20240628
10.5281/zenodo.11862672

Level 2, presentation 2

Esther Plomp & Bosun Obileye

Sharing Open Data

10.5281/zenodo.11862672

Bosun Obileye
Engagement Specialist

Esther Plomp
Data Steward

@ToothFAIRy Mastodon
EstherPlomp LinkedIn
e.plomp@tudelft.nl

TRAINERS

Agenda

FAIR/CARE (Esther)

What repository (Bosun)

File formats (Bosun)

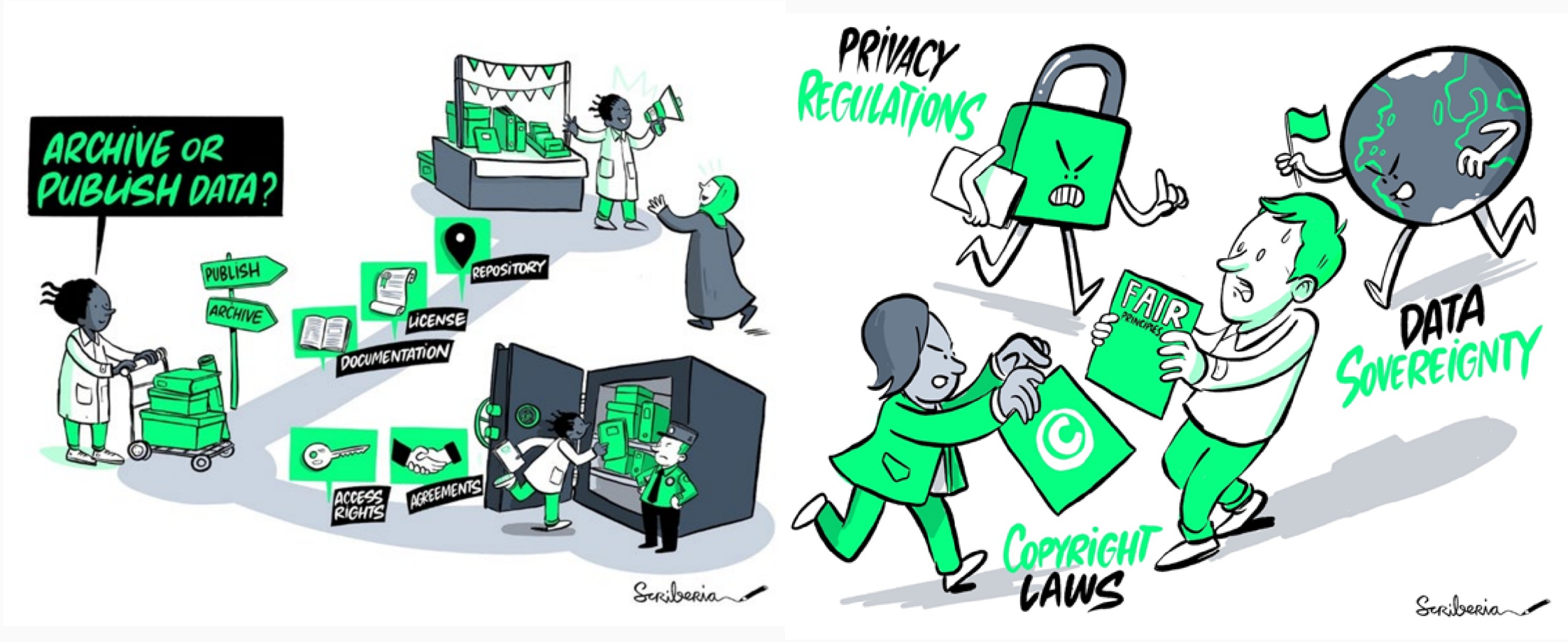
Reusable (Esther):

- Licenses
- Documentation

Data citation/linking research outputs (Esther)

10.5281/zenodo.11862672

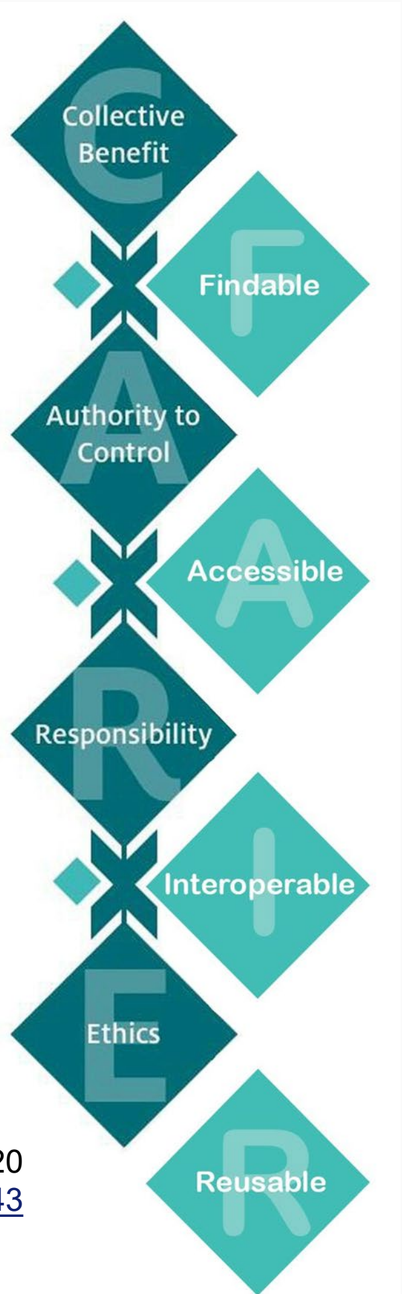
SHOULD YOU SHARE?



10.5281/zenodo.11862672

CARE

Carroll et al. 2020
https://doi.org/10.5334/dsj-2020-043



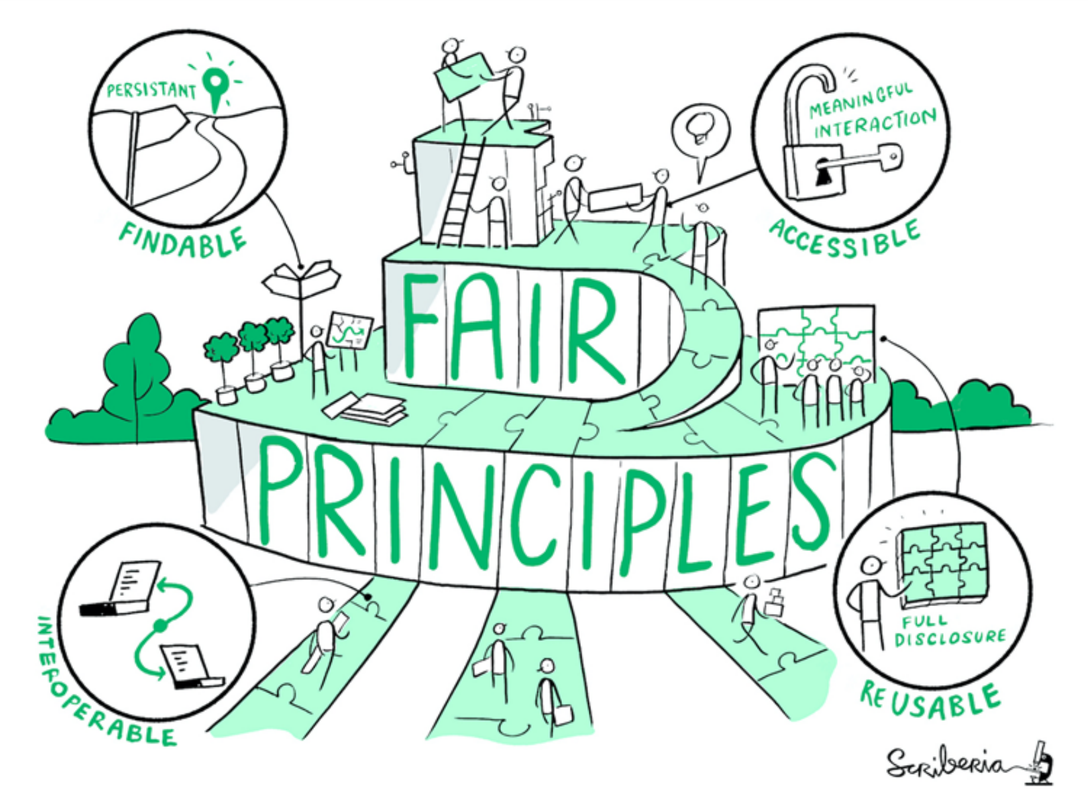
Adapted from Carroll et al. 2020
https://doi.org/10.5334/dsj-2020-043

FAIR

The Turing Way: FAIR

10.5281/zenodo.11862672

FAIR



F

Findable

Descriptive metadata and persistent identifier

A

Accessible

Data could be openly available OR authentication and authorisation procedures are necessary

I

Interoperable

Data needs to be integrated with other data and interoperate with applications or workflows

R

Reusable

Documentation and license

10.5281/zenodo.11862672

DATA REPOSITORIES

The Turing Way - selecting an appropriate repository

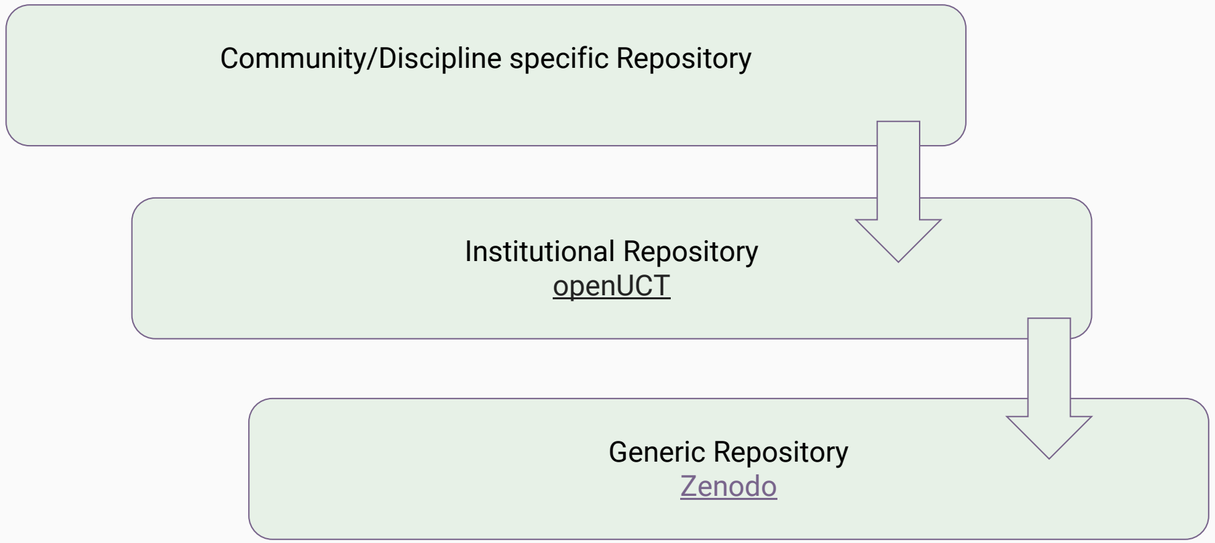
Data should be submitted to **domain or discipline specific**, community recognised, repository where possible.

- Discipline specific data repositories are likely to have more functionalities for the type of data that you would like to share
- Community standards make the data more FAIR!

A **general purpose repository** can be used when there are no suitable discipline specific repositories.

Most of them assign **persistent identifiers** for data to become findable

May place requirements on **file formats** shared



With Open Data the data repository is responsible for long term preservation and access!

10.5281/zenodo.11862672

DATA REPOSITORIES EXAMPLES

Zenodo

General repository for data, code, presentations, reports, training materials
Uploads up to 50 GB

Figshare

General repository
Uploads up to 20 GB

Harvard Dataverse

General repository
Uploads up to 1 TB

Dryad

General repository
Uploads up to 300 GB

AfricArXiv

For African research that provides a platform for African scientists
Free of cost

And more..

Find more using FAIRsharing or re3data

10.5281/zenodo.11862672

Exercise 1

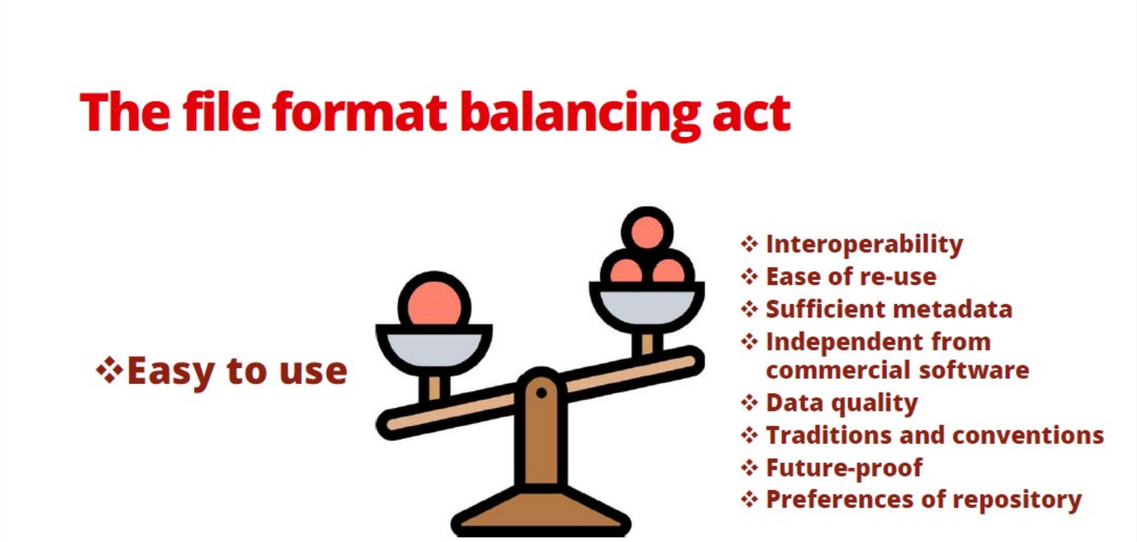
What data repository do you plan to use?

Or what type of repository have you already used?

10.5281/zenodo.11862672

FILE FORMATS

- **Proprietary data formats** often are not readable without the corresponding (commercial) software and may become obsolete in the future. Example include Word (.doc, .docx, etc), Excel (.xls, xlsx,etc), Wave for audio (.wmv)
- **Open Data Format**: *“a freely available published specification which places no restrictions, monetary or otherwise, upon its use”*. - https://opendefinition.org/ofd Examples include .txt, .jpg, .mp3
- **Conversion** from proprietary to open formats is often possible, but may result in some loss of data



https://zenodo.org/doi/10.5281/zenodo.10013234

10.5281/zenodo.11862672

Exercise 2

What file formats will you share the data in?

Are these formats proprietary or open? Why?

10.5281/zenodo.11862672

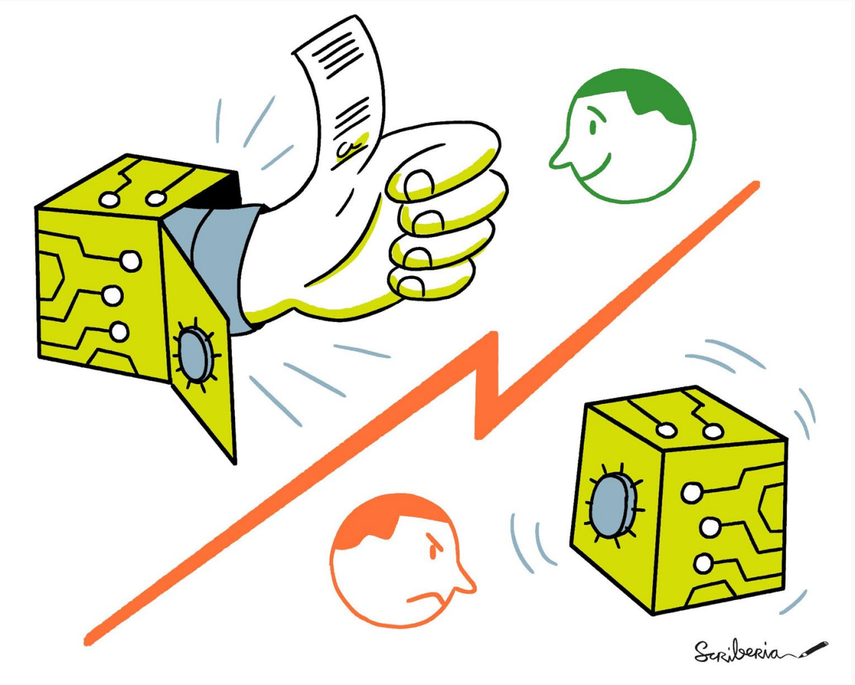
REUSABLE

- License
- Documentation

Also helps if you follow recommended research data management practices!

LICENSES

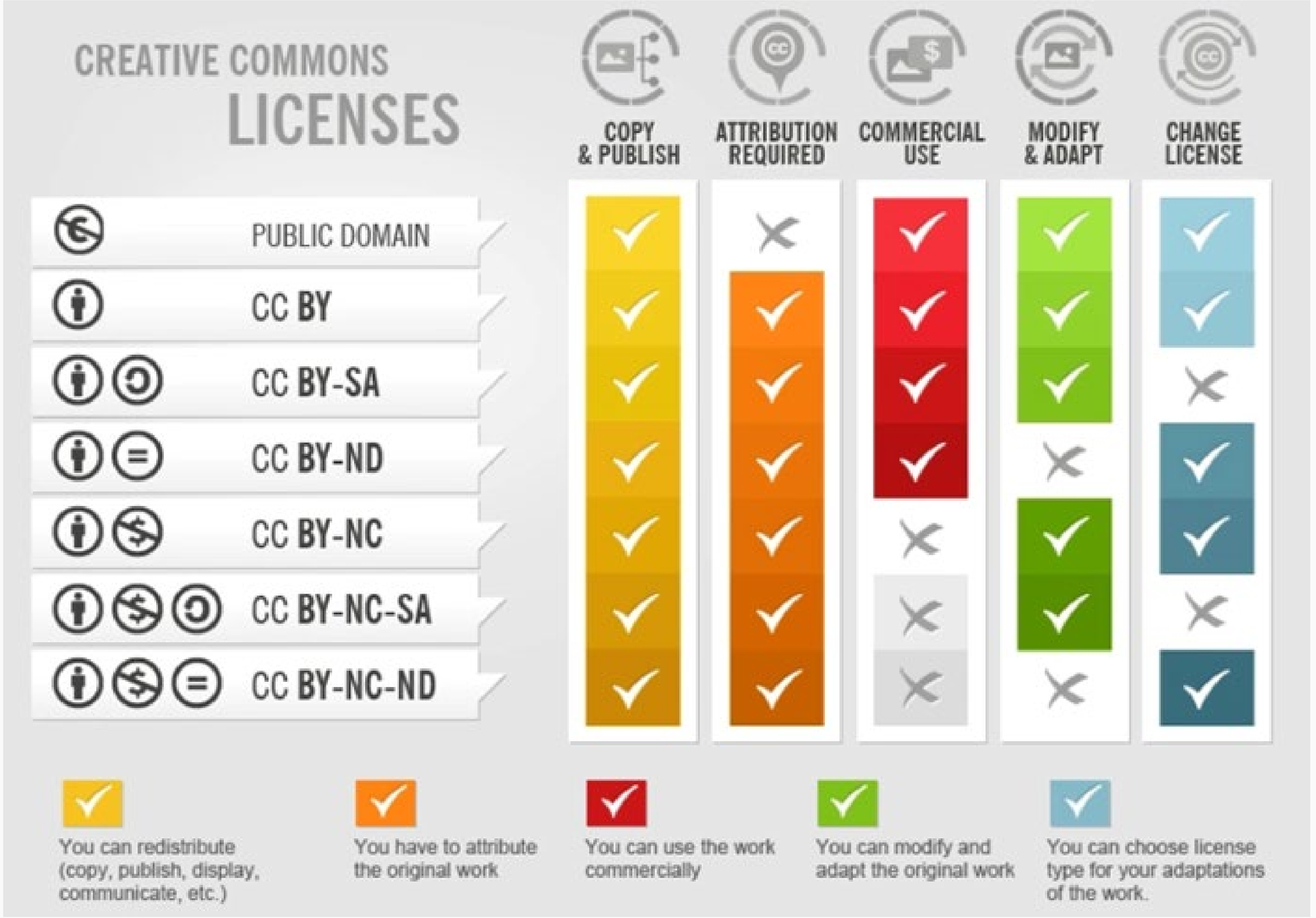
- Formalised agreement of **what reusers can do** with the data/software
- If there is no license it doesn't mean it can be reused - the opposite!
- Software and data have different licenses
- Traditional Knowledge Licenses



10.5281/zenodo.11862672

LICENSES - DATA

Creative Commons License Chooser



CC-BY-SA
https://foter.com/blog/how-to-attribute-creative-commons-photos

10.5281/zenodo.11862672

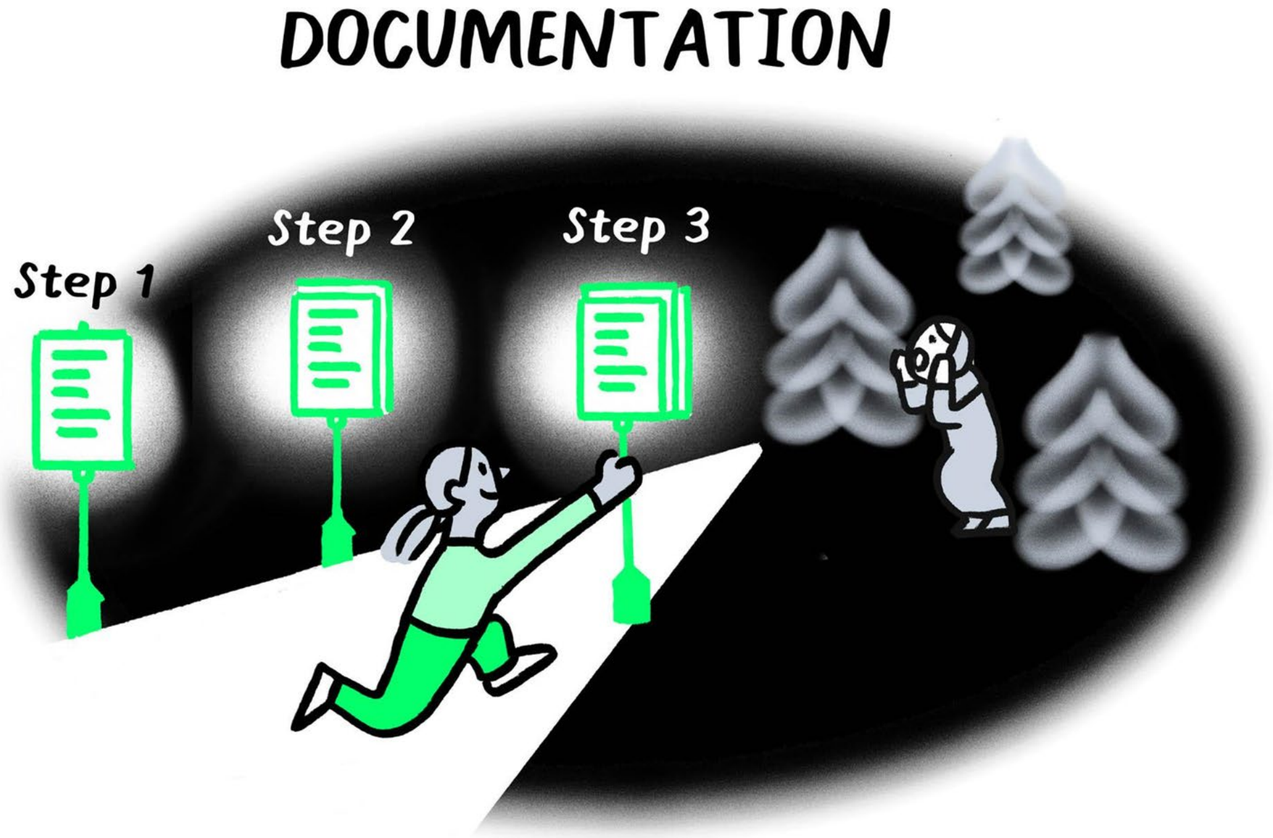
Exercise 3

What data license do you plan to share the data under? (see exercise level 1)

10.5281/zenodo.11862672

DOCUMENTATION

Documentation provides context for your work. It allows your collaborators, colleagues and future you to understand **what has been done and why**.



10.5281/zenodo.11862672

README

Write it in an **open format** such as .txt or .md (Markdown)

Make it clear **what the README file is documenting** (also add this to the README file):

- Project documentation: place the README file in the root folder
- File documentation: add the name of that file to the title of the README file.

Structure it with **defined sections**

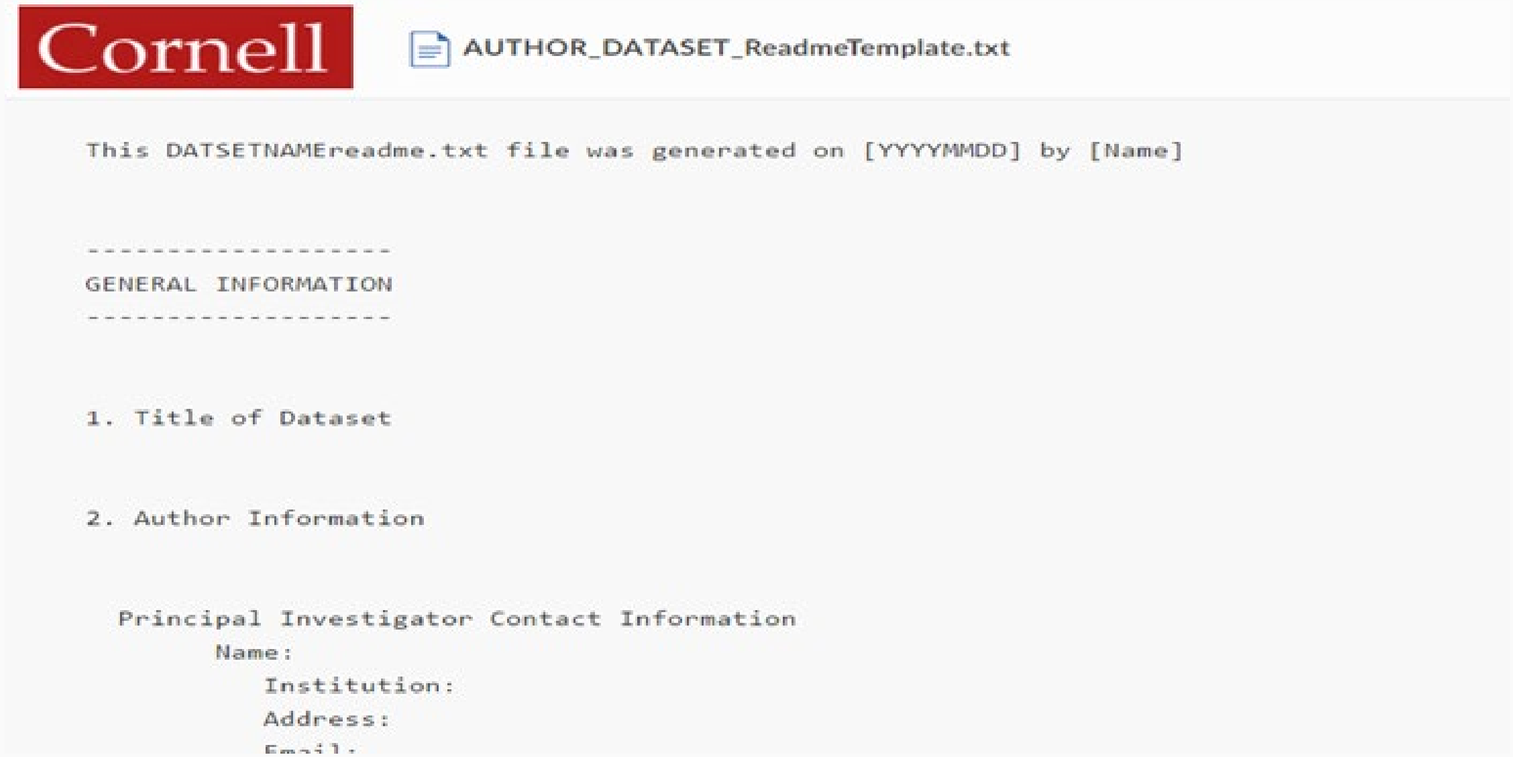
- General information
- Methodological information
- Sharing and access information

Tip: **Create a template** that you can re-use with multiple projects, datasets or files!

10.5281/zenodo.11862672

README - DATA

Use a README file to put your data into context:



https://data.research.cornell.edu/content/readme

More information:
https://datadryad.org/docs/README.md

For Software README files, see level 1 slide 23

10.5281/zenodo.11862672

https://data.research.cornell.edu/content/readme

Exercise 4

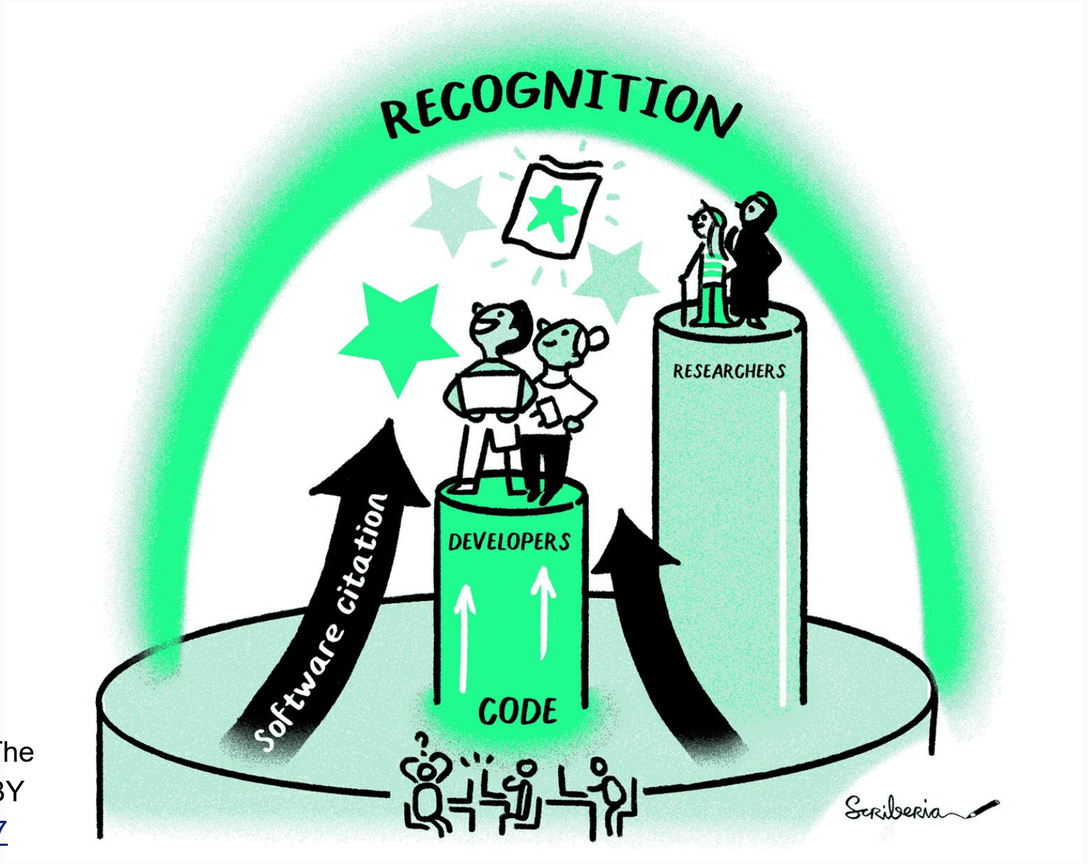
Write a README file for a (potential) dataset you want to share.

Exchange README files and provide feedback!

10.5281/zenodo.11862672

DATA CITATION

- In your **documentation**, include what data you collected and what data has been reused. This can be indicated in the:
  - Research article (citations and data availability statement)
  - New dataset
  - README files
- **Reused data should be cited** in the research article to ensure credit/counting for metrics!

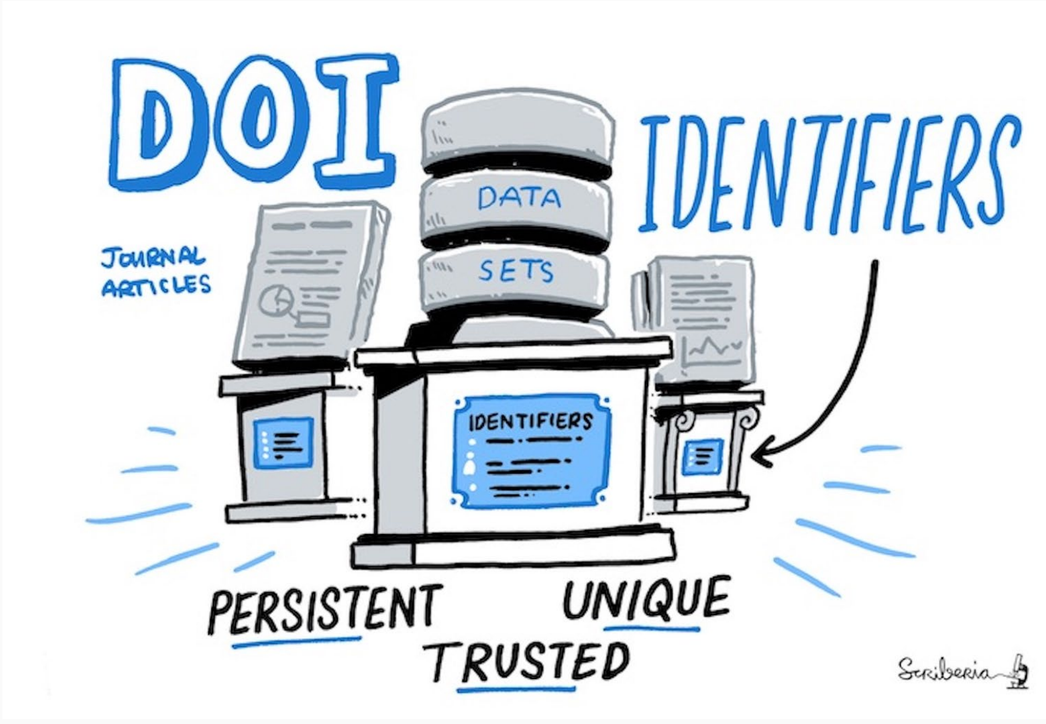


10.5281/zenodo.11862672

HOW TO CITE DATA

A citation includes the following information:

- Author
- Title
- Year of publication
- Publisher (for data, this is often the data repository where it is housed)
- Version (if indicated)
- Access information (a URL or DOI)



10.5281/zenodo.11862672

DATA AVAILABILITY STATEMENT

- A place where you can repeat in more detail what data has been reused, as well as the citations to those sources.
- See The Turing Way for some example Data Availability Statements

“The data that support the findings of this study are openly available in [repository name] at http://doi.org/[doi].”

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in Tables 2-4 as well as openly available at the 4TU.Centre for Research Data (Plomp, Verdegaal-Warmerdam, & Davies, 2020, http://doi.org/10.4121/uuid:f6dc4f20-a6e0-4b2f-b2f8-b79a4f9061c3).

10.5281/zenodo.11862672

LINKING RESEARCH OBJECTS

How do you link the data, code and article together?

- Linking Research Objects – The Turing Way

Always check whether the persistent identifier of the data/code is listed in the article (Data Availability Statement) and in the references! It is important to cite data like articles.

10.5281/zenodo.11862672

QUESTIONS

&

ANSWERS?

10.5281/zenodo.11862672

THANK YOU! QUESTIONS?

Bosun

bosun.obileye@datacite.org

Esther

e.plomp@tudelft.nl

RESOURCES

- Steps to share your data

CREDITS

This presentation template is free for everyone to use thanks to the following:

Slides Carnival

for the presentation template

Pexels, Pixabay

for the photos

HAPPY DESIGNING!