



20240529

[10.5281/zenodo.10994248](https://doi.org/10.5281/zenodo.10994248)

Esther Plomp & Bosun Obileye

Open Data: Principles & Tools

Level 1, presentation 2

Bosun Obileye
Engagement Specialist



Esther Plomp
Data Steward



[@ToothFAIRy](#) Mastodon
[EstherPlomp](#) LinkedIn
e.plomp@tudelft.nl



TRAINERS

Agenda

CARE

FAIR

Data Management Plans

RDM tools

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



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

The CARE Principles

The CARE Principles for Indigenous Data Governance are people and purpose-oriented, reflecting the crucial role of data in advancing Indigenous innovation and self-determination.

The Trend

The current movement toward open data and open science does not fully engage with Indigenous Peoples rights and interests.

The Impact

Existing principles within the open data movement (e.g. FAIR principle) primarily focus on characteristics of data that will facilitate increased data sharing among entities while ignoring power differentials and historical contexts.



What are Indigenous Data?

Data, information and knowledges, in any format, that impacts Indigenous peoples, nations, and communities at the collective and individual levels.

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

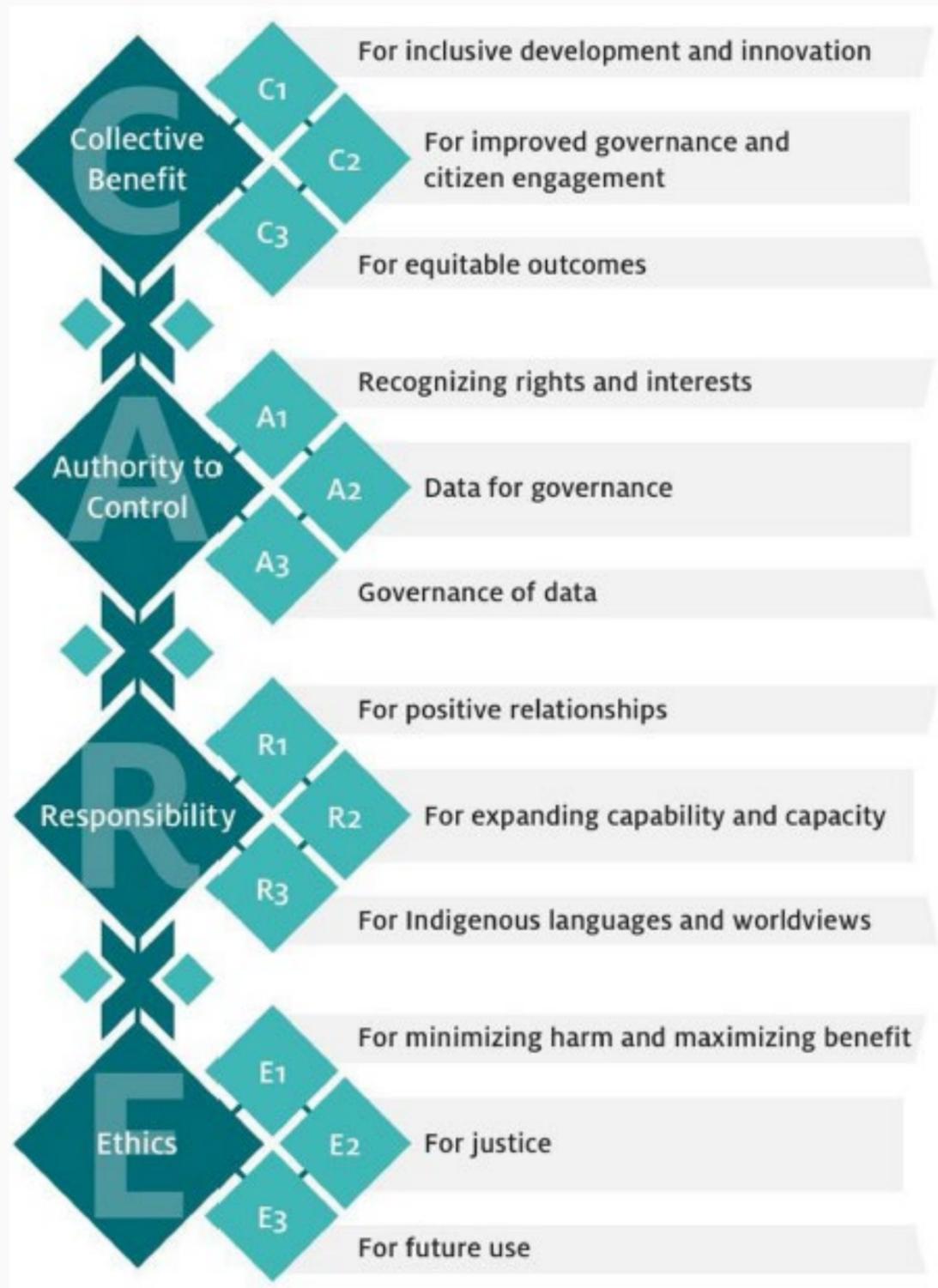
Data about our non-human relations

Land, water, geology, titles, air, soil, sacred ecosystems, territories, plants, animals

Data about us as individuals
Administrative, legal, health, social, commercial, corporate, services

Data about us as collectives

Traditional and cultural information languages, knowledge systems, ancestral and clan knowledges



CARE

Collective benefit

Data ecosystems shall be designed and function in ways that enable Indigenous Peoples to derive benefit from the data.

Responsibility

Those working with Indigenous data have a responsibility to share how those data are used to support Indigenous Peoples' self determination and collective benefit.

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

Authority to Control

Indigenous Peoples' rights and interests in Indigenous data must be recognised.

Ethics

Indigenous Peoples' rights and wellbeing should be the primary concern across the data ecosystem.

FAIR PRINCIPLES



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

FINDABLE

- Descriptive metadata
- Persistent identifier
 - > Data repositories



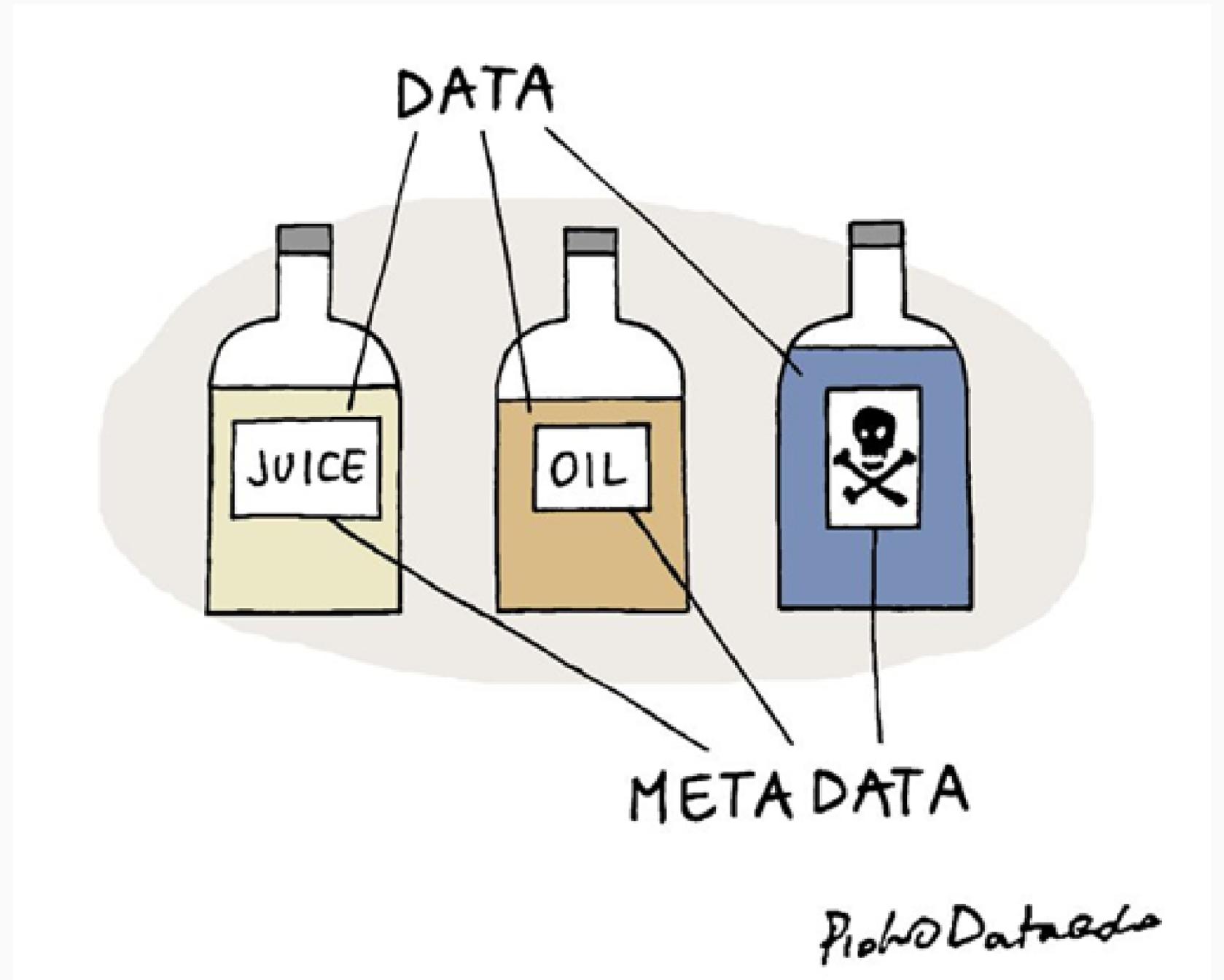
This illustration is created by Scriberia with The Turing Way community.
Used under a CC-BY 4.0 license. DOI: [10.5281/zenodo.3332807](https://doi.org/10.5281/zenodo.3332807)

METADATA

Metadata = Information about data

– Contextual information:

- Who created the data?
- What does the data contain?
- When? Where?
- Why? For what purpose?
- How were data created?
- Size? Standards?



DATA REPOSITORIES

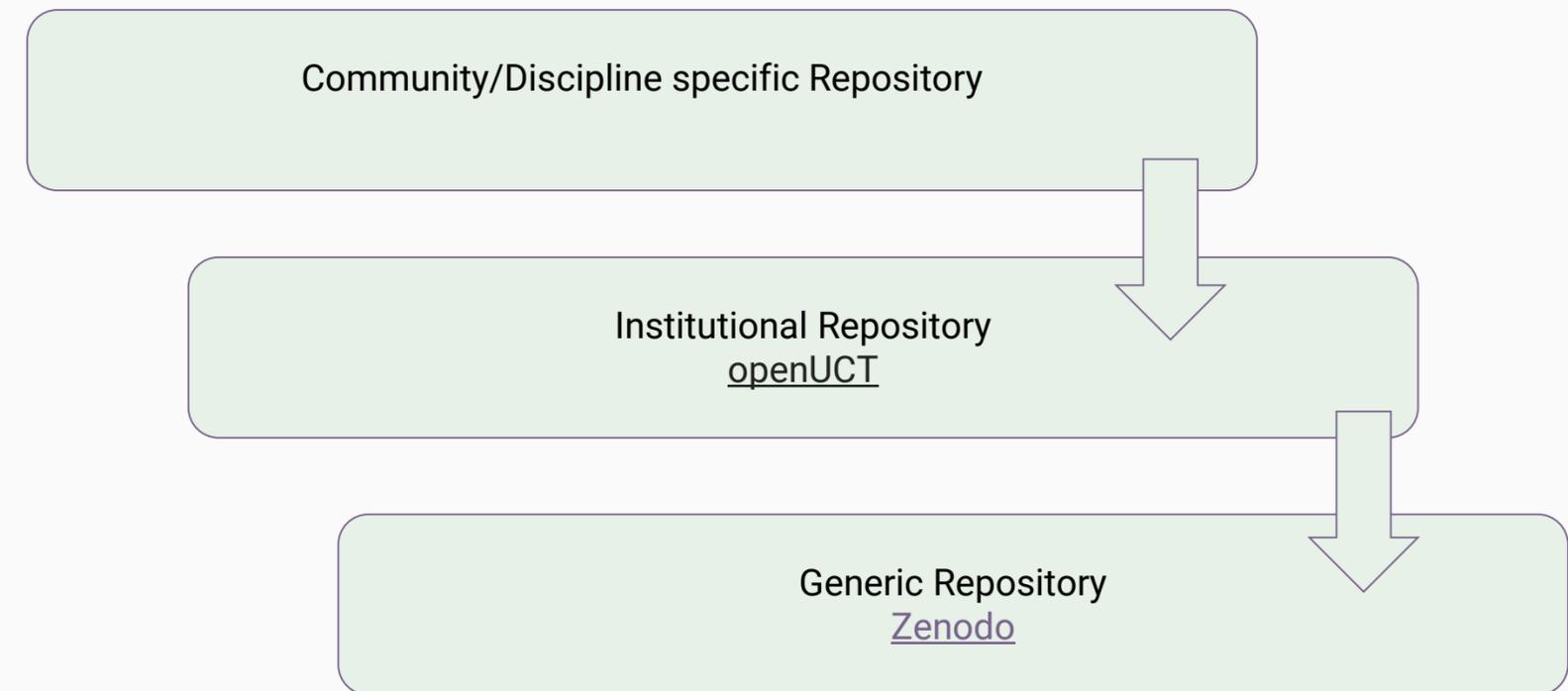
[The Turing Way - selecting an appropriate repository](#)

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



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

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

And more..

Find more using
[FAIRsharing](#) or [re3data](#)

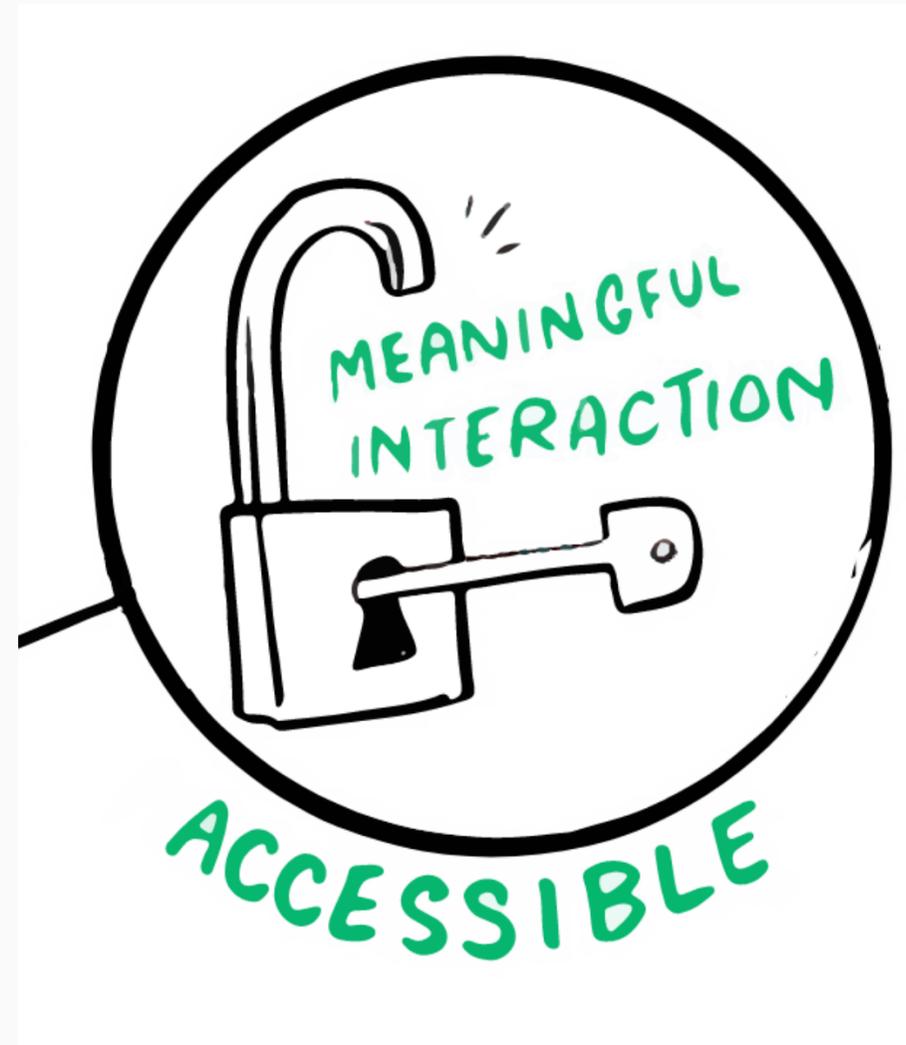
AfricArXiv

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

ACCESSIBLE

FAIR

Data needs to be accessible: could be available under **restricted access**.



This illustration is created by Scriberia with The Turing Way community. Used under a CC-BY 4.0 license. DOI: [10.5281/zenodo.3332807](https://doi.org/10.5281/zenodo.3332807)

OPEN DATA

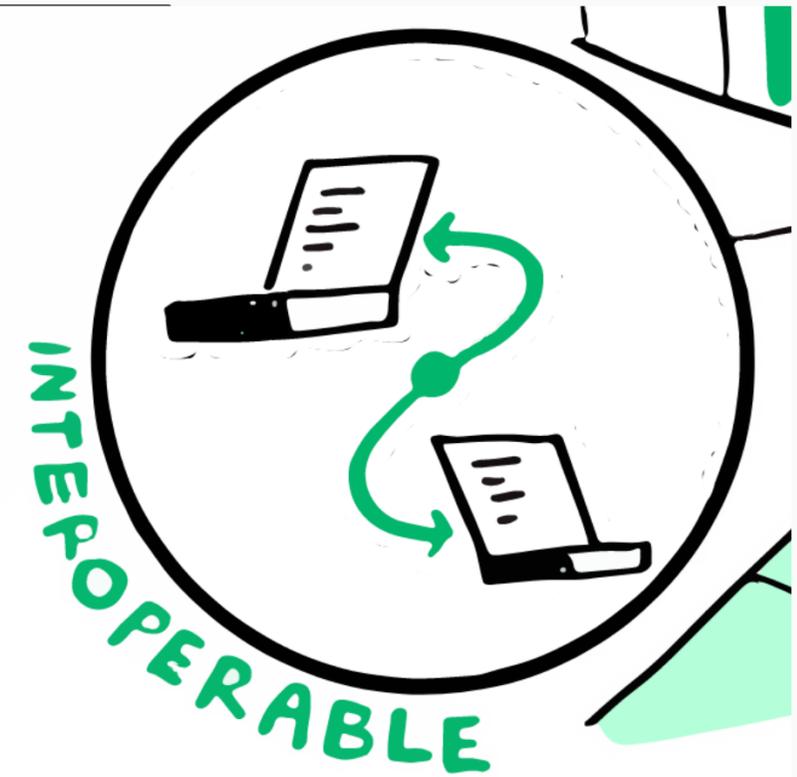
Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike.

[Open Data Handbook](#)

INTEROPERABLE

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

- Similar file formats
- Metadata standards



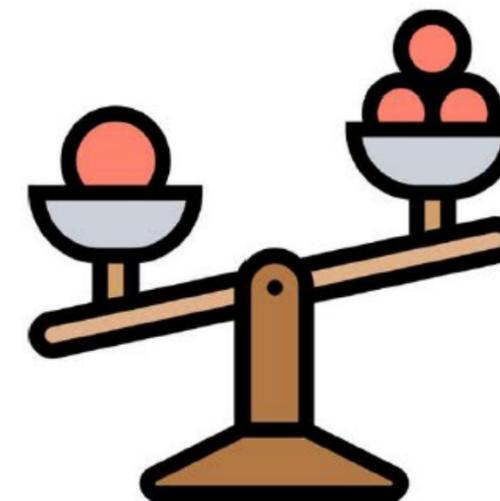
This illustration is created by Scriberia with The Turing Way community. Used under a CC-BY 4.0 license. DOI: [10.5281/zenodo.3332807](https://zenodo.org/doi/10.5281/zenodo.3332807)

FILE FORMATS

- **Proprietary data formats** often are not readable without the corresponding (commercial) software and may become obsolete in the future.
- **Open Data Format:** “a freely available published specification which places no restrictions, monetary or otherwise, upon its use”. - <https://opendefinition.org/ofd>
- **Conversion** from proprietary to open formats is often possible, but may result in some loss of data

The file format balancing act

❖ **Easy to use**



- ❖ **Interoperability**
- ❖ **Ease of re-use**
- ❖ **Sufficient metadata**
- ❖ **Independent from commercial software**
- ❖ **Data quality**
- ❖ **Traditions and conventions**
- ❖ **Future-proof**
- ❖ **Preferences of repository**

METADATA STANDARDS

- Standardised/formalised way of structuring **metadata**
- **General standards** such as Dublin Core (author, date, geographical area)
- **Field-specific documentation** and standards, like **Ecological Metadata Language: [EML](#)** (vocabulary and readable XML markup syntax for documenting research data in earth and environmental sciences)
- Find more via:
 - [FAIRsharing.org](https://www.fairsharing.org)
 - [Digital Curation Center](https://www.digicollab.org/)

Exercise 1

Check if metadata
standards are
available for you

[Fairsharing.org](https://fairsharing.org)



REUSABLE

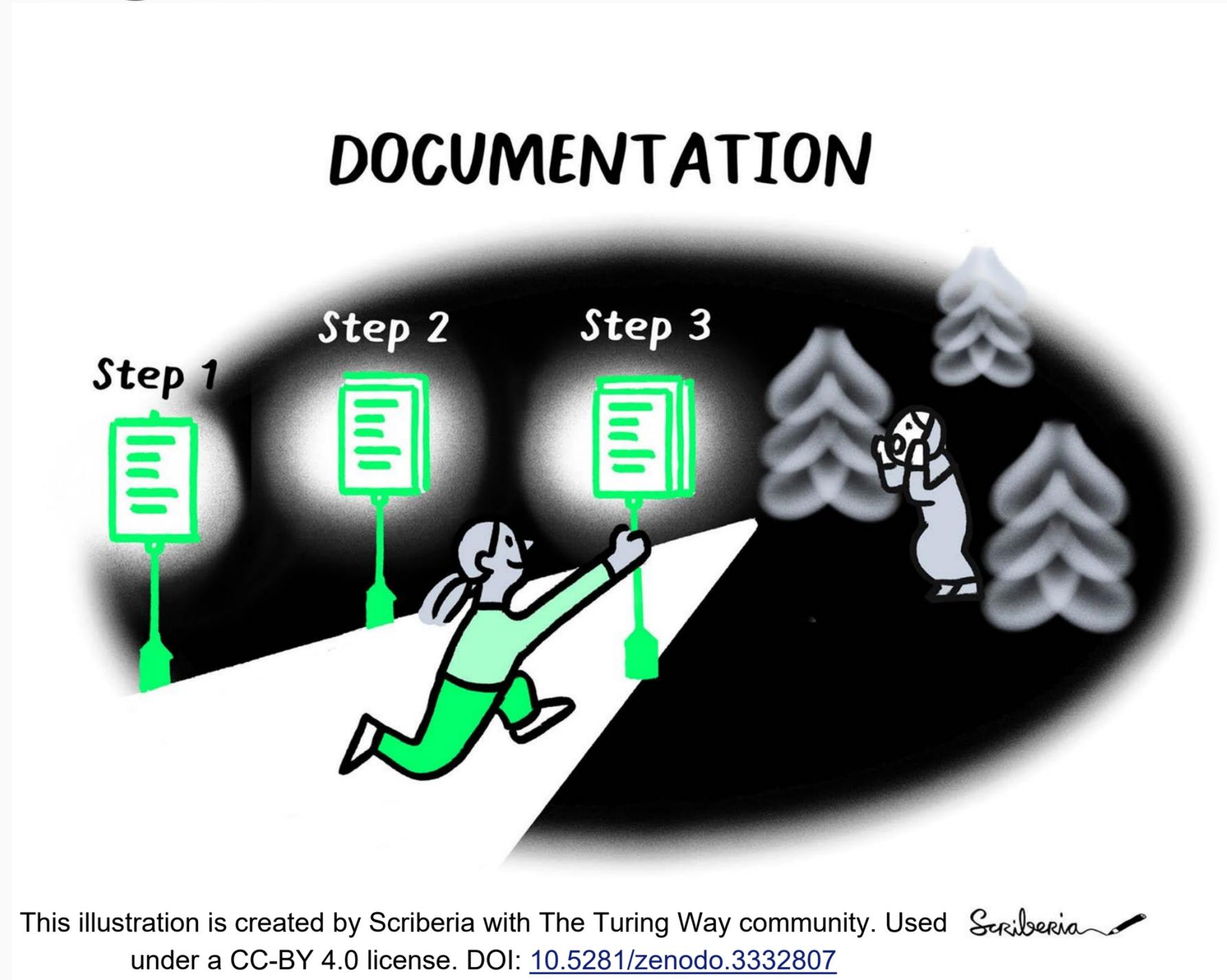
- Documentation
- License



This illustration is created by Scriberia with The Turing Way community. Used under a CC-BY 4.0 license. DOI: [10.5281/zenodo.3332807](https://doi.org/10.5281/zenodo.3332807)

DOCUMENTATION

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



CODE BOOKS

Data dictionaries / code books (**tabular data**) to explain variables

Should include

- Variable name
- Variable explanation
- Measurement unit
- Allowed values
- How missing information is coded

If possible, use standard names for variables!

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.

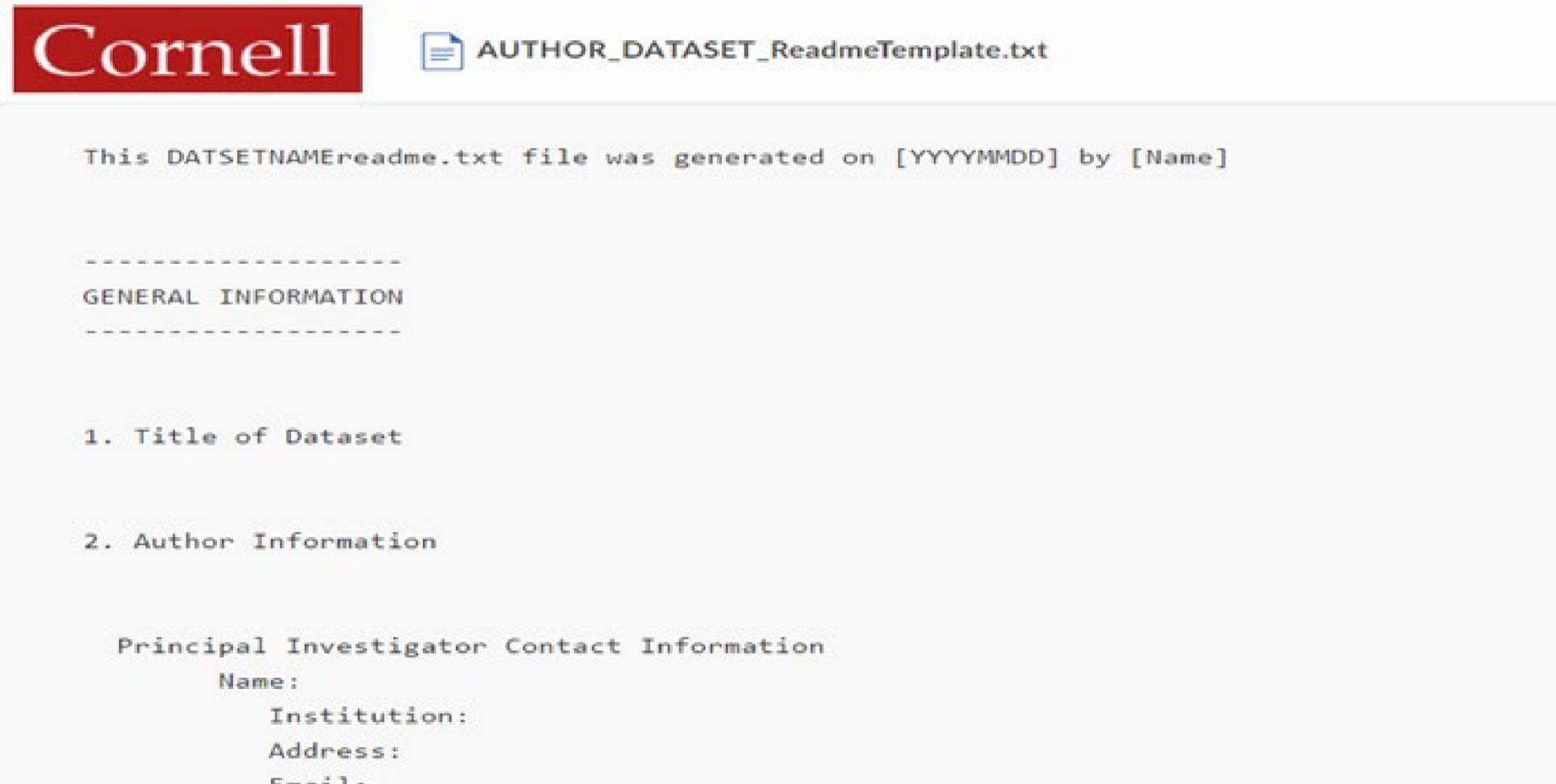
Structure it with **defined sections**

- General information
- Methodological information
- Sharing and access information

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

README - DATA

Use a README file to put your data into context:



Cornell  AUTHOR_DATASET_ReadmeTemplate.txt

```
This DATSETNAMEREADME.txt file was generated on [YYYYMMDD] by [Name]

-----
GENERAL INFORMATION
-----

1. Title of Dataset

2. Author Information

Principal Investigator Contact Information
  Name:
  Institution:
  Address:
  Email:
```

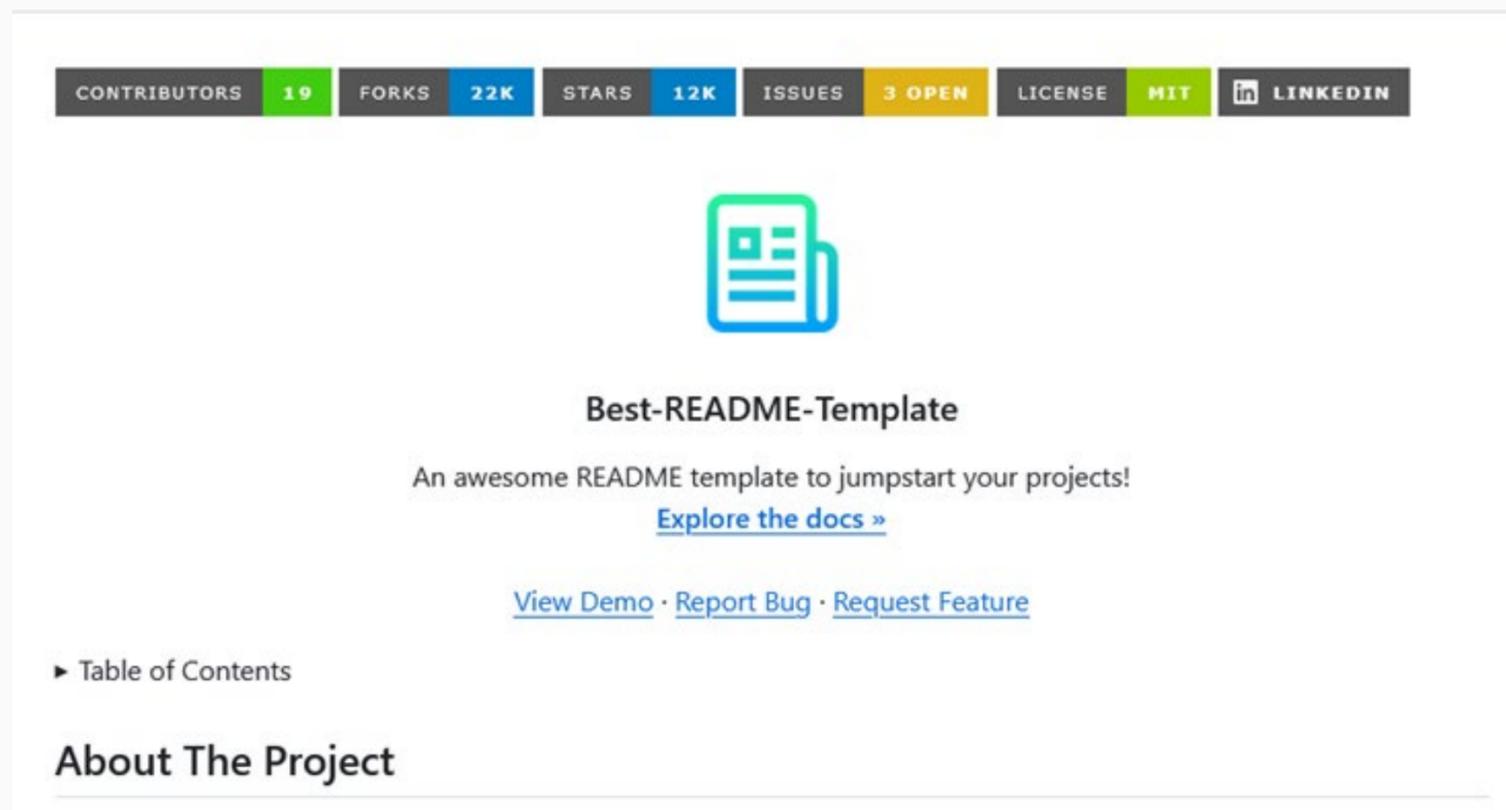
More information:

<https://datadryad.org/docs/README.md>

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

README - SOFTWARE

Use a README file to put your software into context:



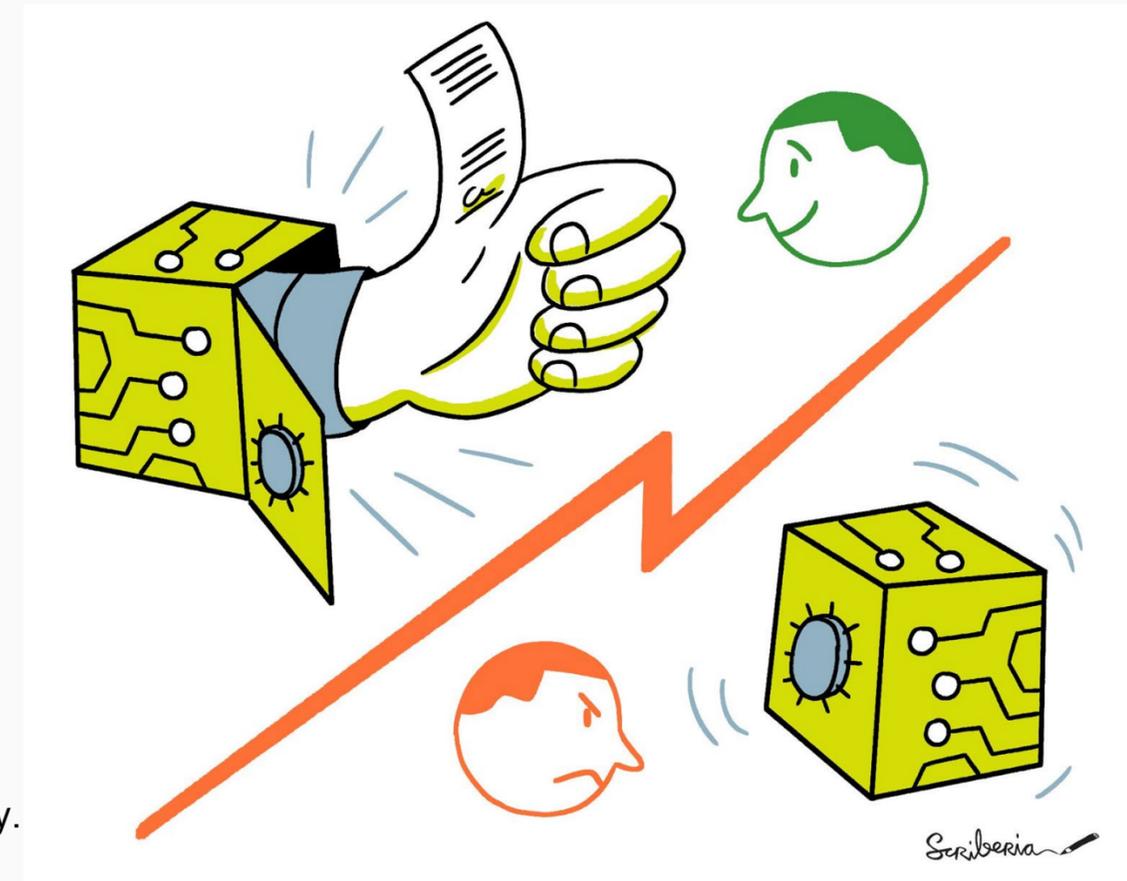
The screenshot shows the GitHub repository page for 'Best-README-Template'. At the top, there is a navigation bar with the following statistics: CONTRIBUTORS 19, FORKS 22K, STARS 12K, ISSUES 3 OPEN, LICENSE MIT, and a LINKEDIN icon. Below this is a large blue icon of a document with a checklist. The repository name 'Best-README-Template' is displayed in bold, followed by the description 'An awesome README template to jumpstart your projects!' and a link 'Explore the docs >'. Below the description are links for 'View Demo', 'Report Bug', and 'Request Feature'. At the bottom left, there is a 'Table of Contents' link and a section titled 'About The Project'.

Examples:

- <https://github.com/othneildrew/Best-README-Template>
- https://ha0ye.github.io/CW21-README-tips/template_README.html
- <https://github.com/manuGil/fair-code>

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



This illustration is created by Scriberia with The Turing Way community.
Used under a CC-BY 4.0 license. DOI: [10.5281/zenodo.3332807](https://zenodo.org/doi/10.5281/zenodo.3332807)

LICENSES - DATA

Creative Commons License Chooser

CREATIVE COMMONS LICENSES

		 COPY & PUBLISH	 ATTRIBUTION REQUIRED	 COMMERCIAL USE	 MODIFY & ADAPT	 CHANGE LICENSE
	PUBLIC DOMAIN	✓	✗	✓	✓	✓
	CC BY	✓	✓	✓	✓	✓
	CC BY-SA	✓	✓	✓	✓	✗
	CC BY-ND	✓	✓	✓	✗	✓
	CC BY-NC	✓	✓	✗	✓	✓
	CC BY-NC-SA	✓	✓	✗	✓	✗
	CC BY-NC-ND	✓	✓	✗	✗	✓

 You can redistribute (copy, publish, display, communicate, etc.)

 You have to attribute the original work

 You can use the work commercially

 You can modify and adapt the original work

 You can choose license type for your adaptations of the work.

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

LICENSES - SOFTWARE

- Choose an open source license
- tl;dr Legal

copyleft = more restrictive than permissive, because derivative code must be open source and distributed under the **same or equivalent license**

							
Type		Permissive	Permissive	Permissive	Copyleft	Copyleft	Copyleft
Provides copyright protection		✓ TRUE					
Can be used in commercial applications		✓ TRUE					
Provides an explicit patent license		✓ TRUE	✗ FALSE				
Can be used in proprietary (closed source) projects		✓ TRUE	✓ TRUE	✓ TRUE	✗ FALSE	✗ FALSE partially	✗ FALSE for web
Popular open-source and free projects		Kubernetes Swift Firebase	Django React Flutter	Angular.js jQuery, .NET Core Laravel	Joomla Notepad++ MySQL	Qt SharpDevelop	SugarCRM Launchpad

<https://moqod.com/understanding-open-source-and-free-software-licensing/>

Exercise 2

What license do
you prefer?

Data licenses?
Software licenses?



~~FAIR~~

A standard

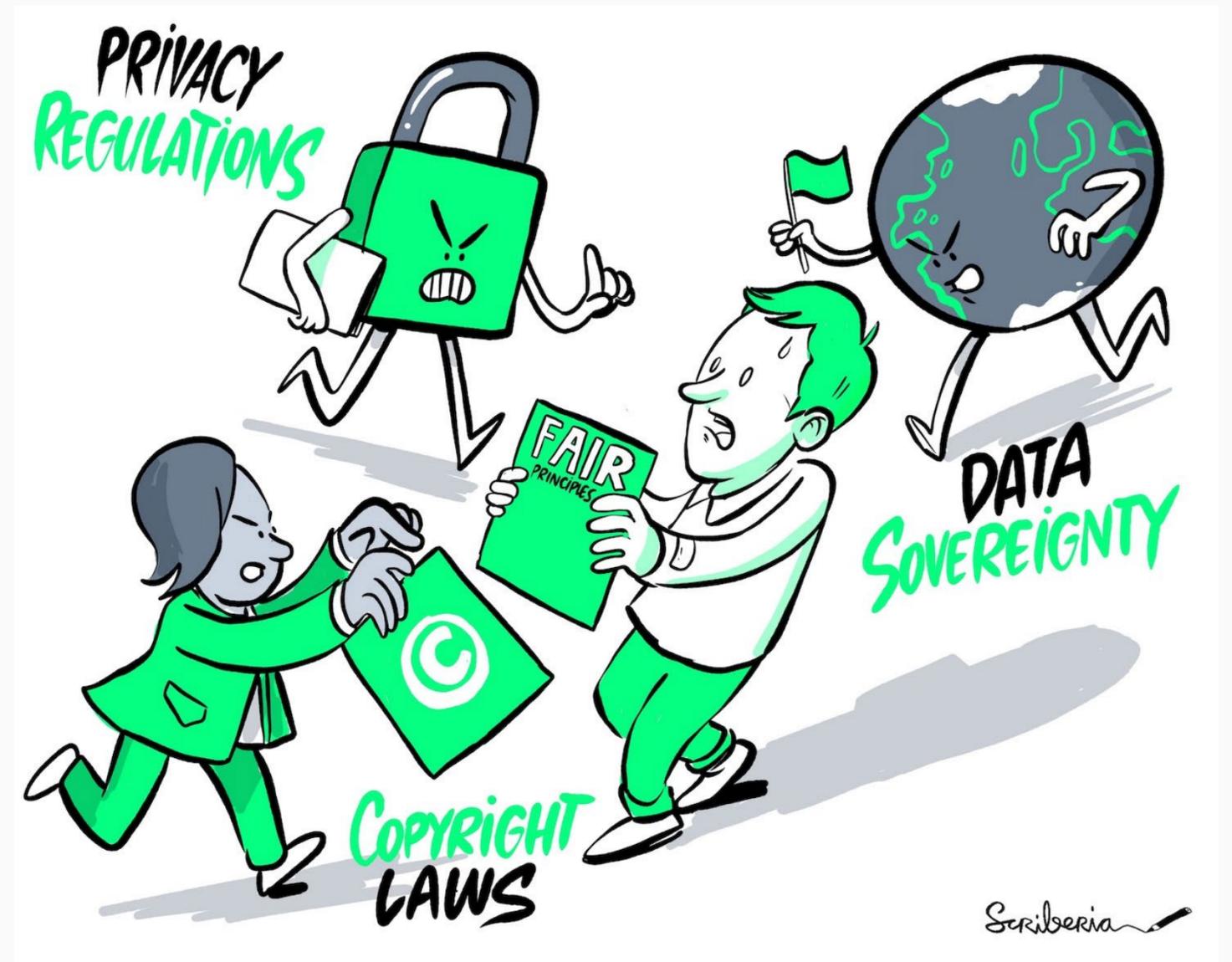
These are recommendations to make data more reusable.

Open Data

FAIR data needs to be accessible: could be available under restricted access.

A tool

There's no one size fits all tool - There might be different tools that enable FAIR data within different disciplines or research workflows.



This illustration is created by Scriberia with The Turing Way community. Used under a CC-BY 4.0 license. DOI: [10.5281/zenodo.3332807](https://doi.org/10.5281/zenodo.3332807)

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.

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>).



This illustration is created by Scriberia with The Turing Way community. Used under a CC-BY 4.0 license. DOI: [10.5281/zenodo.3332807](https://doi.org/10.5281/zenodo.3332807)

QUESTIONS
&
ANSWERS?

DATA MANAGEMENT PLANS

Planning how to manage data

- Roles & Responsibilities
- Planning and data collection
- Type of Data, Documentation
- Sharing and security
- Storage/Backup solutions
- Preservation
- Reuse
- Costs

See [The Turing Way](#) for more details



This illustration is created by Scriberia with The Turing Way community. Used under a CC-BY 4.0 license. DOI: [10.5281/zenodo.3332807](https://doi.org/10.5281/zenodo.3332807)

OPEN SCIENCE FRAMEWORK

- Private / public data sharing
- Individual size limit: 5 GB, no limits on the amount of files



<https://osf.io/>

GITHUB/LAB

- Private / public code sharing
- Private repositories may be limited on free accounts



<https://github.com/>



<https://www.gida-global.org/care>

Be FAIR and CARE!

THANK YOU! QUESTIONS?



Bosun

bosun.obileye@datacite.org

Esther

e.plomp@tudelft.nl

RESOURCES

- www.dmptools.org
- <https://dataseer.ai/>

CREDITS

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



for the presentation template

Pexels, Pixabay

for the photos

HAPPY DESIGNING!
