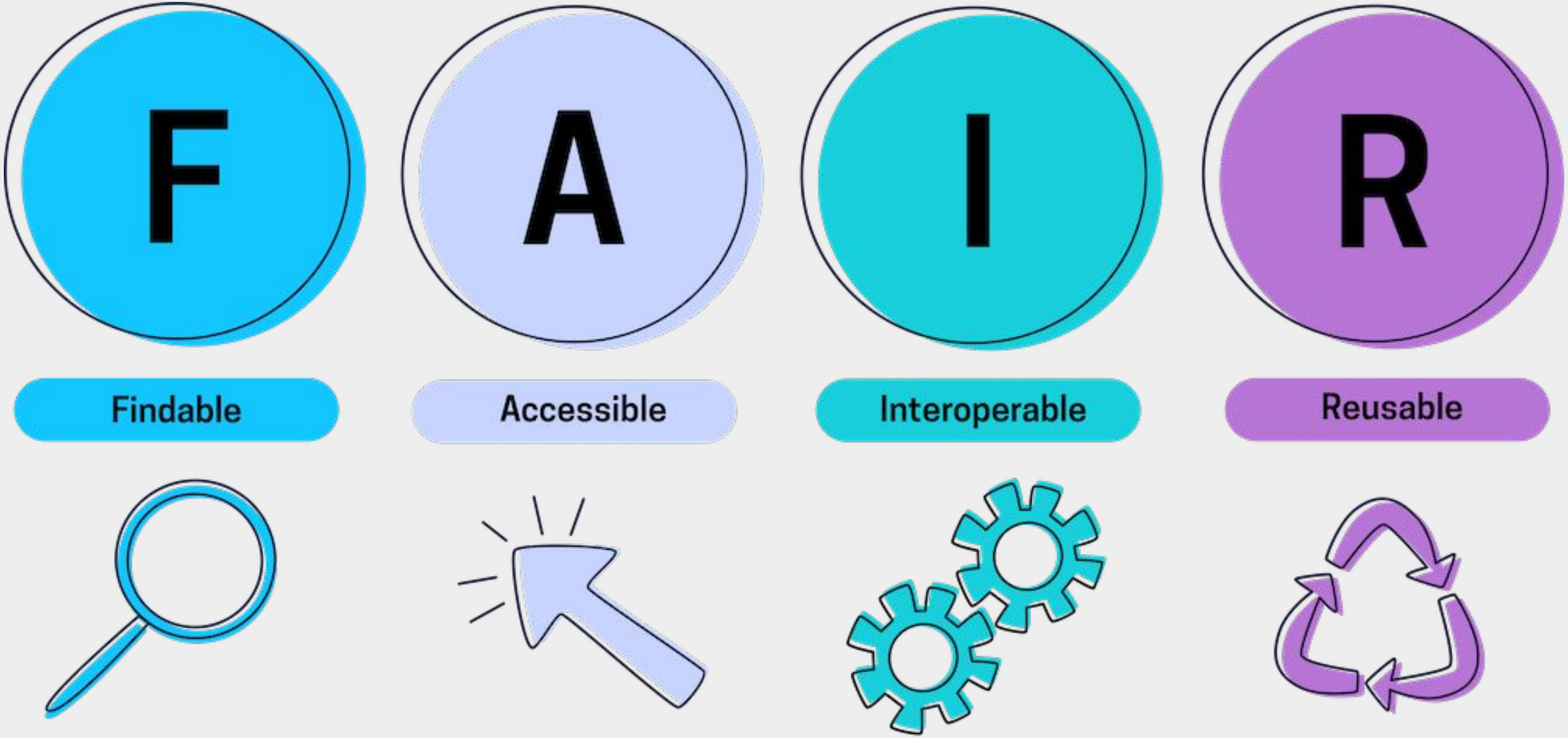
the collaborative pad

http://tiny.cc/maribor



```
--------------------------------------------------------------------------------------
Welcome to this space!
Please enter your name, if you want, using the little icon on the right side - you'll
get assigned a color, and you can start writing down here :)

Please be considerate of each other, kindness and respect are a must, way more than
FAIR data ;)
--------------------------------------------------------------------------------------
```

the FAIR principles as guidance for data stewardship



Findable

Accessible

Interoperable

Reusable

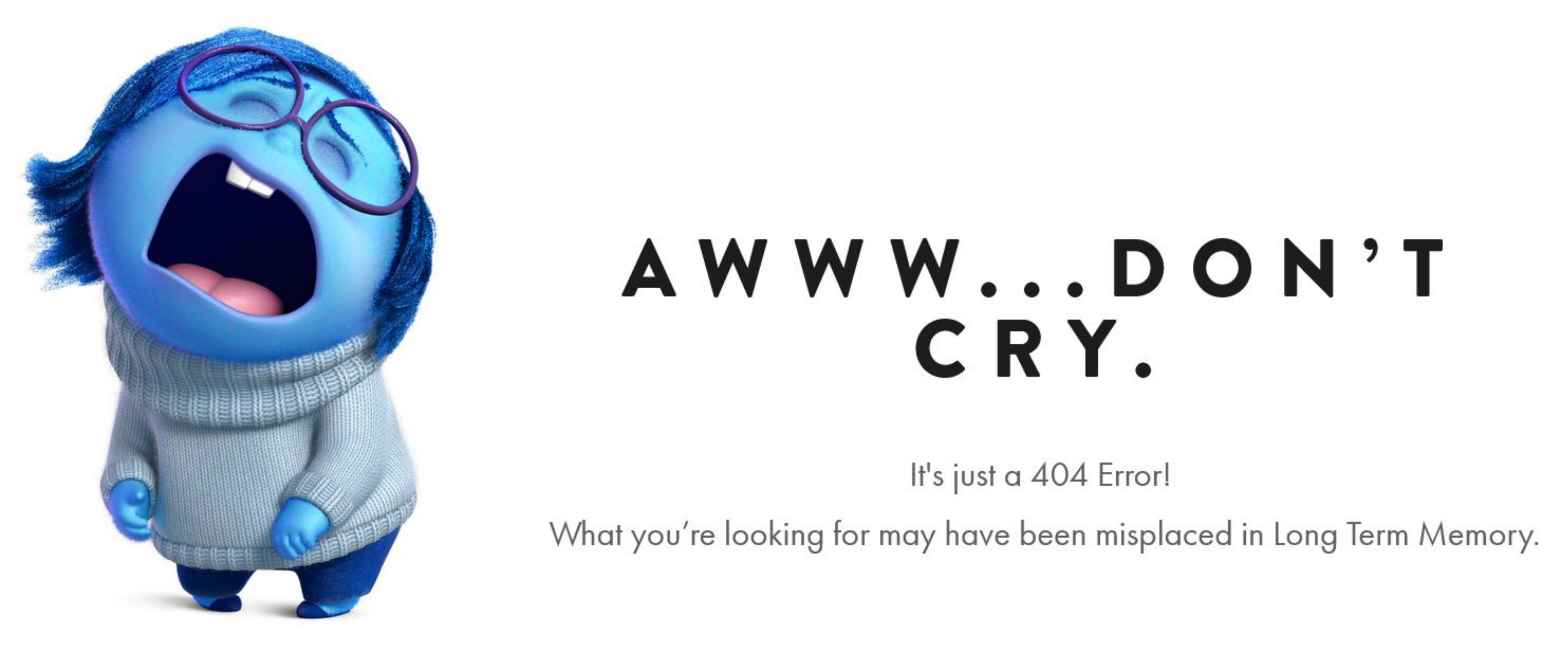
F is for Findable

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

have you ever landed on a 404 page?



AWWW...DON'T CRY.

It's just a 404 Error!

What you're looking for may have been misplaced in Long Term Memory.

https://www.pixar.com/404

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

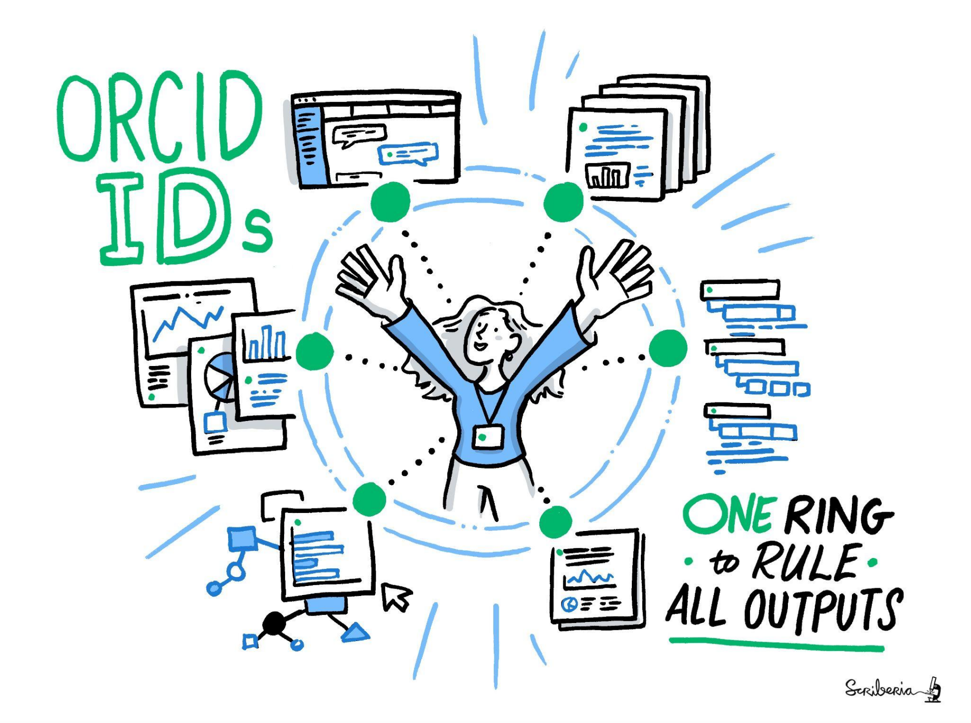
A persistent identifier (PID) is a long-lasting reference to a digital resource and provides the information required to reliably identify, verify and locate your research data, eliminating many misunderstandings.

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.

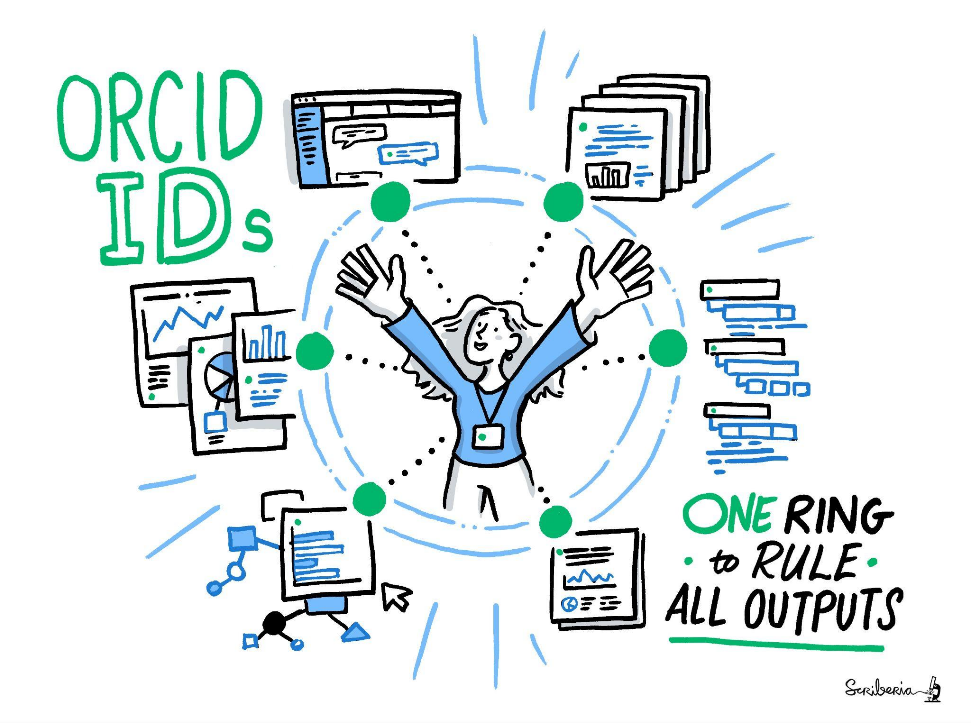
a PID for researchers



the Open Researcher and Contributor ID

Image; https://orcid.org/

a PID for researchers



the Open Researcher and Contributor ID

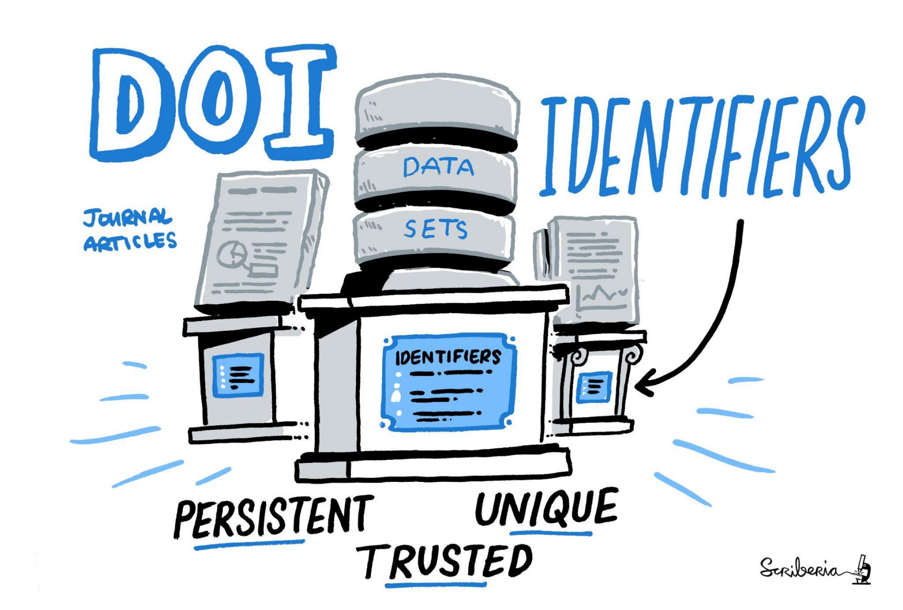
https://orcid.org/0000-0003-3699-1195

- ➔ do you have an ORCID?
- ➔ do you use it as part of your affiliation when submitting articles to journals?
- ➔ what do you think are the benefits of having an ORCID?

Image; https://orcid.org/

anatomy of a DOI

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



resolver service

doi

directory indicator +prefix (assigning body)

suffix resource

Picture

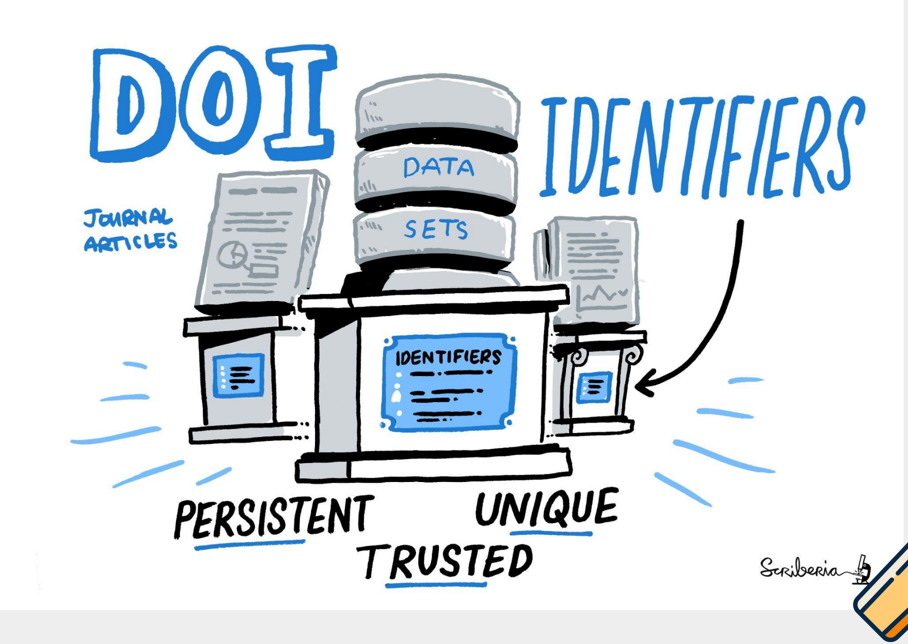
anatomy of a DOI

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

resolver service

directory indicator +prefix (assigning body)

suffix resource



- ➔ go to https://www.doi.org/ and resolve some DOI's
- ➔ you can use some examples from https://www.doi.org/demos.html or use this one from me: 10.5281/zenodo.7260977

Picture

no PID? no 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? no 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? no 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?

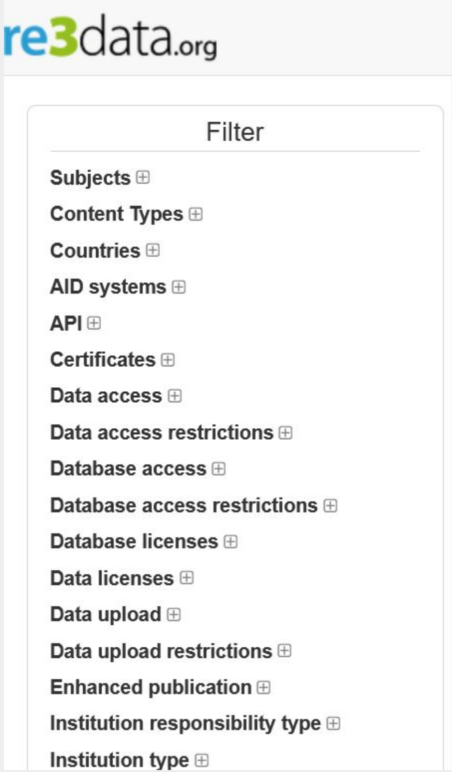
deposit your data in a trusted repository that issues a PID

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

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

zenodo

re3data listed repositories



re3data.org

Filter

- 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 ⊞
- Institution responsibility type ⊞
- Institution type ⊞

➔ go to https://www.re3data.org/search and look for a repository that could host the type of data for your research, or some data you are interested in

➔ check the persistent identifier and the data access field

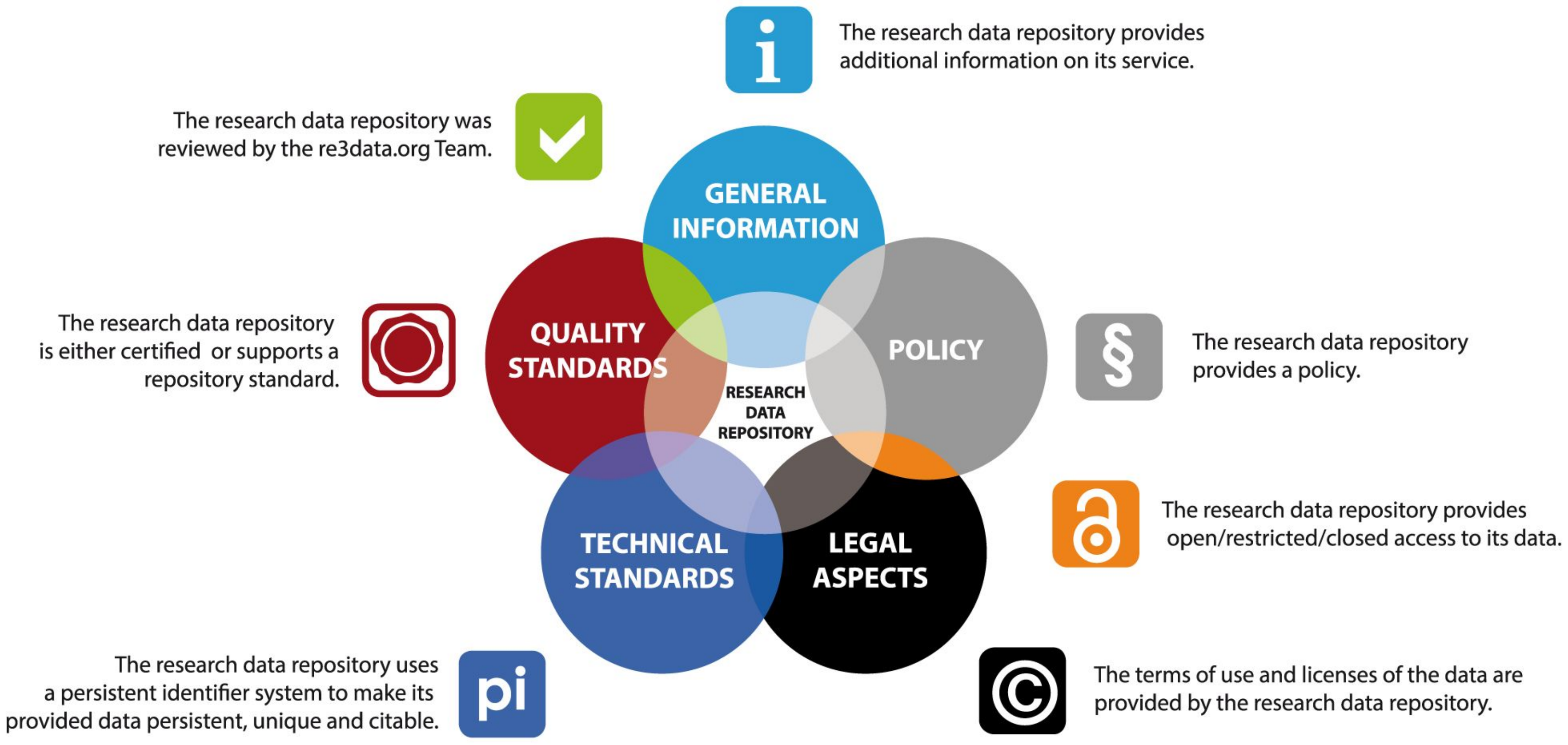
➔ which other fields are there? something you want to know more about? let's discuss!

re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

https://www.re3data.org/

aspects of a research data repository



The research data repository provides additional information on its service.

The research data repository was reviewed by the re3data.org Team.

The research data repository is either certified or supports a repository standard.

The research data repository provides a policy.

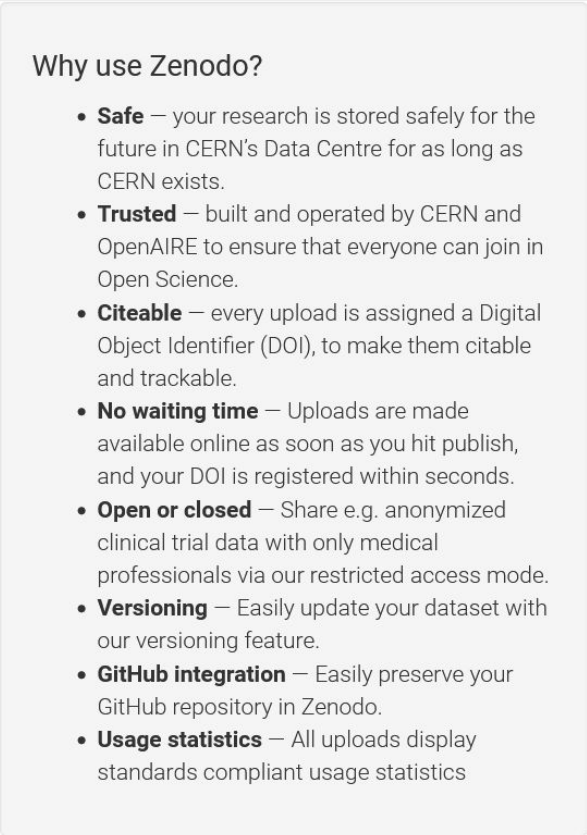
The research data repository provides open/restricted/closed access to its data.

The research data repository uses a persistent identifier system to make its provided data persistent, unique and citable.

The terms of use and licenses of the data are provided by the research data repository.

Picture

Zenodo: a general-purpose repository



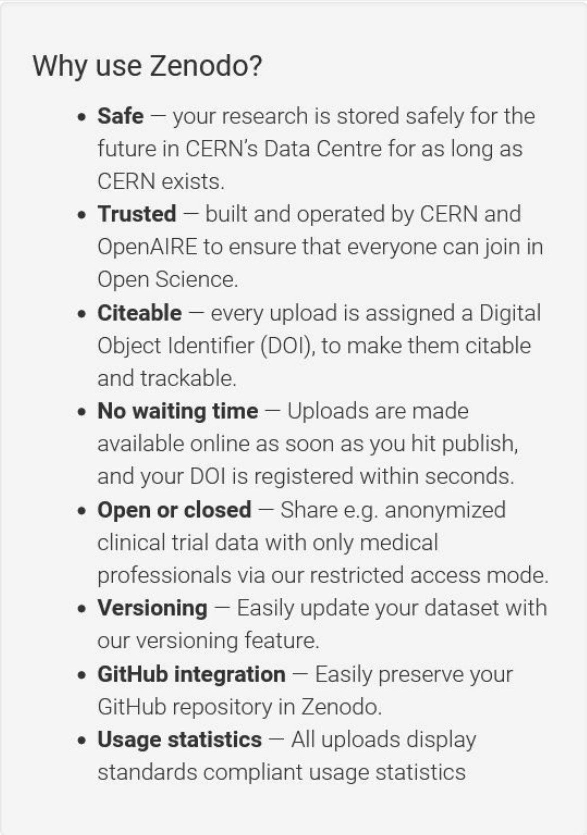
Why use Zenodo?

- **Safe** — your research is stored safely for the future in CERN's Data Centre for as long as CERN exists.
- **Trusted** — built and operated by CERN and OpenAIRE to ensure that everyone can join in Open Science.
- **Citeable** — every upload is assigned a Digital Object Identifier (DOI), to make them citable and trackable.
- **No waiting time** — Uploads are made available online as soon as you hit publish, and your DOI is registered within seconds.
- **Open or closed** — Share e.g. anonymized clinical trial data with only medical professionals via our restricted access mode.
- **Versioning** — Easily update your dataset with our versioning feature.
- **GitHub integration** — Easily preserve your GitHub repository in Zenodo.
- **Usage statistics** — All uploads display standards compliant usage statistics

zenodo

https://zenodo.org/

Zenodo: a general-purpose repository



Why use Zenodo?

- **Safe** — your research is stored safely for the future in CERN's Data Centre for as long as CERN exists.
- **Trusted** — built and operated by CERN and OpenAIRE to ensure that everyone can join in Open Science.
- **Citeable** — every upload is assigned a Digital Object Identifier (DOI), to make them citable and trackable.
- **No waiting time** — Uploads are made available online as soon as you hit publish, and your DOI is registered within seconds.
- **Open or closed** — Share e.g. anonymized clinical trial data with only medical professionals via our restricted access mode.
- **Versioning** — Easily update your dataset with our versioning feature.
- **GitHub integration** — Easily preserve your GitHub repository in Zenodo.
- **Usage statistics** — All uploads display standards compliant usage statistics

https://zenodo.org/

zenodo

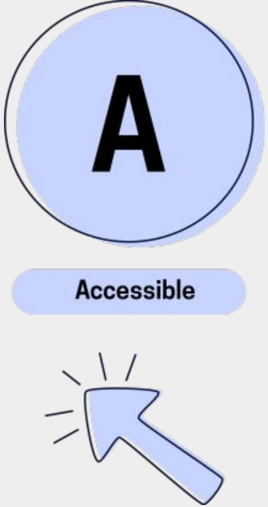
- ➔ go to https://sandbox.zenodo.org/ and either register or login
- ➔ we will practice on the Zenodo sandbox (and not on the "real" one)
- ➔ this is because once DOIs are created on the "real" service, they cannot be easily removed or modified (which makes sense!)
- ➔ on the sandbox service, data are deleted after a certain time

A is for accessible

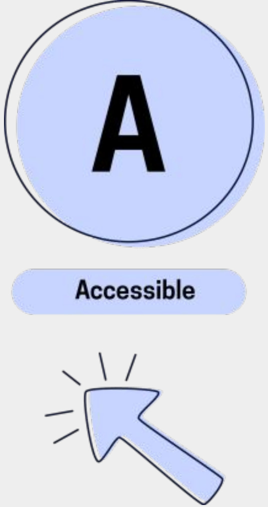
accessible does not imply open

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



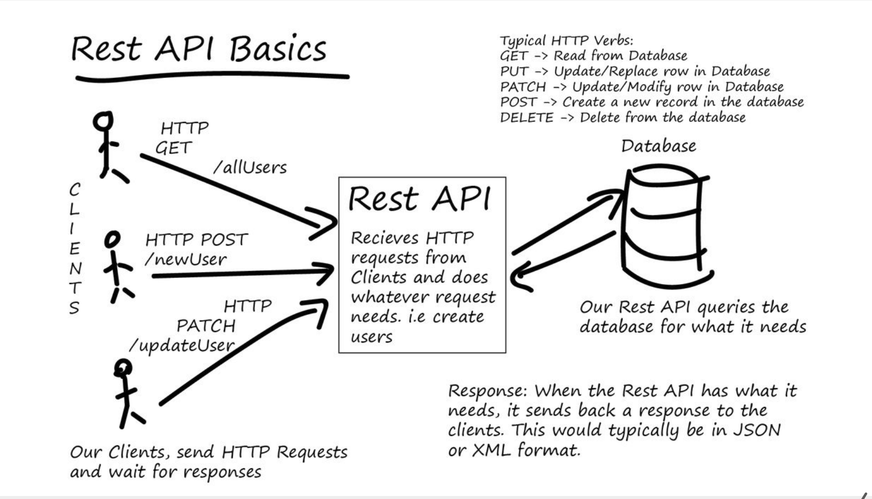
A is for accessible



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



Typical HTTP Verbs:
GET -> Read from Database
PUT -> Update/Replace row in Database
PATCH -> Update/Modify row in Database
POST -> Create a new record in the database
DELETE -> Delete from the database

Recieves HTTP requests from Clients and does whatever request needs. i.e create users

Our Rest API queries the database for what it needs

Response: When the Rest API has what it needs, it sends back a response to the clients. This would typically be in JSON or XML format.

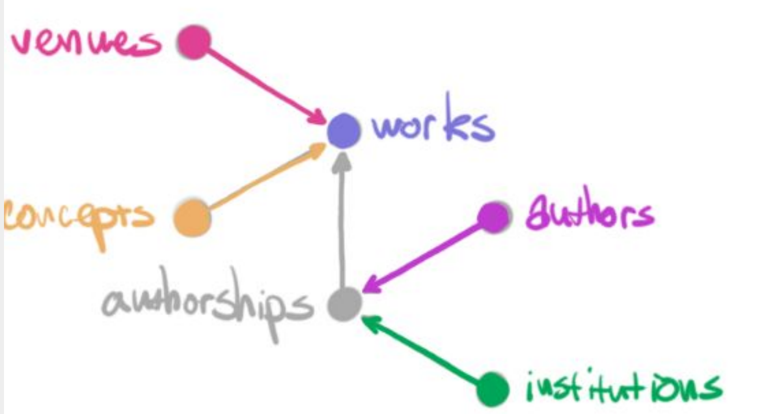
Our Clients, send HTTP Requests and wait for responses

https://en.wikipedia.org/wiki/API

the case of OpenAlex

An open and comprehensive catalog of scholarly papers, authors, institutions, and more.

Figure 1: Sketch of the OpenAlex graph data model.



https://openalex.org/; https://docs.openalex.org/; https://arxiv.org/abs/2205.01833

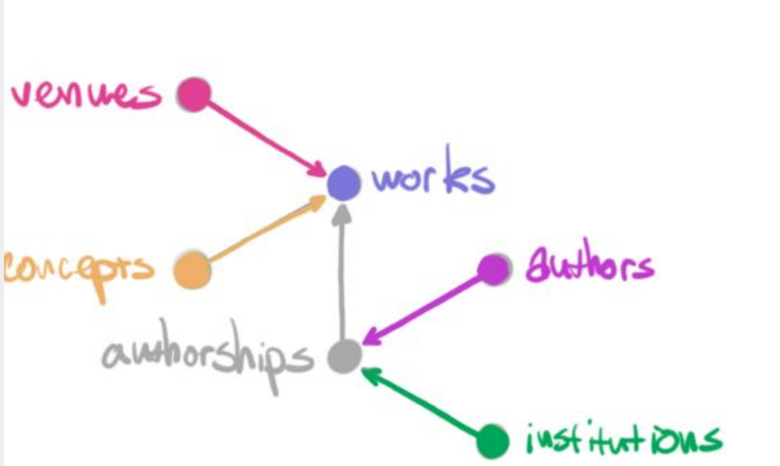
the case of OpenAlex

OpenAlex

An open and comprehensive catalog of scholarly papers, authors, institutions, and more.

our research

Figure 1: Sketch of the OpenAlex graph data model.



- ➔ the API is the primary way to get OpenAlex data; it's free and requires no authentication
- ➔ we will build some API calls and look at the returned outputs
- ➔ https://api.openalex.org/authors/orcid:0000-0003-3699-1195
- ➔ https://api.openalex.org/institutions?filter=display_name.search:university%20of%20maribor
- ➔ https://api.openalex.org/institutions/I37696226
- ➔ https://api.openalex.org/works?filter=institutions.id:https://api.openalex.org/institutions/I37696226,is_paratext:false,type:journal-article,from_publication_date:2020-01-01

https://openalex.org/; https://docs.openalex.org/; https://arxiv.org/abs/2205.01833

I is for Interoperable

data & metadata need to be interoperable: it needs to be possible to combine them with other data & tools

this means that their format needs to be open, and that both data & metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation (controlled vocabularies and ontologies, wherever possible)

I is for Interoperable

data & metadata need to be interoperable: it needs to be possible to combine them with other data & tools

this means that their format needs to be open, and that both data & metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation (controlled vocabularies and ontologies, wherever possible)

PDF — not machine-readable

XLS — machine-readable, but closed, not standard format

CSV — machine-readable in a format that is both open and standard

human-readable

human-readable

data that can be read, processed, by people, human beings

we can easily read a PDF document, but an algorithm/a machine can't because the representation of the data on disk does not reflect the relationship of the data in reality

https://opendatahandbook.org/glossary/en/terms/human-readable/

human-readable *vs* machine-readable

human-readable

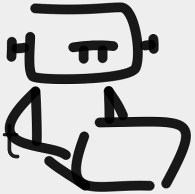
data that can be read, processed, by people, human beings

we can easily read a PDF document, but an algorithm/a machine can't because the representation of the data on disk does not reflect the relationship of the data in reality

machine-readable

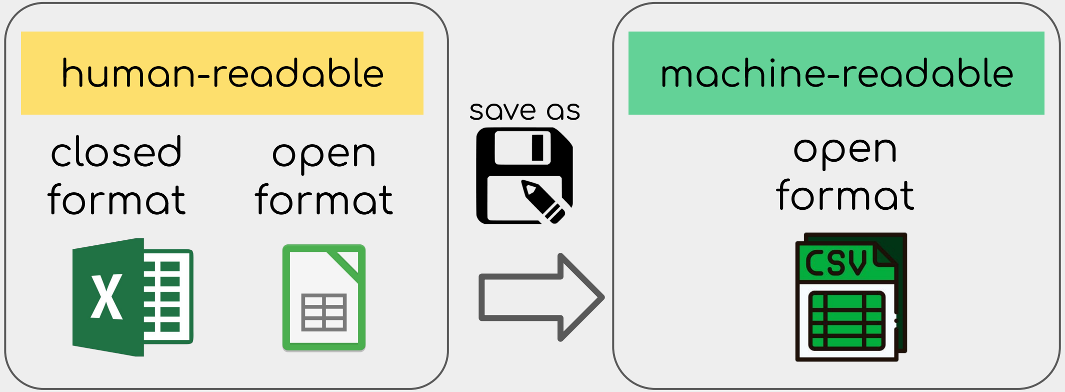
data in a format that can be automatically read and processed by a computer, such as CSV, RDF, JSON, XML, etc.

machine-readable data must be ***structured data***



https://opendatahandbook.org/glossary/en/terms/machine-readable/

data organized in spreadsheets



human-readable

closed format

open format

save as

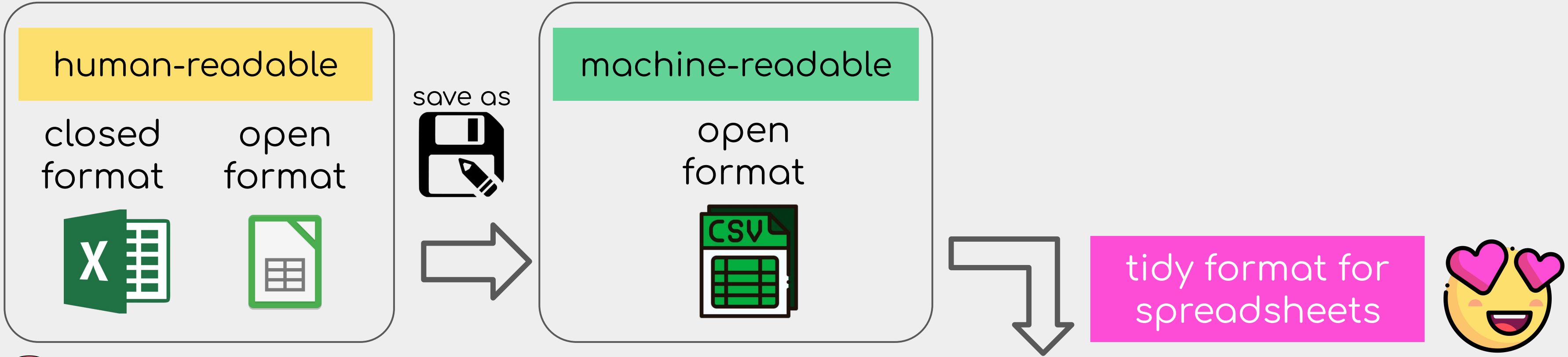
machine-readable

open format

❓ Ask yourself: can anybody open this file and look at the data? Do people need to buy specific software to do so?

https://www.libreoffice.org/discover/what-is-opendocument/

data organized in spreadsheets



human-readable

closed format

open format

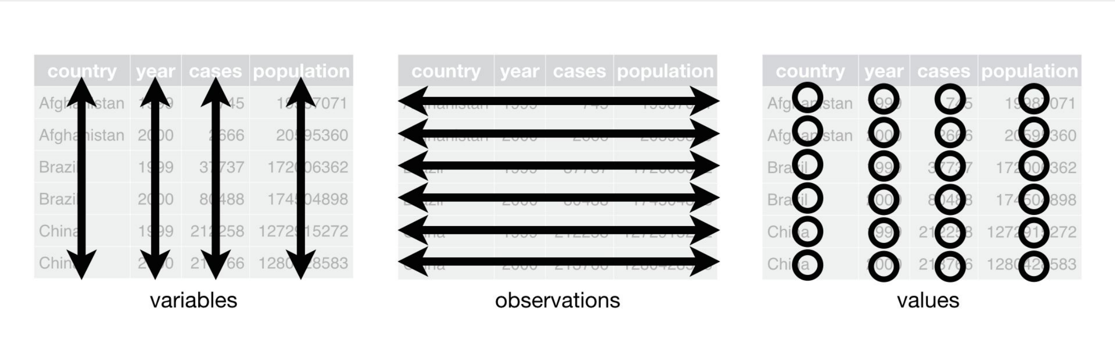
save as

machine-readable

open format

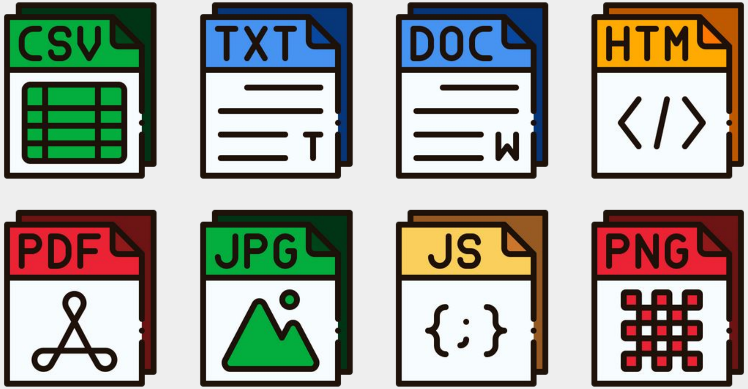
tidy format for spreadsheets

❓Ask yourself: can anybody open this file and look at the data? Do people need to buy specific software to do so?

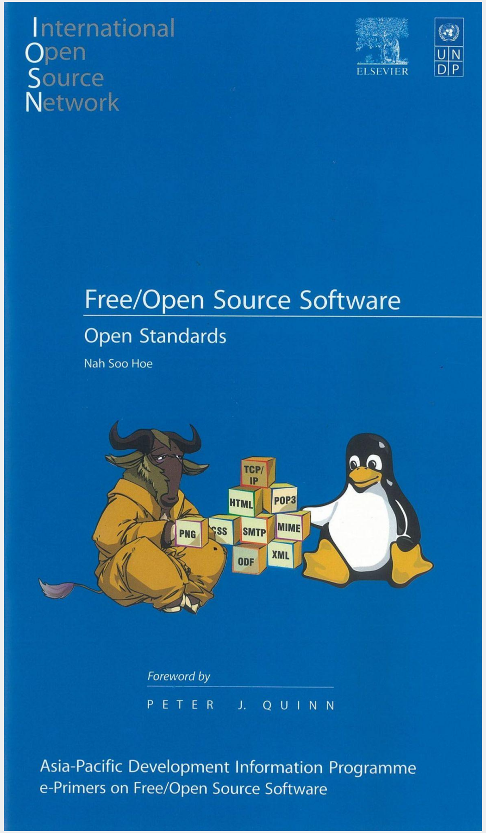


https://the-turing-way.netlify.app/reproducible-research/rdm/rdm-spreadsheets.html

open file formats



- ➔ go to https://en.wikipedia.org/wiki/List_of_open_file_formats: is there a format you don't know?
- ➔ browse the wikibook https://en.wikibooks.org/wiki/FOSS_Open_Standards/Comparison_of_File_Formats and take a look at the Office Document Formats; is PDF an open format?



International
Open
Source
Network

ELSEVIER

UN
DP

Free/Open Source Software

Open Standards

Nah Soo Hoe

Foreword by

PETER J. QUINN

Asia-Pacific Development Information Programme
e-Primers on Free/Open Source Software

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

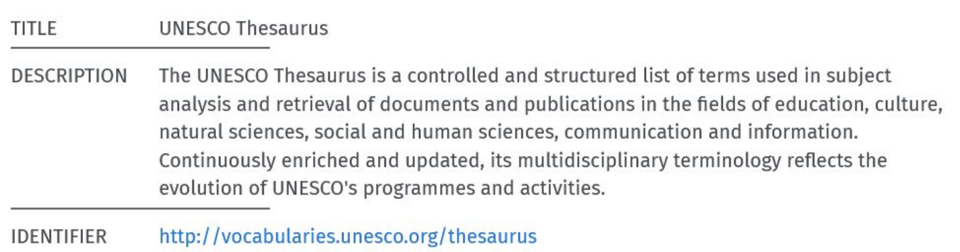
controlled vocabularies for FAIR metadata

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

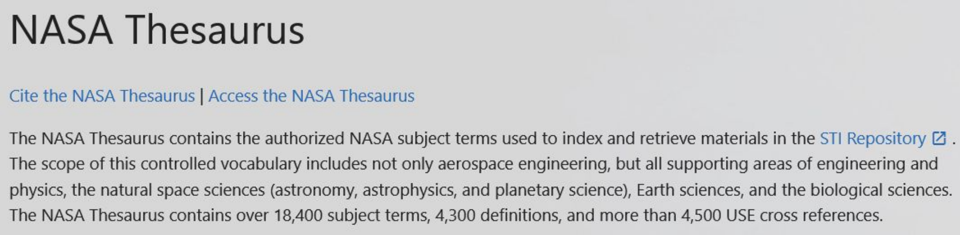
unesco

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



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

https://vocabularies.unesco.org/browser/thesaurus/en/; https://sti.nasa.gov/nasa-thesaurus/#access-the-nasa-thesaurus; https://fairsharing.org/search?fairsharingRegistry=Standard

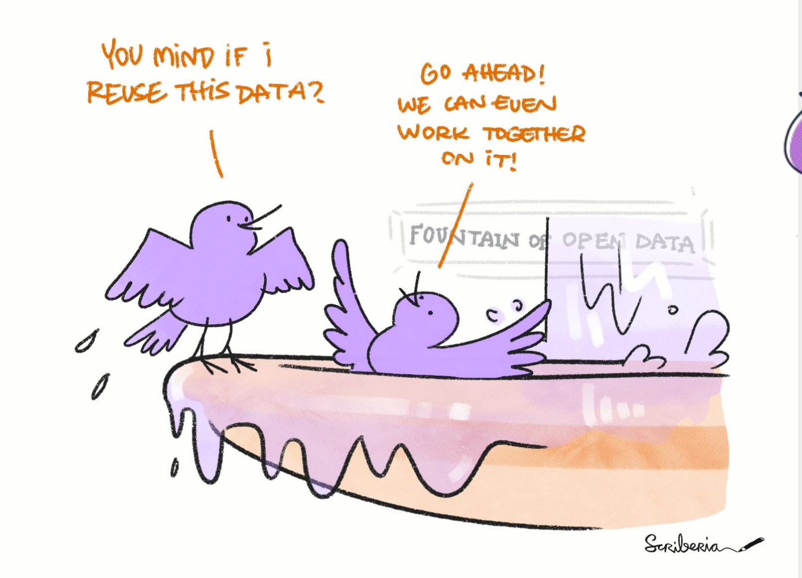
R is for Reusable

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


YOU MIND IF I REUSE THIS DATA?
GO AHEAD! WE CAN EVEN WORK TOGETHER ON IT!
FOUNTAIN OF OPEN DATA
Scriberia

http://doi.org/10.5281/zenodo.3695300

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	
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)

Open Knowledge Foundation

https://opendefinition.org/licenses/

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 Knowledge Foundation

https://opendefinition.org/licenses/

creative commons licenses

creative commons

	Derivatives Can Be Shared	Derivatives Can Be Shared ONLY IF You Share Alike	Derivatives CANNOT Be Shared
Commercial Use Allowed	cc BY	cc BY SA	cc BY ND
Commercial Use NOT Allowed	cc BY NC	cc BY NC SA	cc BY NC ND

https://docs.google.com/presentation/d/15DhTr6A5i6PYWtUriyvxGJYCriobgm_BxvxdL2yol9g/edit#slide=id.g4b305805a9_0_22

creative commons licenses

creative commons

	Derivatives Can Be Shared	Derivatives Can Be Shared ONLY IF You Share Alike	Derivatives CANNOT Be Shared
Commercial Use Allowed	CC BY	CC BY SA	CC BY ND
Commercial Use NOT Allowed	CC BY NC	CC BY NC SA	CC BY NC ND

No Derivative Works ND

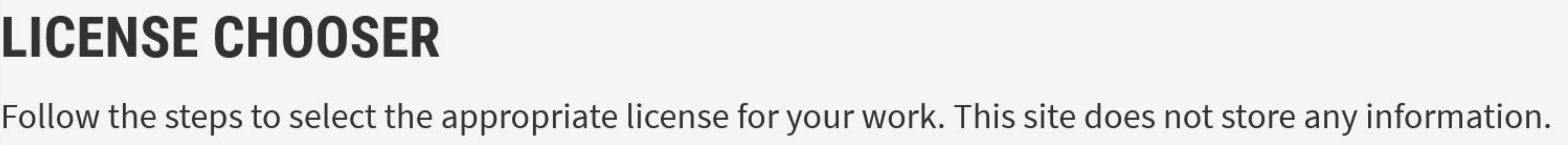
Others can only copy, distribute, display or perform verbatim copies of your work

Non-Commercial NC

Others can copy, distribute, display, perform or remix your work but for non-commercial purposes only.

https://upload.wikimedia.org/wikipedia/commons/c/c7/Free_Knowledge_thanks_to_Creative_Commons_Licenses.pdf

choosing a license



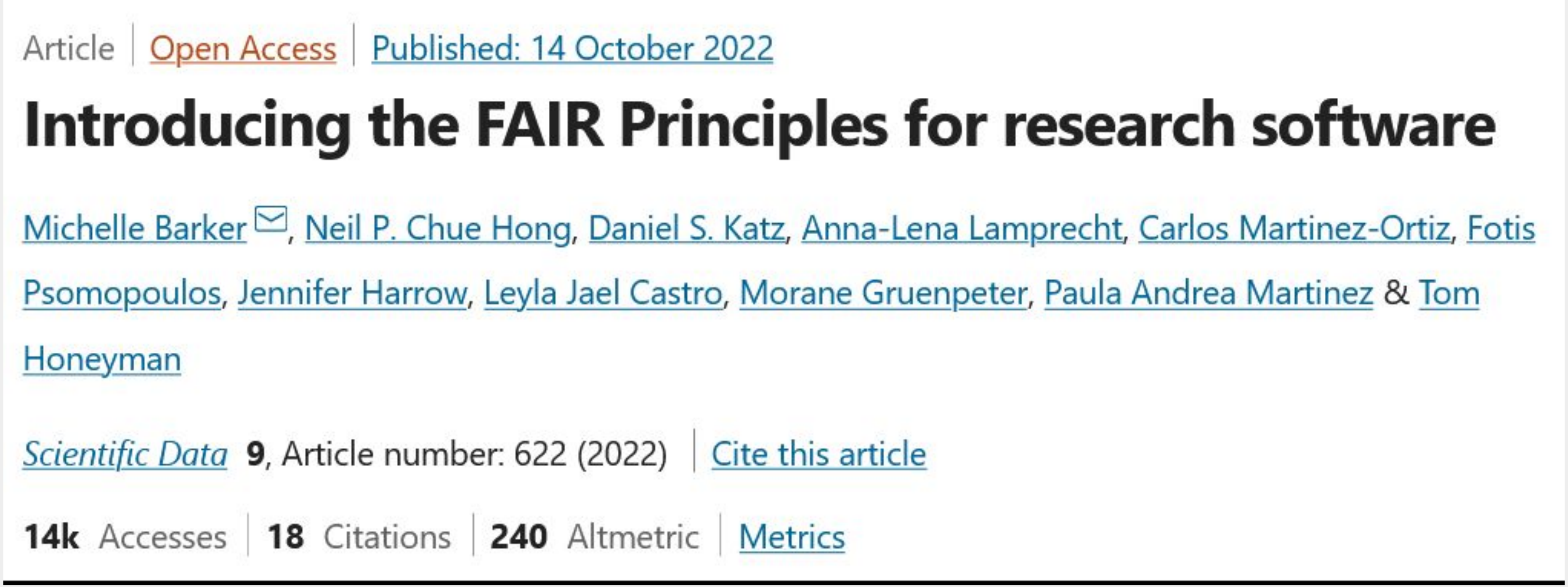
LICENSE CHOOSER

Follow the steps to select the appropriate license for your work. This site does not store any information.

- go to the beta license chooser from Creative Commons: https://chooser-beta.creativecommons.org/ and walk through the steps to choose an appropriate license for your work
- software needs specific licenses: go to https://choosealicense.com/ and take a look at what options are available; select the MIT license and see what it entails
- general tip: avoid writing bespoke licenses!

Choose an open source license

FAIR4RS



Article | Open Access | Published: 14 October 2022

Introducing the FAIR Principles for research software

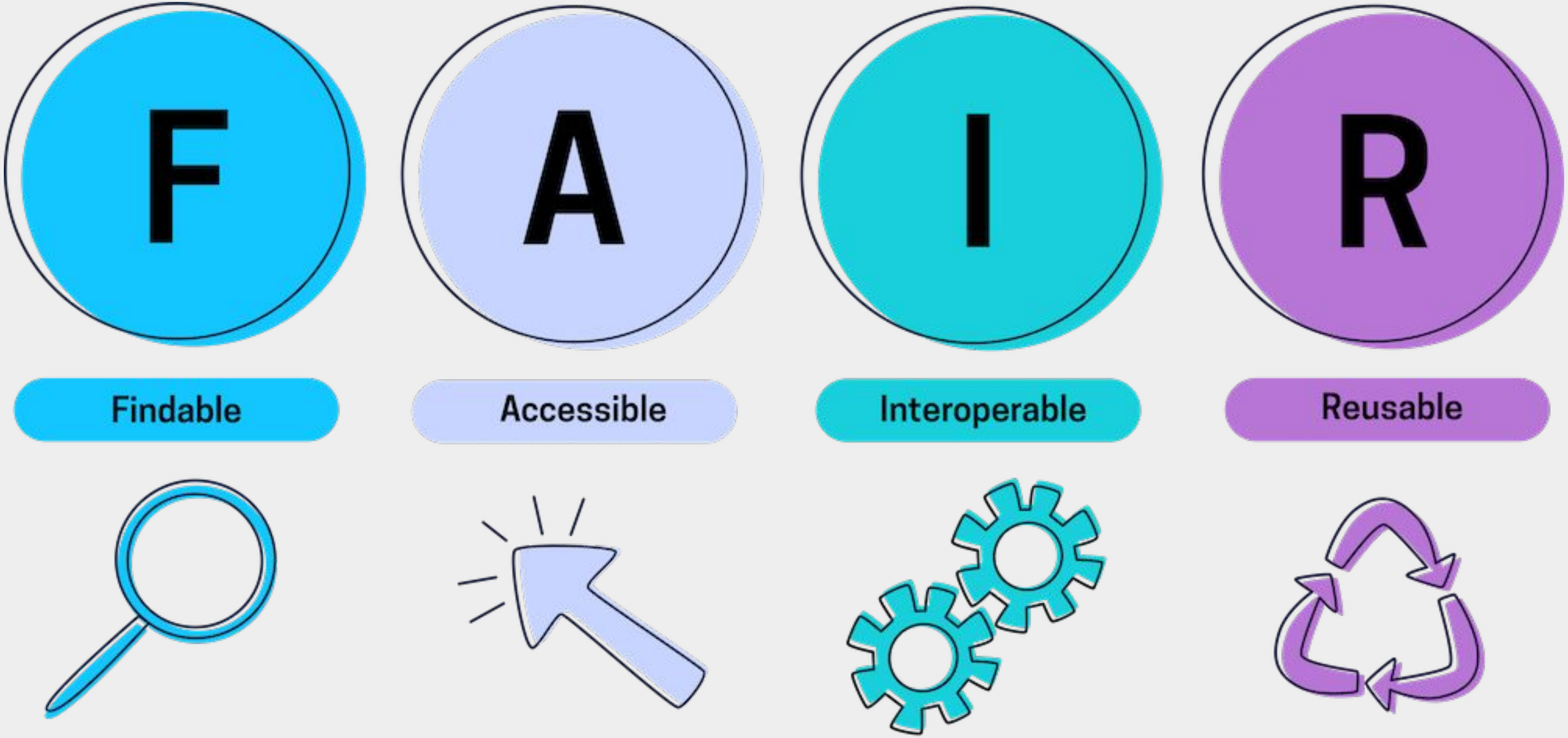
Michelle Barker, Neil P. Chue Hong, Daniel S. Katz, Anna-Lena Lamprecht, Carlos Martinez-Ortiz, Fotis Psomopoulos, Jennifer Harrow, Leyla Jael Castro, Morane Gruenpeter, Paula Andrea Martinez & Tom Honeyman

Scientific Data **9**, Article number: 622 (2022) | Cite this article

14k Accesses | **18** Citations | **240** Altmetric | Metrics

https://www.nature.com/articles/s41597-022-01710-x#Sec2

a short recap on FAIR



F A I R

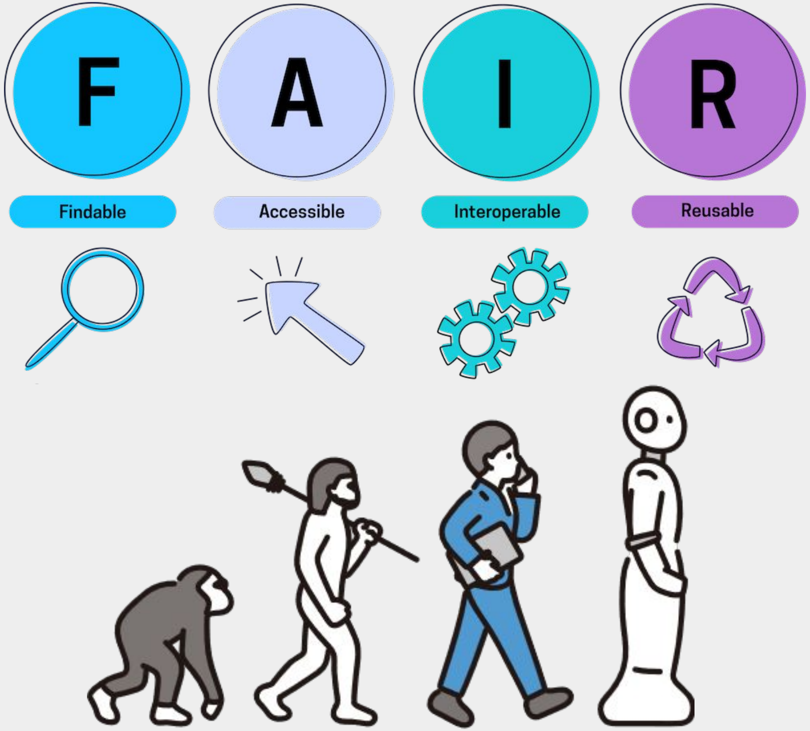
Findable

Accessible

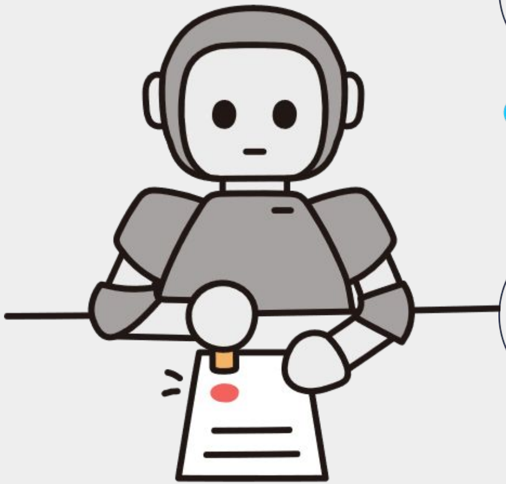
Interoperable

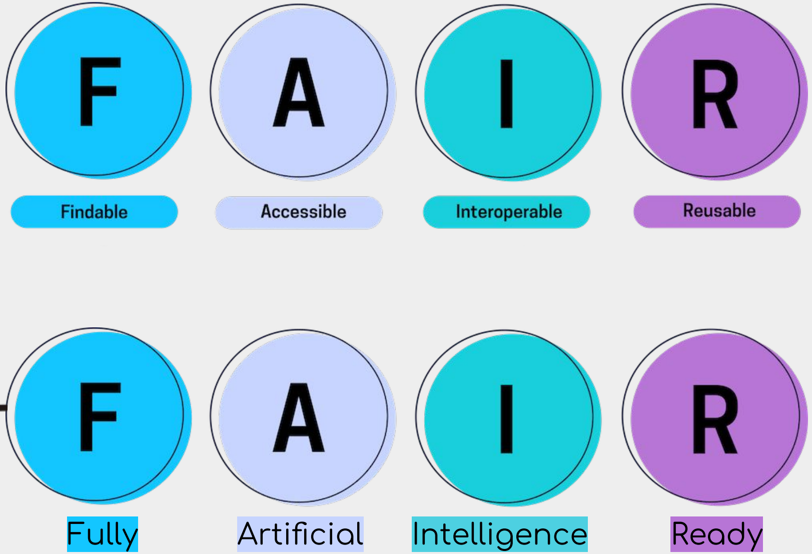
Reusable

FAIR for people and for machines



FAIR for people and for machines





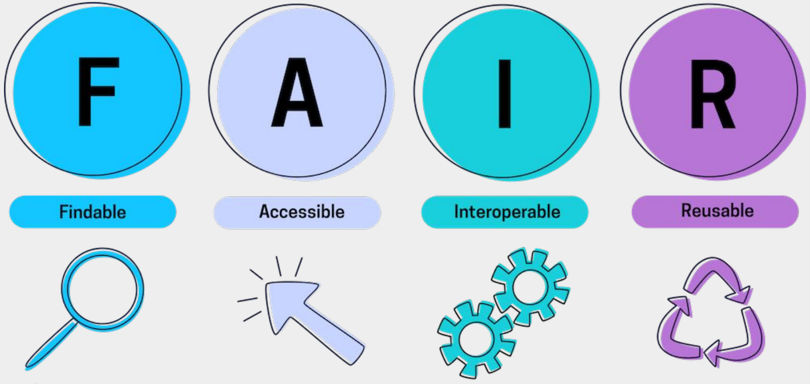
F A I R

Findable Accessible Interoperable Reusable

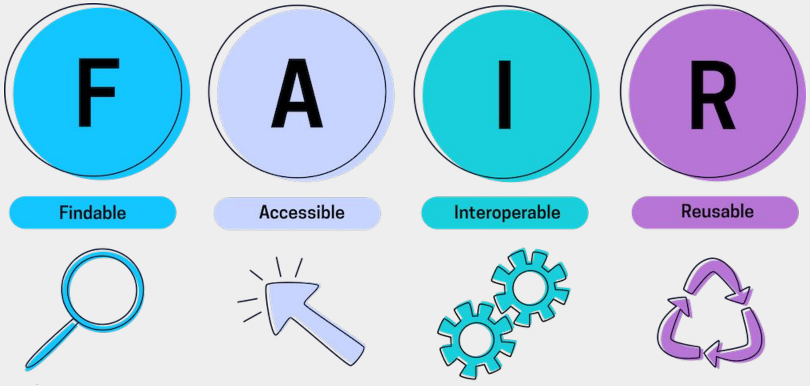
F A I R

Fully Artificial Intelligence Ready

FAIR is absolutely great



FAIR with a pinch of love is even better



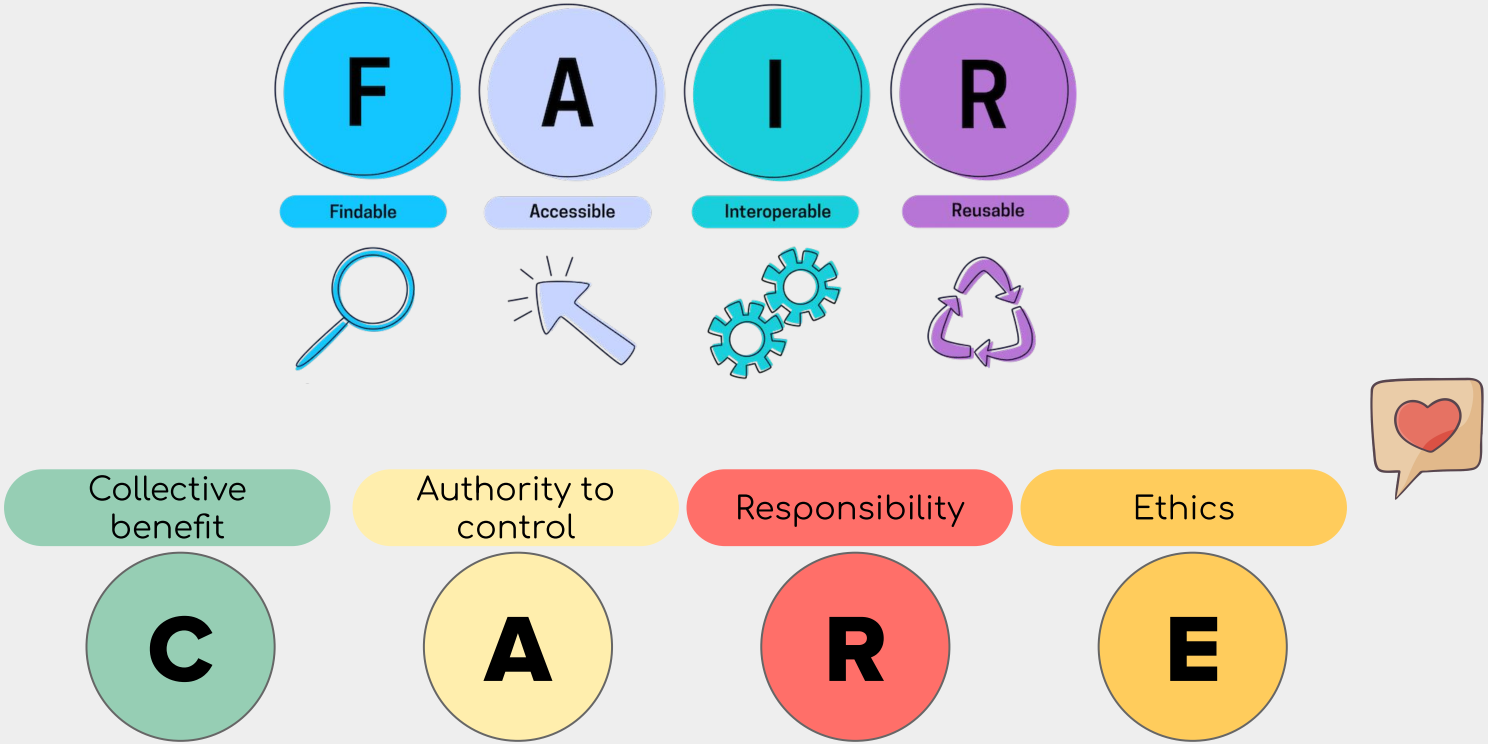
Collective benefit

Authority to control

Responsibility

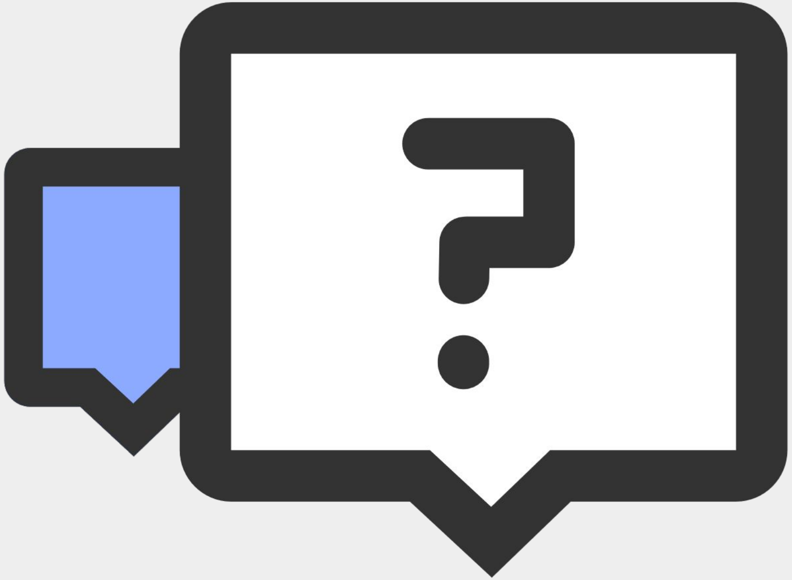
Ethics

from FAIR to CARE



https://www.nature.com/articles/s41597-021-00892-0

Time for Questions



the (rough) agenda

Introduction to research data and FAIR - 12h15-12:45

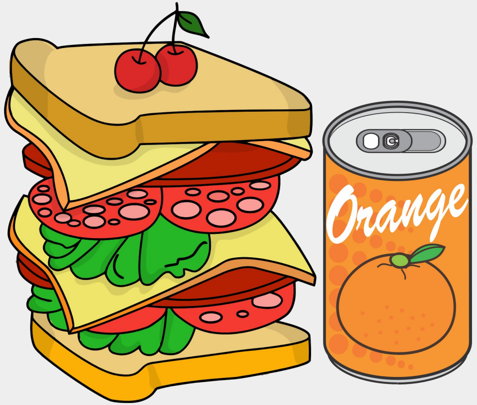
The FAIR principles in action - 13h00-14h00

- F for Findable
- A for Accessible
- I for Interoperable
- R for Reusable

Lunch 14h00-15h00

FAIRify (your) data - 15h00-16h15

TIME FOR LUNCH!



the (rough) agenda

Introduction to research data and FAIR - 12h15-12:45

The FAIR principles in action - 13h00-14h00

F for Findable
A for Accessible
I for Interoperable
R for Reusable

Lunch 14h00-15h00

FAIRify (your) data - 15h00-16h15

BEFORE WE START...
CAN WE DISCUSS THE DATASETS YOU SHARED WITH ME OPENLY IN THE CLASS?

useful links

http://tiny.cc/maribor

http://tiny.cc/maribordata

tabular data

https://allisonhorst.com/

tabular data: messy or tidy?

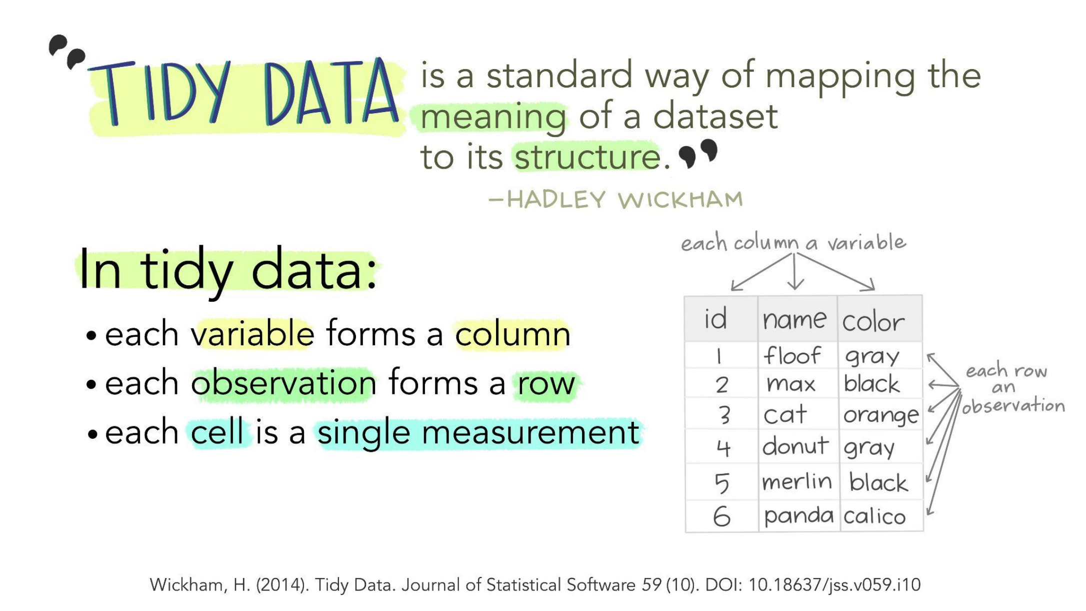
Data is often acquired and represented in various shapes and sizes, but it is most commonly received in the form of data tables (tabular data).

A dataset is messy or tidy depending on how rows, columns and tables are matched up with observations, variables and types.

tabular data: messy or tidy?

Data is often acquired and represented in various shapes and sizes, but it is most commonly received in the form of data tables (tabular data).

A dataset is messy or tidy depending on how rows, columns and tables are matched up with observations, variables and types.



"TIDY DATA is a standard way of mapping the meaning of a dataset to its structure."

—HADLEY WICKHAM

In tidy data:

- each variable forms a column
- each observation forms a row
- each cell is a single measurement

each column a variable

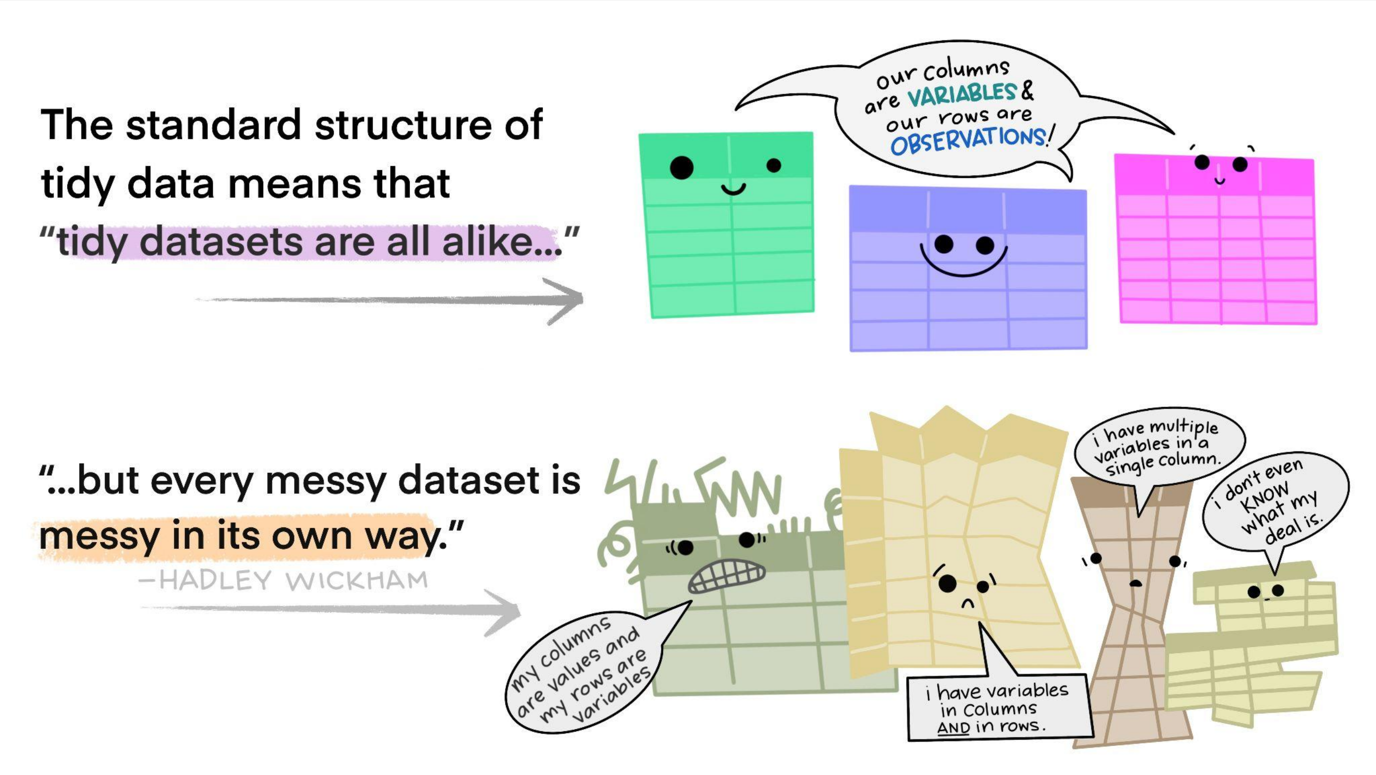
id	name	color
1	floof	gray
2	max	black
3	cat	orange
4	donut	gray
5	merlin	black
6	panda	calico

each row an observation

Wickham, H. (2014). Tidy Data. Journal of Statistical Software 59 (10). DOI: 10.18637/jss.v059.i10

https://www.jstatsoft.org/article/view/v059i10; https://openscapes.org/blog/2020-10-12-tidy-data/

tidy datasets are all alike (and happy!)



The standard structure of tidy data means that "tidy datasets are all alike…"

"…but every messy dataset is messy in its own way."

—HADLEY WICKHAM

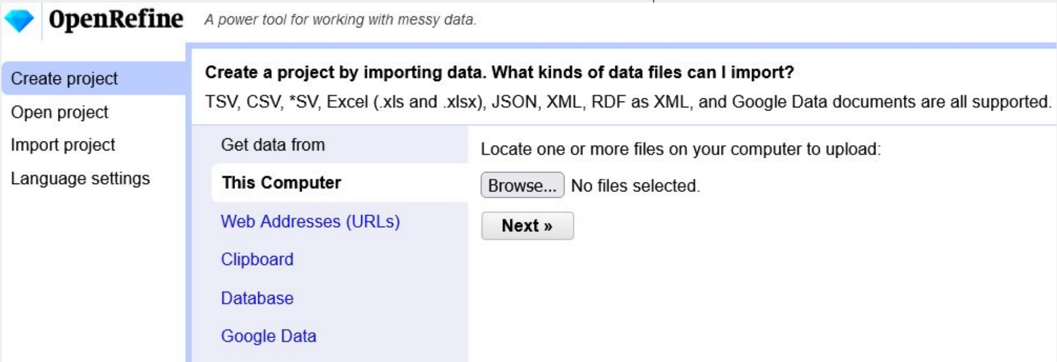
our columns are VARIABLES & our rows are OBSERVATIONS!
my columns are values and my rows are variables
i have variables in columns AND in rows.
i have multiple variables in a single column.
i don't even KNOW what my deal is.

https://openscapes.org/blog/2020-10-12-tidy-data/

what we need: some (toy) data and a tool

http://tiny.cc/maribordata

OpenRefine



A power tool for working with messy data.

Create a project by importing data. What kinds of data files can I import?

TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported.

Locate one or more files on your computer to upload:

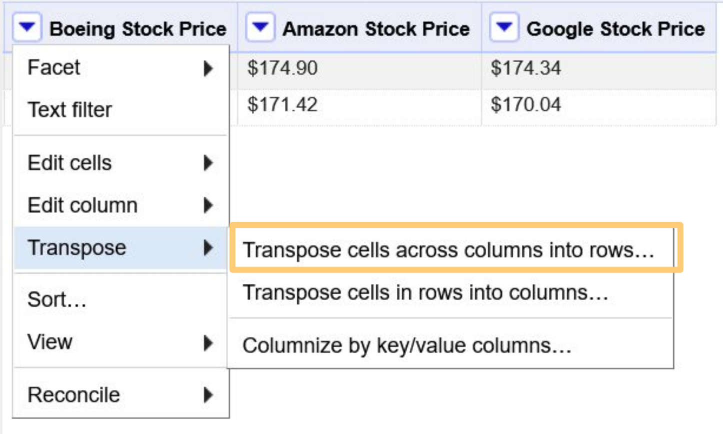
No files selected.

https://openrefine.org/

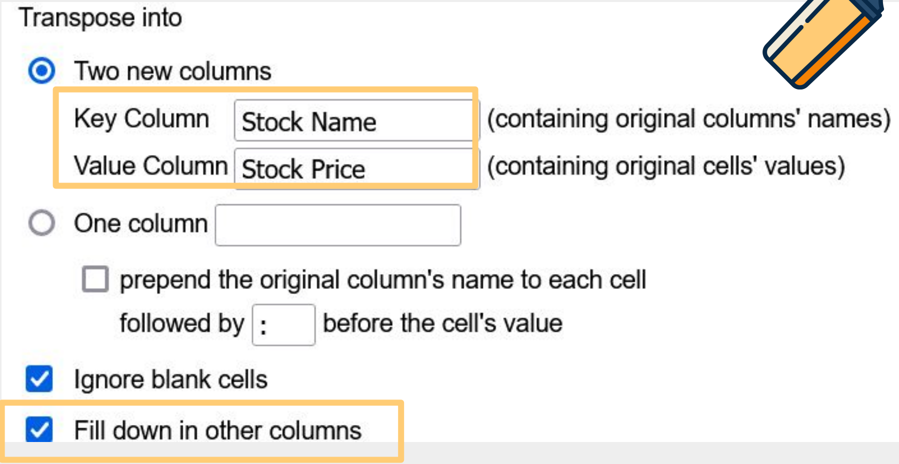
let's tidy up some data

OpenRefine

- ➔ open the untidy1.csv file: why is this data not tidy?
- ➔ let's tidy it up in OpenRefine: launch the software (this opens http://127.0.0.1:3333 in your default browser), import the file, and create a Project (you can leave the default name)
- ➔ select the Boeing Stock Price and follow the steps highlighted below
- ➔ what has happened to the data?



Amazon Stock Price	Google Stock Price
$174.90	$174.34
$171.42	$170.04


Transpose into

● Two new columns

Key Column: Stock Name (containing original columns' names)

Value Column: Stock Price (containing original cells' values)

○ One column

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

☑ Ignore blank cells

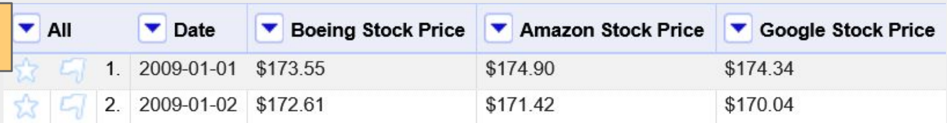
☑ Fill down in other columns

let’s tidy up some data

OpenRefine

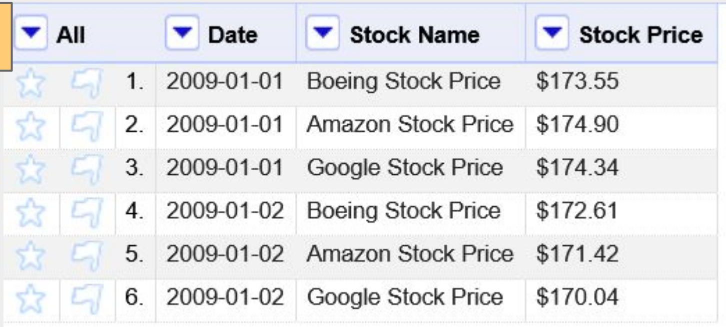
- ➔ remember, in a tidy format, each variable should be its own column
- ➔ while this is not the case in 1, it is the case in 2 (see images below)
- ➔ the last thing to do is to change the stock names: to do this, follow the steps highlighted in 3

1

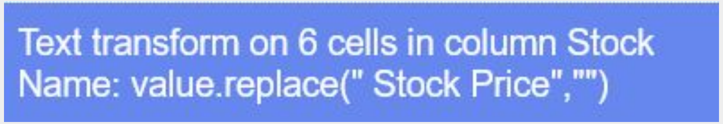


All		Date	Boeing Stock Price	Amazon Stock Price	Google Stock Price
	1.	2009-01-01	$173.55	$174.90	$174.34
	2.	2009-01-02	$172.61	$171.42	$170.04

2

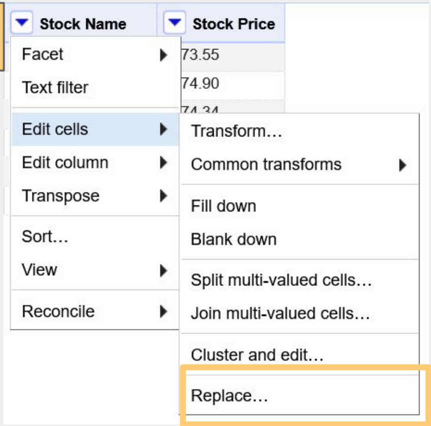


All		Date	Stock Name	Stock Price
	1.	2009-01-01	Boeing Stock Price	$173.55
	2.	2009-01-01	Amazon Stock Price	$174.90
	3.	2009-01-01	Google Stock Price	$174.34
	4.	2009-01-02	Boeing Stock Price	$172.61
	5.	2009-01-02	Amazon Stock Price	$171.42
	6.	2009-01-02	Google Stock Price	$170.04



Text transform on 6 cells in column Stock Name: value.replace(" Stock Price","")

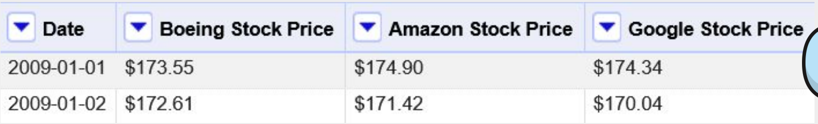
3



our first tidy dataset

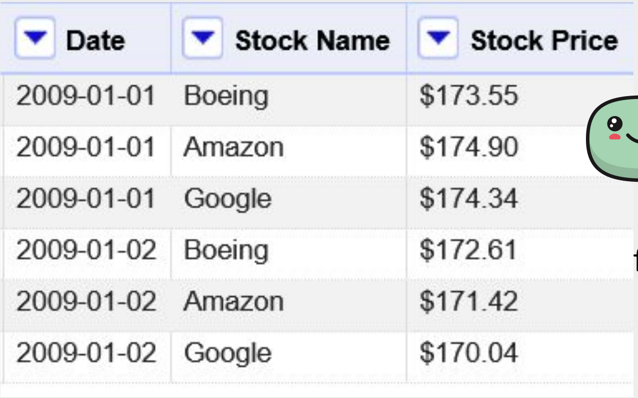
this was a very small dataset (just two rows), and we could have also cleaned it up manually

for larger datasets, however, this would become very inefficient and laborious, so tools like OpenRefine, or R libraries like tidyR, become very powerful



Date	Boeing Stock Price	Amazon Stock Price	Google Stock Price
2009-01-01	$173.55	$174.90	$174.34
2009-01-02	$172.61	$171.42	$170.04

this untidy format is also known as wide format



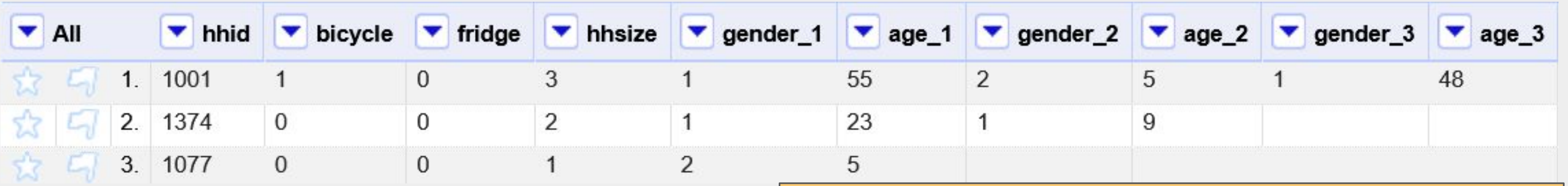
Date	Stock Name	Stock Price
2009-01-01	Boeing	$173.55
2009-01-01	Amazon	$174.90
2009-01-01	Google	$174.34
2009-01-02	Boeing	$172.61
2009-01-02	Amazon	$171.42
2009-01-02	Google	$170.04

this tidy format is also known as long format

https://tidyr.tidyverse.org/; https://byuidatascience.github.io/python4ds/tidy-data.html

a more complex case

- ➔ open now the untidy2.csv file: why is this dataset not tidy?
- ➔ the variables have two different units of observation: household and household member; as a result, we have multiple columns for a single variable (look at the age and gender columns)
- ➔ the trick here is to create two tables, one for each unit of observation
- ➔ open the file in OpenRefine, and perform a transpose operation like explained below



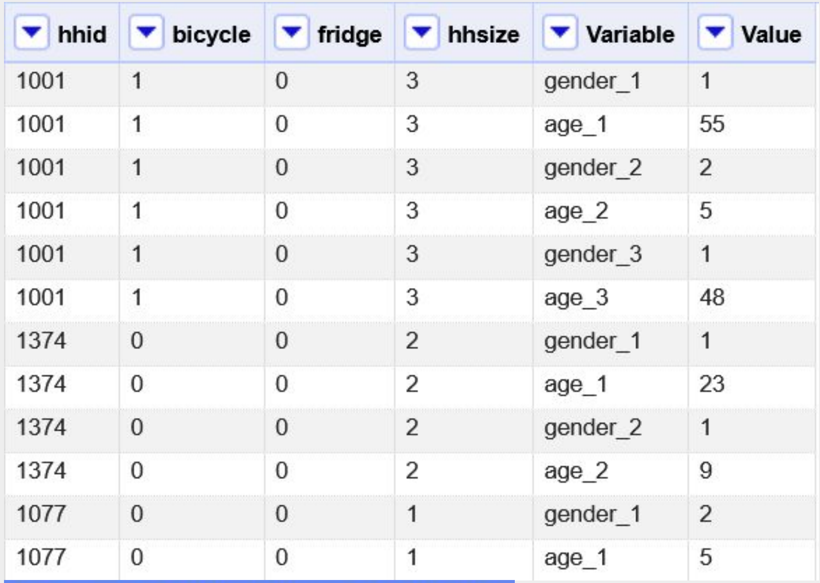
All	hhid	bicycle	fridge	hhsize	gender_1	age_1	gender_2	age_2	gender_3	age_3
1.	1001	1	0	3	1	55	2	5	1	48
2.	1374	0	0	2	1	23	1	9		
3.	1077	0	0	1	2	5				

Transpose cells in columns starting with gender_1 into rows in two new columns named Variable and Value

household 1077 only has one member, so the columns for gender_2, age_2, gender_3, and age_3 are empty (missing values)

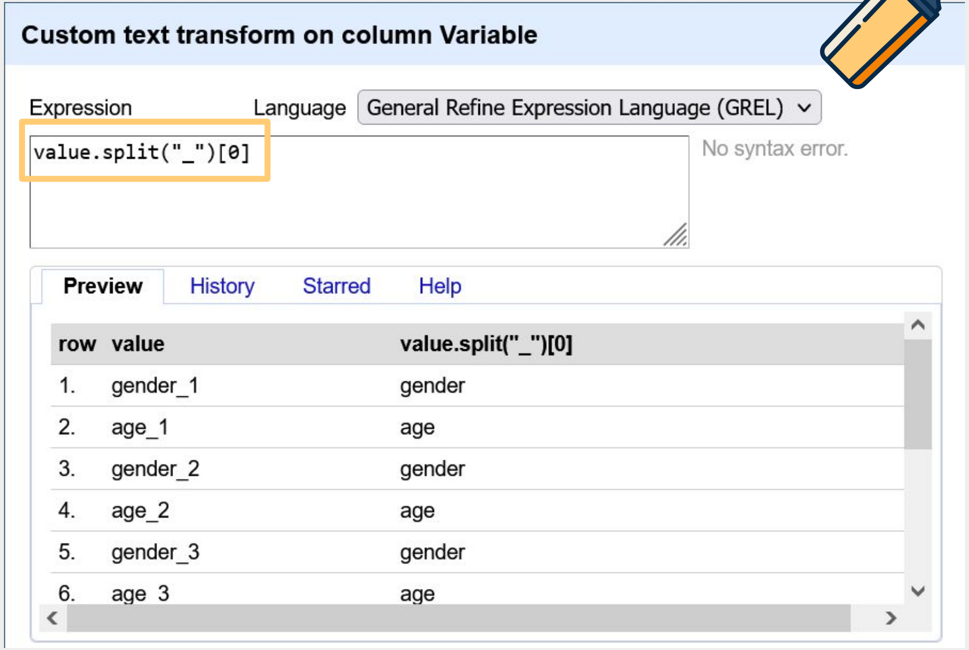
a more complex case

➔ after the transpose, edit the Variable column with the expression below



hhid	bicycle	fridge	hhsize	Variable	Value
1001	1	0	3	gender_1	1
1001	1	0	3	age_1	55
1001	1	0	3	gender_2	2
1001	1	0	3	age_2	5
1001	1	0	3	gender_3	1
1001	1	0	3	age_3	48
1374	0	0	2	gender_1	1
1374	0	0	2	age_1	23
1374	0	0	2	gender_2	1
1374	0	0	2	age_2	9
1077	0	0	1	gender_1	2
1077	0	0	1	age_1	5

Transpose cells in columns starting with gender_1 into rows in two new columns named Variable and Value



Custom text transform on column Variable

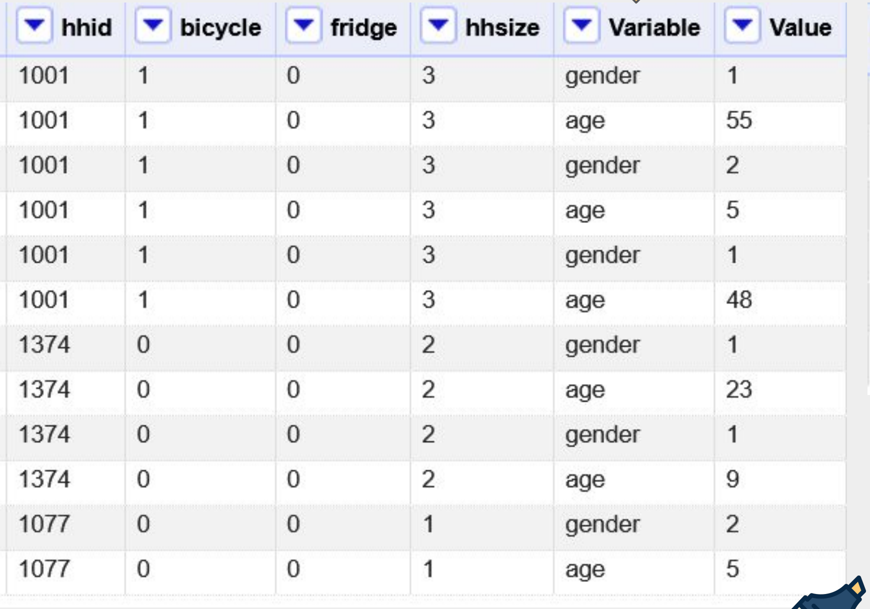
Expression: Language: General Refine Expression Language (GREL)

```
value.split("_")[0]
```

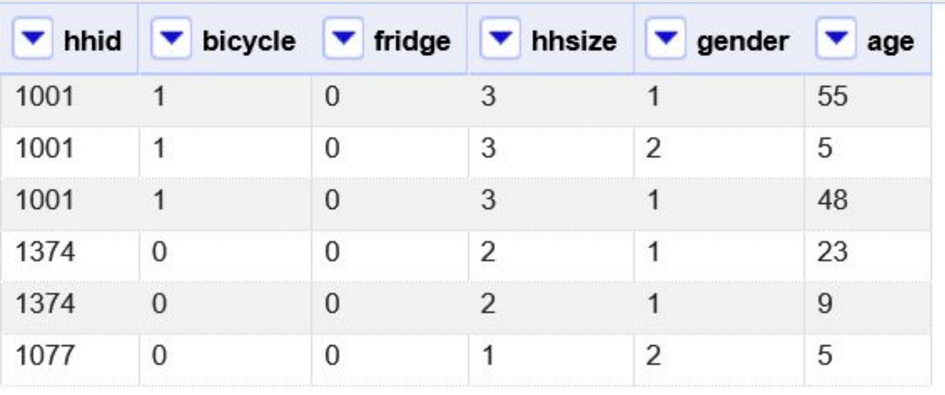
No syntax error.

row	value	value.split("_")[0]
1.	gender_1	gender
2.	age_1	age
3.	gender_2	gender
4.	age_2	age
5.	gender_3	gender
6.	age_3	age

a more complex case



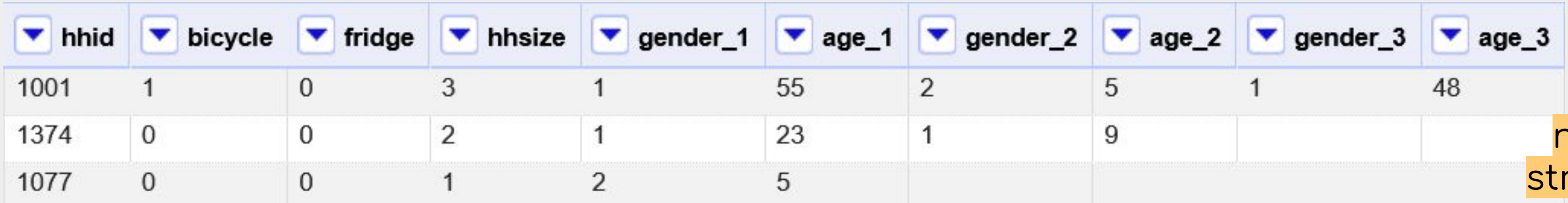
hhid	bicycle	fridge	hhsize	Variable	Value
1001	1	0	3	gender	1
1001	1	0	3	age	55
1001	1	0	3	gender	2
1001	1	0	3	age	5
1001	1	0	3	gender	1
1001	1	0	3	age	48
1374	0	0	2	gender	1
1374	0	0	2	age	23
1374	0	0	2	gender	1
1374	0	0	2	age	9
1077	0	0	1	gender	2
1077	0	0	1	age	5



hhid	bicycle	fridge	hhsize	gender	age
1001	1	0	3	1	55
1001	1	0	3	2	5
1001	1	0	3	1	48
1374	0	0	2	1	23
1374	0	0	2	1	9
1077	0	0	1	2	5

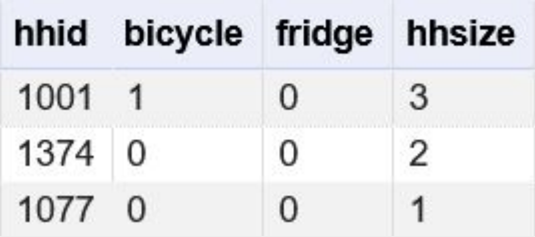
- ➔ next step would be to transpose again so that the variable column becomes two separate columns, gender and age
- ➔ finally, identifiers for the household members need to be created
- ➔ it is key to have identifiers on both side!

from a nested dataset to multiple datasets

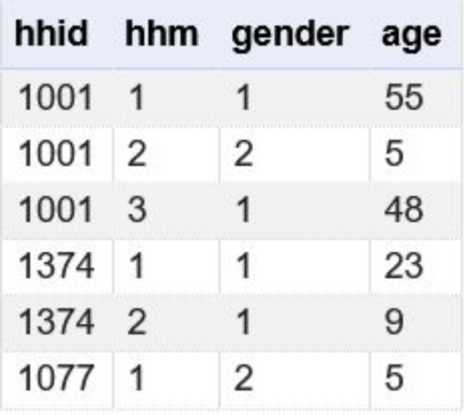


hhid	bicycle	fridge	hhsize	gender_1	age_1	gender_2	age_2	gender_3	age_3
1001	1	0	3	1	55	2	5	1	48
1374	0	0	2	1	23	1	9		
1077	0	0	1	2	5				

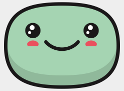
nested structure!



hhid	bicycle	fridge	hhsize
1001	1	0	3
1374	0	0	2
1077	0	0	1

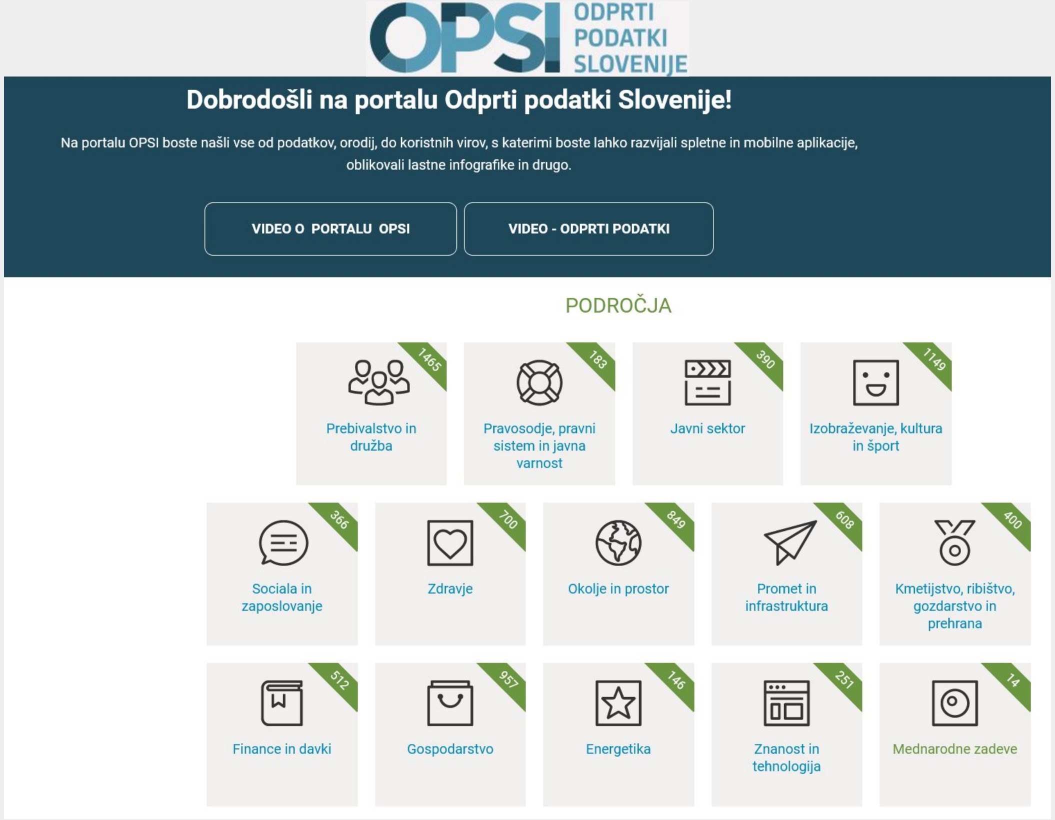


hhid	hhm	gender	age
1001	1	1	55
1001	2	2	5
1001	3	1	48
1374	1	1	23
1374	2	1	9
1077	1	2	5



one table for each unit of observation

the open data portal of Slovenia



OPSI ODPRTI PODATKI SLOVENIJE

Dobrodošli na portalu Odprti podatki Slovenije!

Na portalu OPSI boste našli vse od podatkov, orodij, do koristnih virov, s katerimi boste lahko razvijali spletne in mobilne aplikacije, oblikovali lastne infografike in drugo.

VIDEO O PORTALU OPSI | VIDEO - ODPRTI PODATKI

PODROČJA

- Prebivalstvo in družba 1465
- Pravosodje, pravni sistem in javna varnost 183
- Javni sektor 390
- Izobraževanje, kultura in šport 1149
- Sociala in zaposlovanje 366
- Zdravje 700
- Okolje in prostor 849
- Promet in infrastruktura 608
- Kmetijstvo, ribištvo, gozdarstvo in prehrana 400
- Finance in davki 512
- Gospodarstvo 957
- Energetika 146
- Znanost in tehnologija 251
- Mednarodne zadeve 14

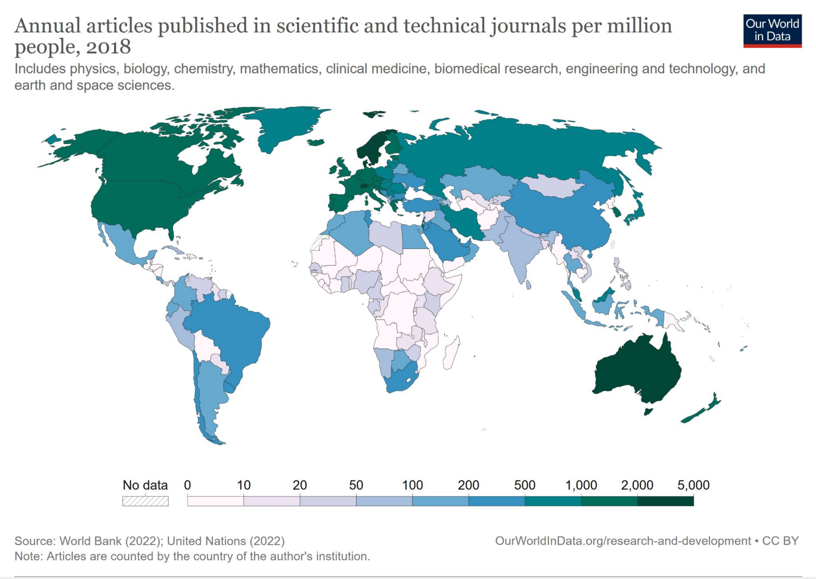
https://podatki.gov.si/

our world in data

Research and data to make progress against the world's largest problems

3653 charts across 297 topics

All free: open access and open source



Annual articles published in scientific and technical journals per million people, 2018

Includes physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences.

Source: World Bank (2022); United Nations (2022)

OurWorldInData.org/research-and-development • CC BY

Note: Articles are counted by the country of the author's institution.

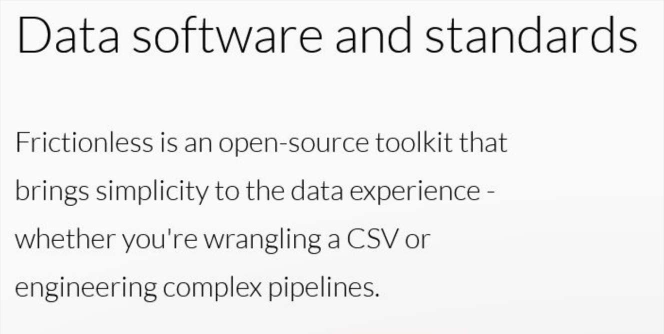
- ➔ we will now work with the scientific-publications-per-million.csv file
- ➔ this data contains "scientific and technical journal articles per million people" from 2000 to 2018
- ➔ the data can be downloaded from Our world in data, but I have already put it in our data directory
- ➔ we will annotate this dataset and publish it on the web

https://ourworldindata.org; https://ourworldindata.org/grapher/scientific-publications-per-million?time=2018 http://tiny.cc/maribordata

let’s discuss the datasets together: we can use the collaborative pad to write down some questions and answers, or any thoughts!

Frictionless data

FRICTIONLESS DATA
STANDARDS AND TOOLING



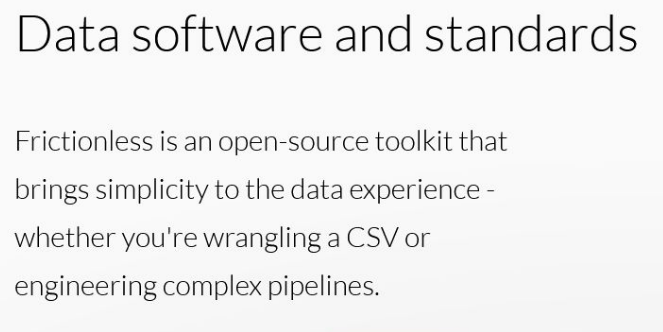
Data software and standards

Frictionless is an open-source toolkit that brings simplicity to the data experience - whether you're wrangling a CSV or engineering complex pipelines.

https://frictionlessdata.io/; https://frictionlessdata.io/introduction/

Frictionless data & data packages

FRICTIONLESS DATA
STANDARDS AND TOOLING



Data software and standards

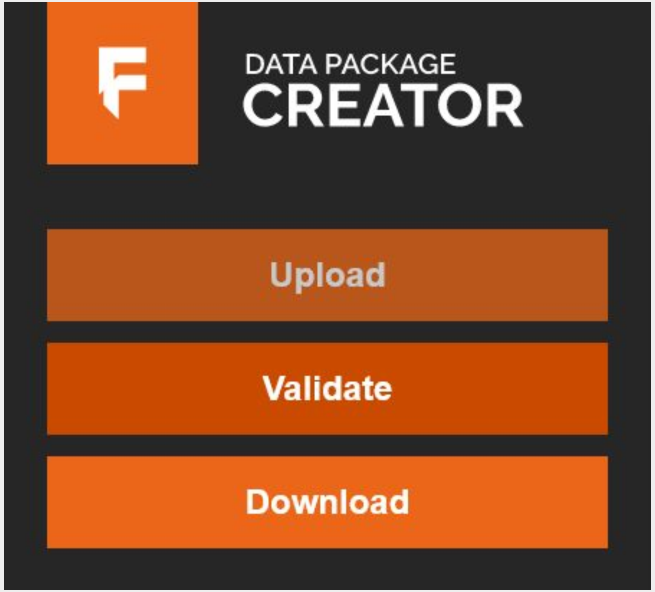
Frictionless is an open-source toolkit that brings simplicity to the data experience - whether you're wrangling a CSV or engineering complex pipelines.

Data Package is a format that makes it possible to put data and relevant information that provides context about it, in one container before you share it

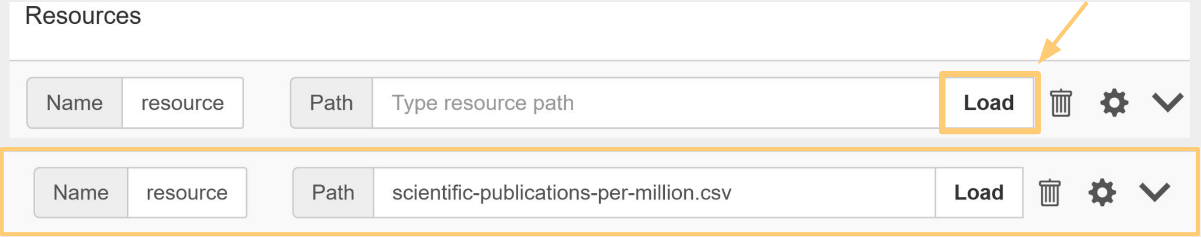
all contextual information, like the metadata and the data schema, is published in a JSON file named *datapackage.json*

https://specs.frictionlessdata.io/data-package/

the data package creator



- ➔ we'll use the Data Package Creator, an online service that facilitates the creation and editing of data packages
- ➔ go to: https://create.frictionlessdata.io/
- ➔ there are several ways to create a data package - if your data resource is publicly available, like on GitHub/Gitlab or in a data repository, you can obtain the URL and paste it in the Path section
- ➔ we will load a resource from our CSV file

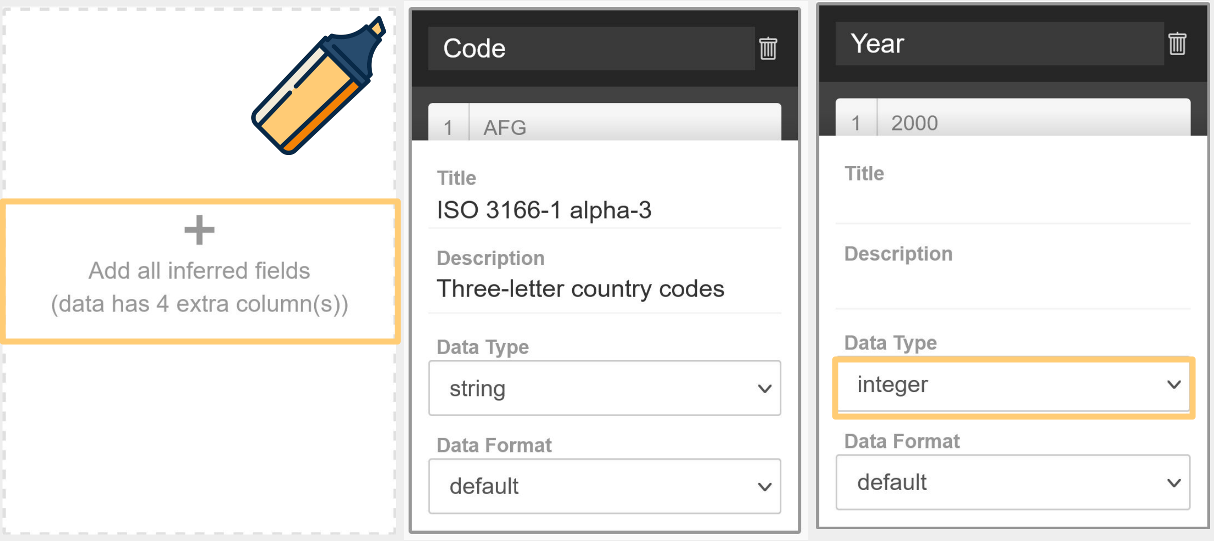


Resources

Name: resource | Path: Type resource path | Load

Name: resource | Path: scientific-publications-per-million.csv | Load

annotate the data



+ Add all inferred fields
(data has 4 extra column(s))

Code

1 AFG

Title
ISO 3166-1 alpha-3

Description
Three-letter country codes

Data Type
string

Data Format
default

Year

1 2000

Title

Description

Data Type
integer

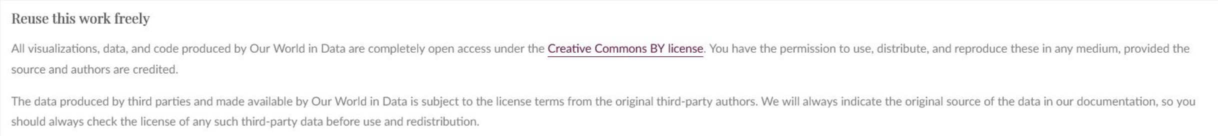
Data Format
default

- the tool is relatively smart, so it will automatically infer the fields of your data
- however, it is still up to you to add Title and Description to the columns, and to make sure that the Data Types are inferred correctly
- check the data type of the column Year: is it correct?

https://en.wikipedia.org/wiki/ISO_3166-1_alpha-3

take care of the metadata

- ➔ add metadata to the dataset
  name, title, description, author, license
- ➔ because this is data already published somewhere else, we need to make sure we respect the original terms and conditions

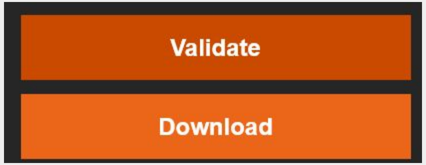


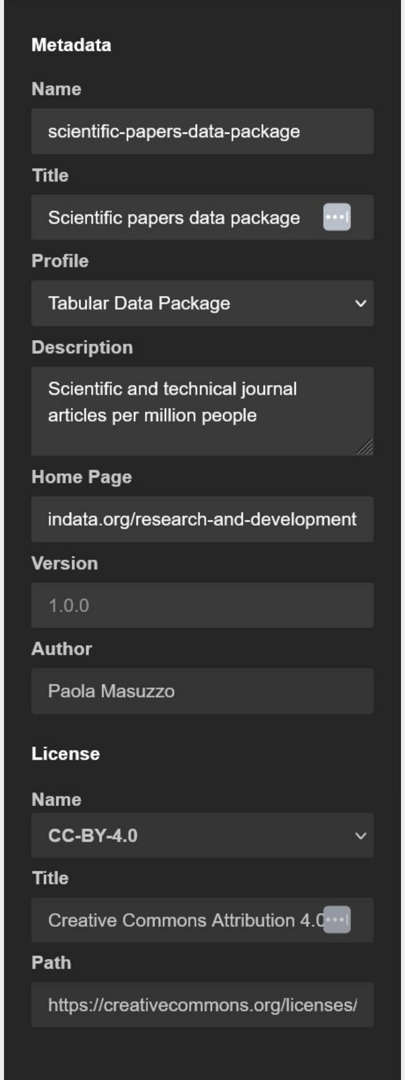
Reuse this work freely

All visualizations, data, and code produced by Our World in Data are completely open access under the Creative Commons BY license. You have the permission to use, distribute, and reproduce these in any medium, provided the source and authors are credited.

The data produced by third parties and made available by Our World in Data is subject to the license terms from the original third-party authors. We will always indicate the original source of the data in our documentation, so you should always check the license of any such third-party data before use and redistribution.

- ➔ we are then finally ready to validate and download the data package!
- ➔ the package is a *json* file which you can then open in any text editor, or, even better, in your web browser





Metadata

Name: scientific-papers-data-package

Title: Scientific papers data package

Profile: Tabular Data Package

Description: Scientific and technical journal articles per million people

Home Page: indata.org/research-and-development

Version: 1.0.0

Author: Paola Masuzzo

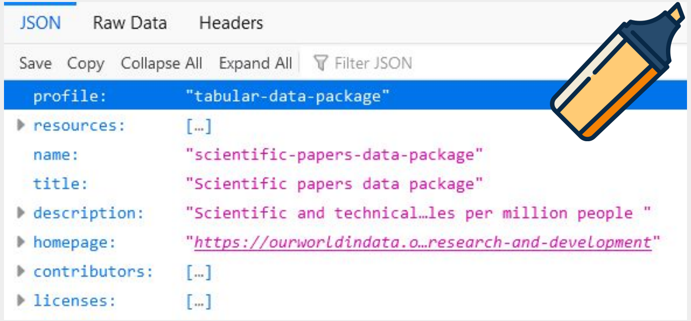
License

Name: CC-BY-4.0

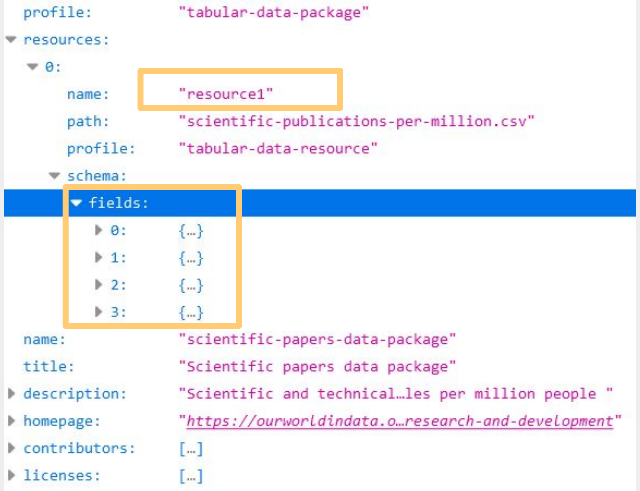
Title: Creative Commons Attribution 4.0

Path: https://creativecommons.org/licenses/

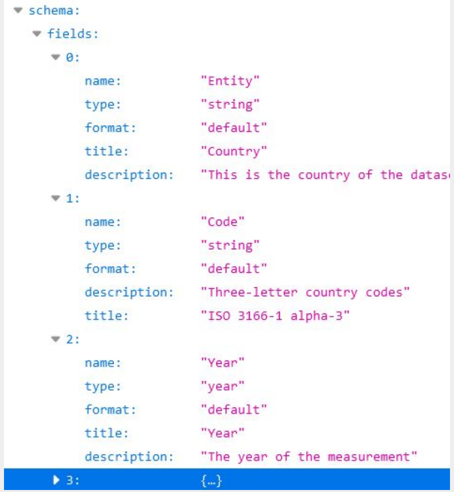
the data package



```
profile:        "tabular-data-package"
▸ resources:    […]
name:           "scientific-papers-data-package"
title:          "Scientific papers data package"
▸ description:  "Scientific and technical…les per million people "
▸ homepage:     "https://ourworldindata.o…research-and-development"
▸ contributors: […]
▸ licenses:     […]
```



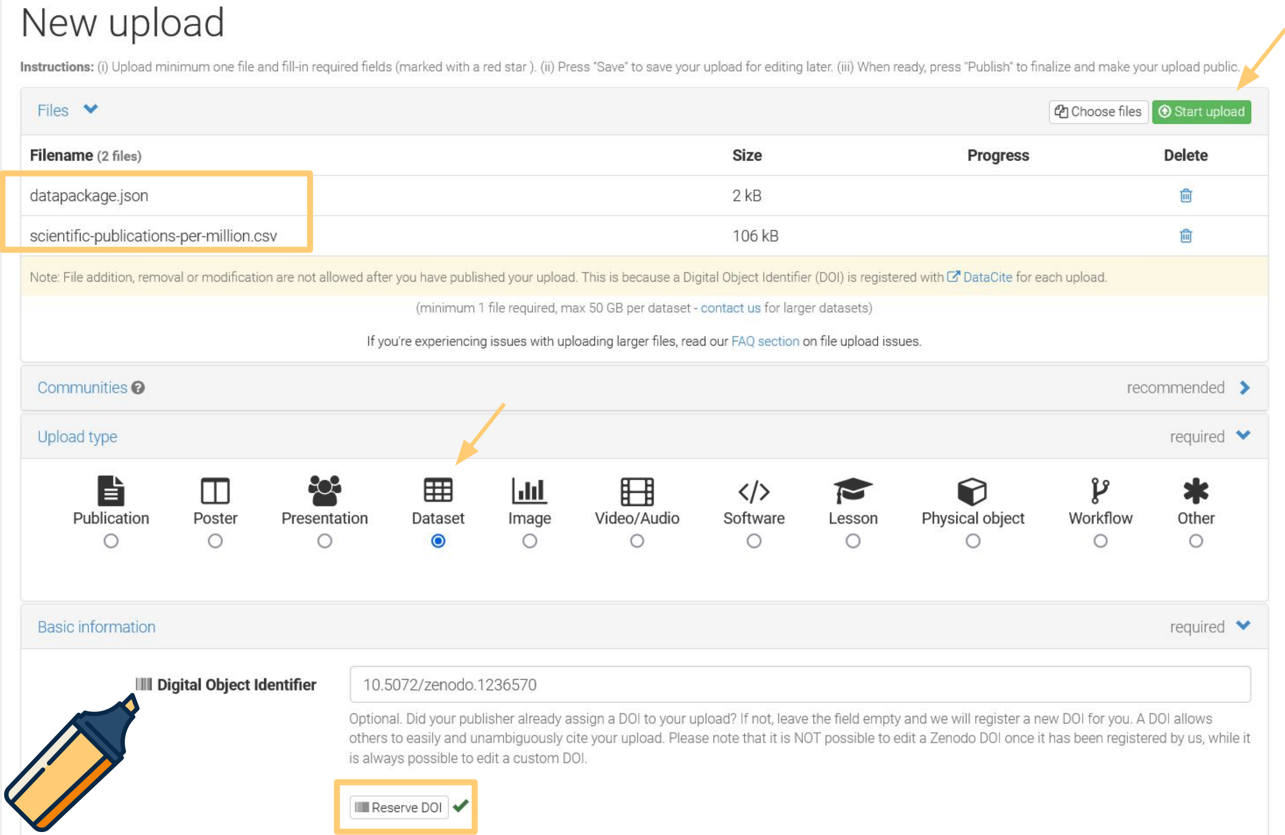
```
  profile:           "tabular-data-package"
▾ resources:
  ▾ 0:
      name:          "resource1"
      path:          "scientific-publications-per-million.csv"
      profile:       "tabular-data-resource"
    ▾ schema:
      ▾ fields:
        ▸ 0:         {…}
        ▸ 1:         {…}
        ▸ 2:         {…}
        ▸ 3:         {…}
  name:              "scientific-papers-data-package"
  title:             "Scientific papers data package"
▸ description:       "Scientific and technical…les per million people "
▸ homepage:          "https://ourworldindata.o…research-and-development"
▸ contributors:      […]
▸ licenses:          […]
```



```
▾ schema:
  ▾ fields:
    ▾ 0:
        name:          "Entity"
        type:          "string"
        format:        "default"
        title:         "Country"
        description:   "This is the country of the datas
    ▾ 1:
        name:          "Code"
        type:          "string"
        format:        "default"
        description:   "Three-letter country codes"
        title:         "ISO 3166-1 alpha-3"
    ▾ 2:
        name:          "Year"
        type:          "year"
        format:        "default"
        title:         "Year"
        description:   "The year of the measurement"
    ▸ 3:               {…}
```

- ➔ the data package contains 1 resource (the CSV file with our data), and this resource has a well-defined schema with 4 fields
- ➔ when we will publish this dataset on the web, we will publish both the data (the CSV) file, and this *json* file, which acts as a data dictionary
- ➔ this will help other people understand the data, will aid machine readability, and overall make reuse and repurpose of the data much easier

publish the data package



New upload

Instructions: (i) Upload minimum one file and fill-in required fields (marked with a red star). (ii) Press "Save" to save your upload for editing later. (iii) When ready, press "Publish" to finalize and make your upload public.

Files

Filename (2 files)	Size	Progress	Delete
datapackage.json	2 kB		🗑
scientific-publications-per-million.csv	106 kB		🗑

Note: File addition, removal or modification are not allowed after you have published your upload. This is because a Digital Object Identifier (DOI) is registered with DataCite for each upload.

(minimum 1 file required, max 50 GB per dataset - contact us for larger datasets)

If you're experiencing issues with uploading larger files, read our FAQ section on file upload issues.

Communities — recommended

Upload type — required

Publication ○ Poster ○ Presentation ○ Dataset ● Image ○ Video/Audio ○ Software ○ Lesson ○ Physical object ○ Workflow ○ Other ○

Basic information — required

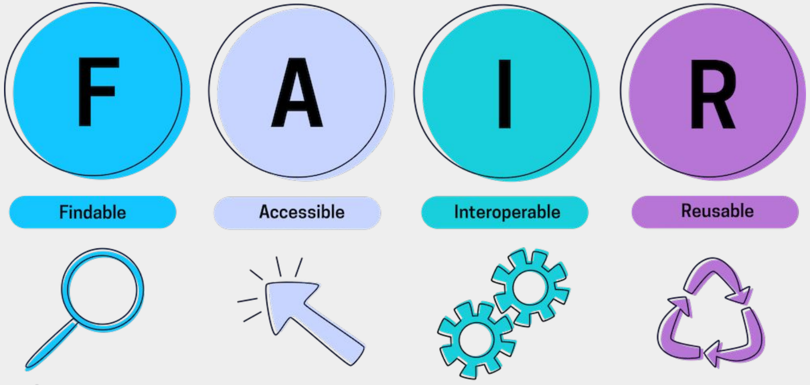
Digital Object Identifier: 10.5072/zenodo.1236570

Optional. Did your publisher already assign a DOI to your upload? If not, leave the field empty and we will register a new DOI for you. A DOI allows others to easily and unambiguously cite your upload. Please note that it is NOT possible to edit a Zenodo DOI once it has been registered by us, while it is always possible to edit a custom DOI.

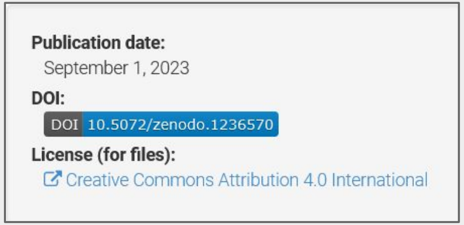
- ➔ let's publish the dataset on the web
- ➔ go to https://sandbox.zenodo.org/ and sign-in
- ➔ choose both files and start the upload
- ➔ select Dataset as upload type, and don't forget to reserve a DOI!
- ➔ go ahead and fill in the other required fields
- ➔ save your upload, ad finally publish it!

CONGRATULATIONS!

a FAIR dataset



a FAIR dataset

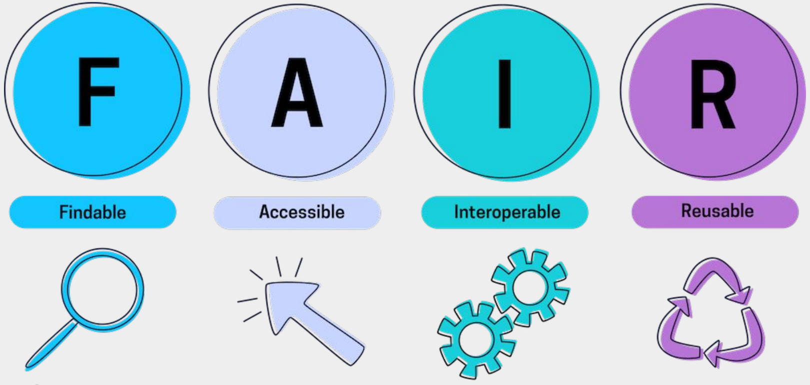


Publication date:
September 1, 2023

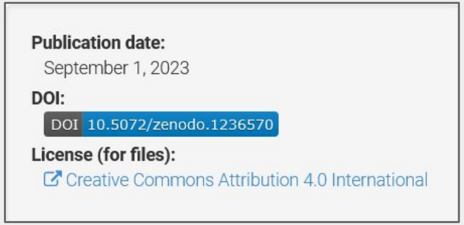
DOI:
DOI 10.5072/zenodo.1236570

License (for files):
Creative Commons Attribution 4.0 International

deposit on Zenodo



a FAIR dataset



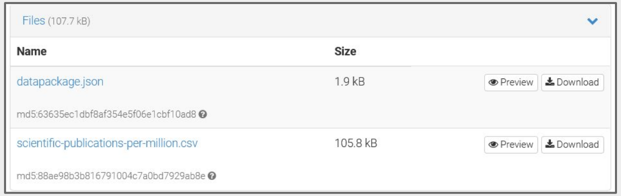
Publication date:
September 1, 2023

DOI:
DOI 10.5072/zenodo.1236570

License (for files):
Creative Commons Attribution 4.0 International

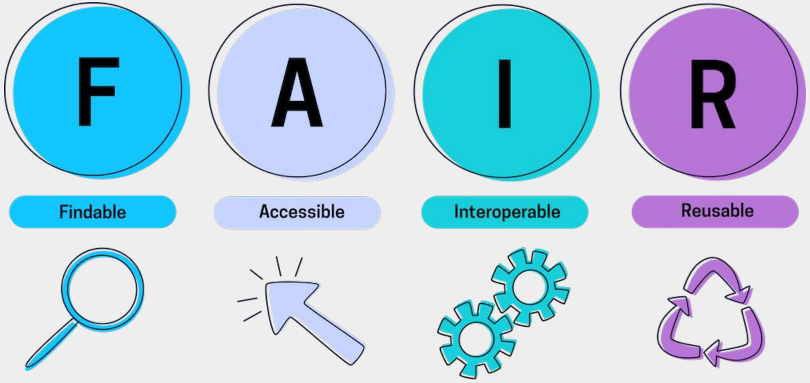
deposit on Zenodo

download from Zenodo



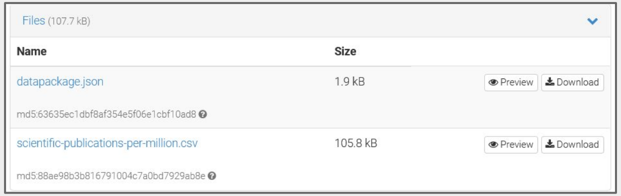
Files (107.7 kB)

Name	Size	
datapackage.json md5:63635ec1dbf8af354e5f06e1cbf10ad8	1.9 kB	Preview Download
scientific-publications-per-million.csv md5:88ae98b3b816791004c7a0bd7929ab8e	105.8 kB	Preview Download



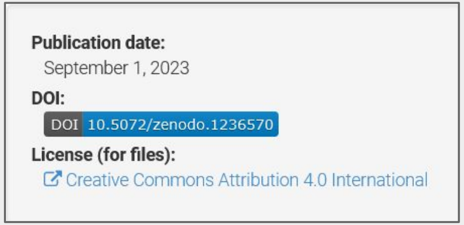
a FAIR dataset

download from Zenodo



Files (107.7 kB)

Name	Size
datapackage.json md5:63635ec1dbf8af354e5f06e1cbf10ad8	1.9 kB
scientific-publications-per-million.csv md5:88ae98b3b816791004c7a0bd7929ab8e	105.8 kB

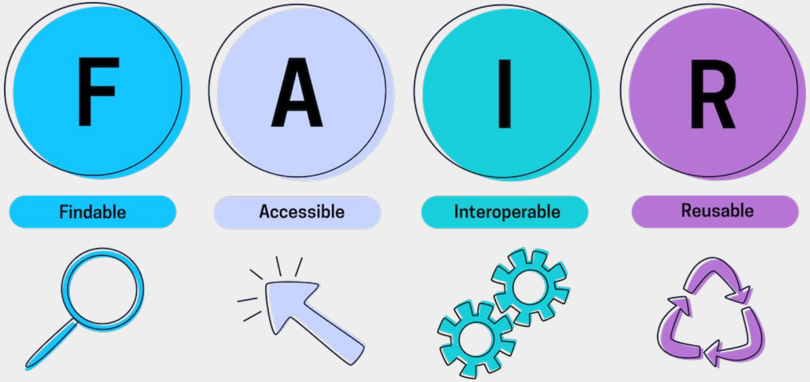


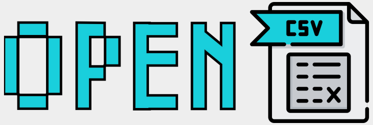
Publication date:
September 1, 2023

DOI:
DOI 10.5072/zenodo.1236570

License (for files):
Creative Commons Attribution 4.0 International

deposit on Zenodo

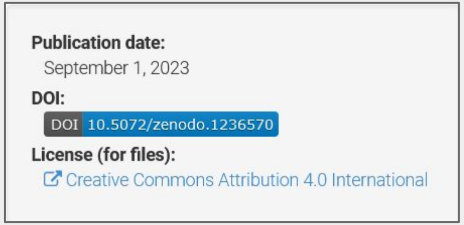




open format (csv)

use of ISO codes

a FAIR dataset



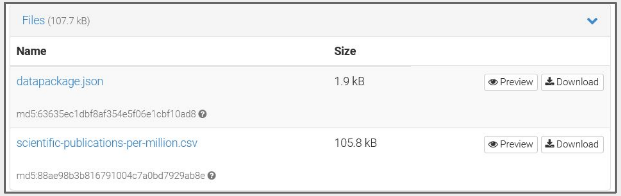
Publication date:
September 1, 2023

DOI:
DOI 10.5072/zenodo.1236570

License (for files):
Creative Commons Attribution 4.0 International

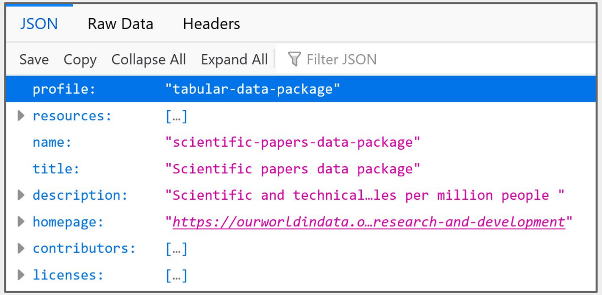
deposit on Zenodo

download from Zenodo

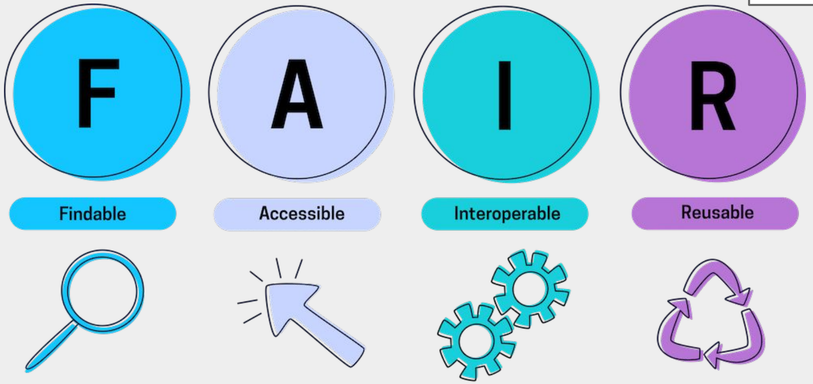


Files (107.7 kB)

Name	Size	
datapackage.json md5:63635ec1dbf8af354e5f06e1cbf10ad8	1.9 kB	Preview Download
scientific-publications-per-million.csv md5:88ae98b3b816791004c7a0bd7929ab8e	105.8 kB	Preview Download



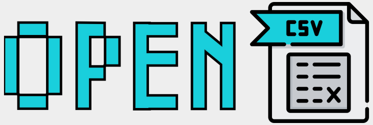
```
profile:        "tabular-data-package"
resources:      […]
name:           "scientific-papers-data-package"
title:          "Scientific papers data package"
description:    "Scientific and technical…les per million people "
homepage:       "https://ourworldindata.o…research-and-development"
contributors:   […]
licenses:       […]
```



clear license

data dictionary

description



open format (csv)

use of ISO codes

time for a little survey, please!

go to menti.com and use the code: **1962 9620**

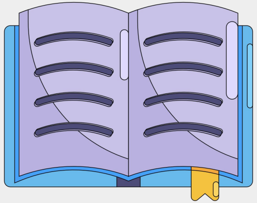
https://www.menti.com/al12w1gyzjxe

☑ ★★★

□ ★★

□

resources



How To FAIR

FAIR Cookbook

Top 10 FAIR Data & Software Things

FAIR training resources | FAIR Data 101

PARTHENOS Guidelines to FAIRify data management and make data reusable

The Turing Way

How to make your data FAIR

FAIRsFAIR

FAIRsharing

CARE Principles — Global Indigenous Data Alliance

https://twitter.com/hashtag/DataHorrorWeek

vectors and icons from https://www.svgrepo.com/

THANK YOU!

Questions?
You can always email me at
paola.masuzzo@gmail.com