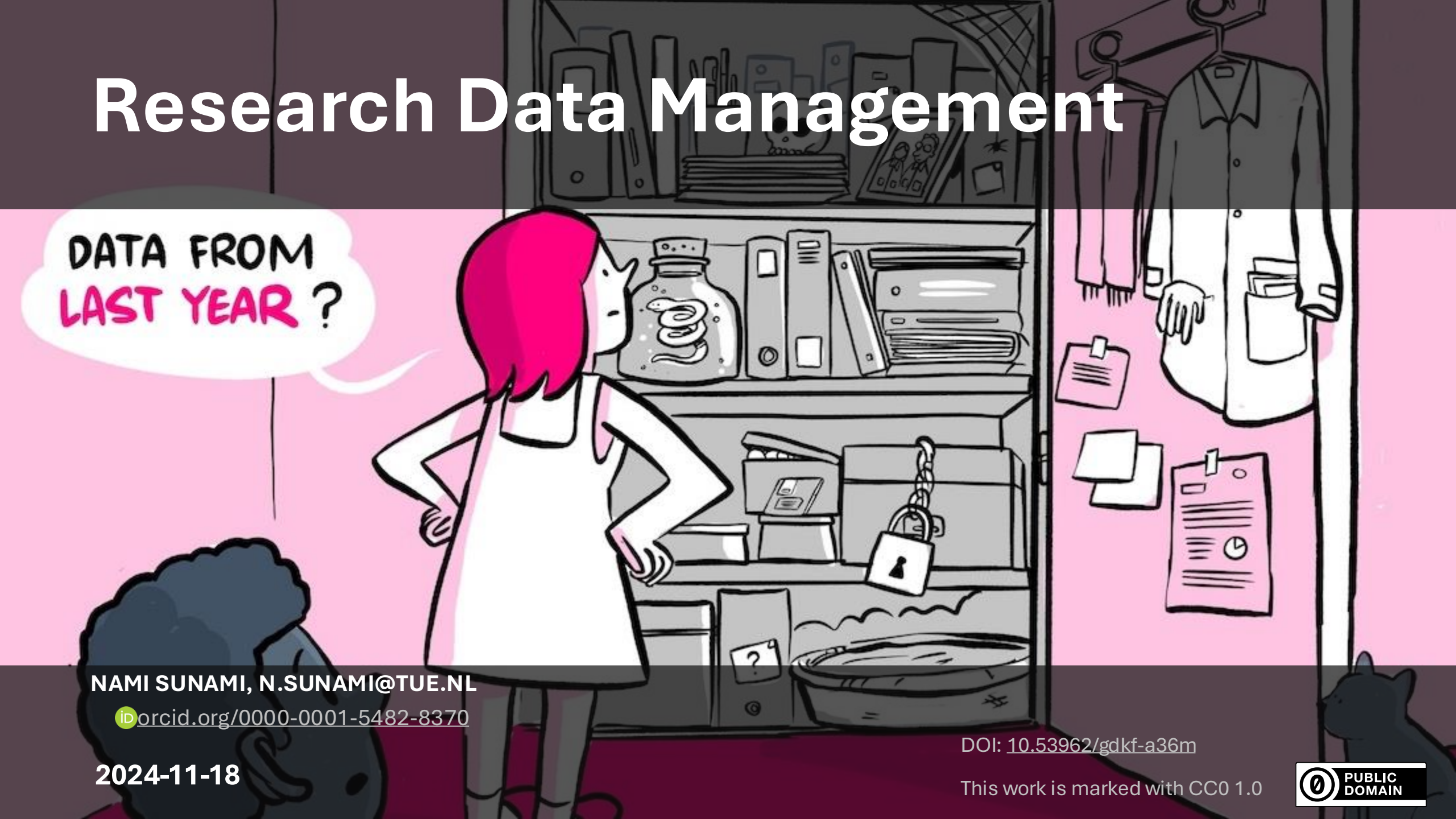
Research Data Management



NAMI SUNAMI, N.SUNAMI@TUE.NL

orcid.org/0000-0001-5482-8370

2024-11-18

DOI: 10.53962/gdkf-a36m

Data Stewards can help you

Liz Guzman-Ramirez
Coordinator

Angela Aleksovska
EE

Davide Nardi
M&CS

Jay Nair
IE&IS

Lucia Forrová
BE

Neda Norouzi
BME, ME

Mariana Oshima Menegon
ID

Nami Sunami
CE&C, APSE

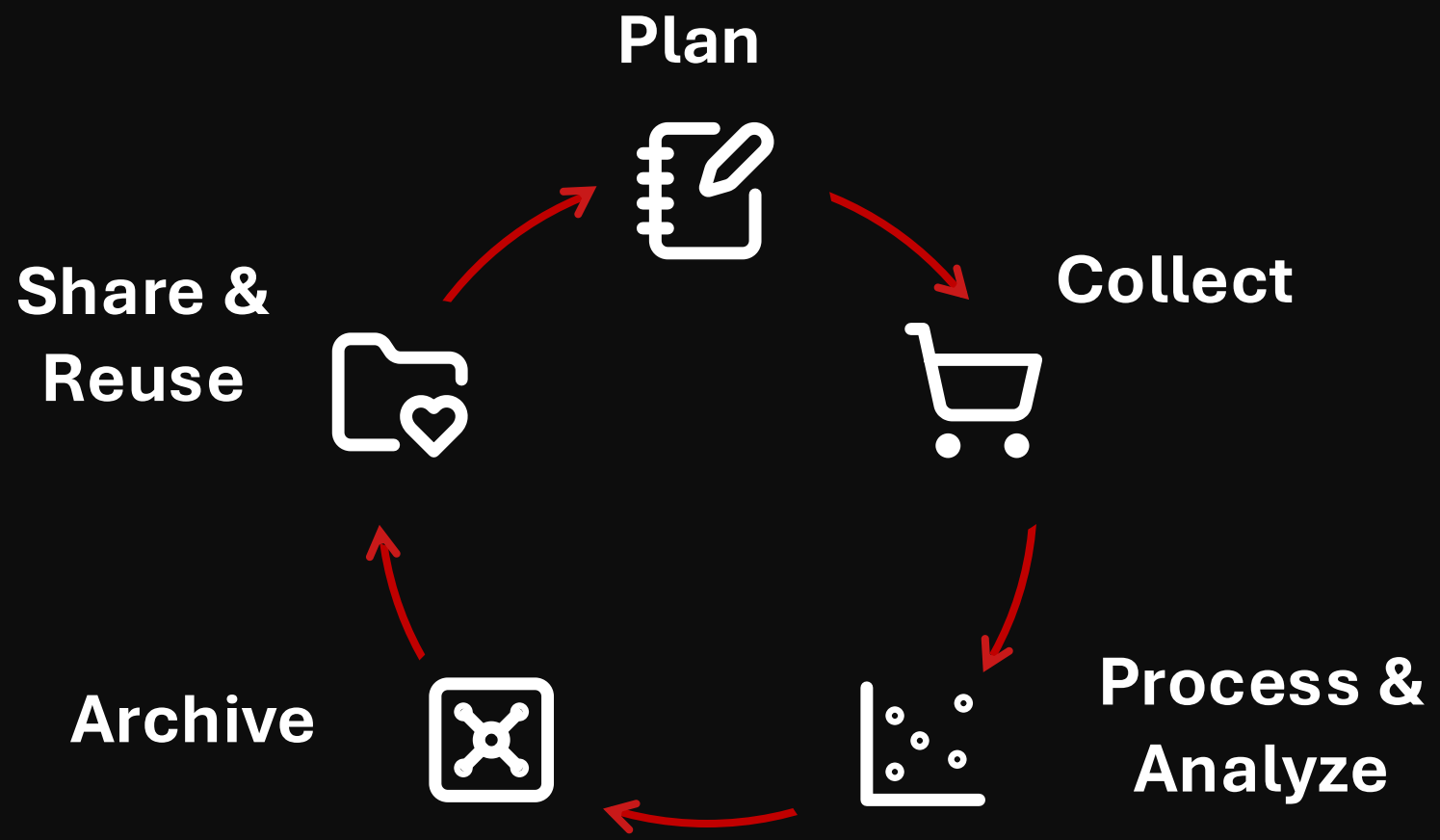
rdmsupport@tue.nl

Research Data Management

Research data management concerns the way you collect, analyze, store, share, archive and publish research data, to satisfy the needs of current and future data users.

Research Data Abandonment

Research Data Management



Adopted from RDMkit: https://rdmkit.elixir-europe.org/data_life_cycle

Why do RDM?

Policy, regulations, requirements

Code of Conduct for Research Integrity

Funder Requirements

GDPR

Code of Conduct for Research Integrity: nwo.nl/en/netherlands-code-conduct-research-integrity
NWO Grant Rules: nwo.nl/en/nwo-grant-rules

Benefit for you

and the research community

Less likely to lose data

Recognition

Fight against scientific fraud

More efficient & sustainable science

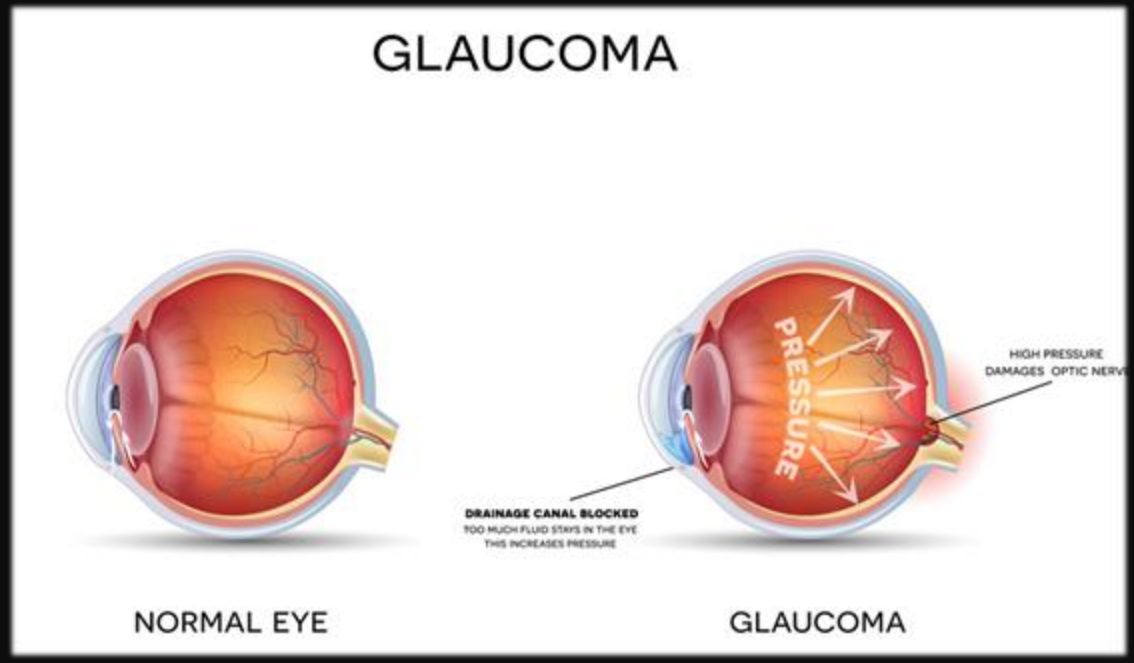
Code of Conduct for Research Integrity: nwo.nl/en/netherlands-code-conduct-research-integrity
NWO Grant Rules: nwo.nl/en/nwo-grant-rules

Example research project

Research Question:

“What causes variation in the rate of development of glaucoma in patients over the age of 50?”

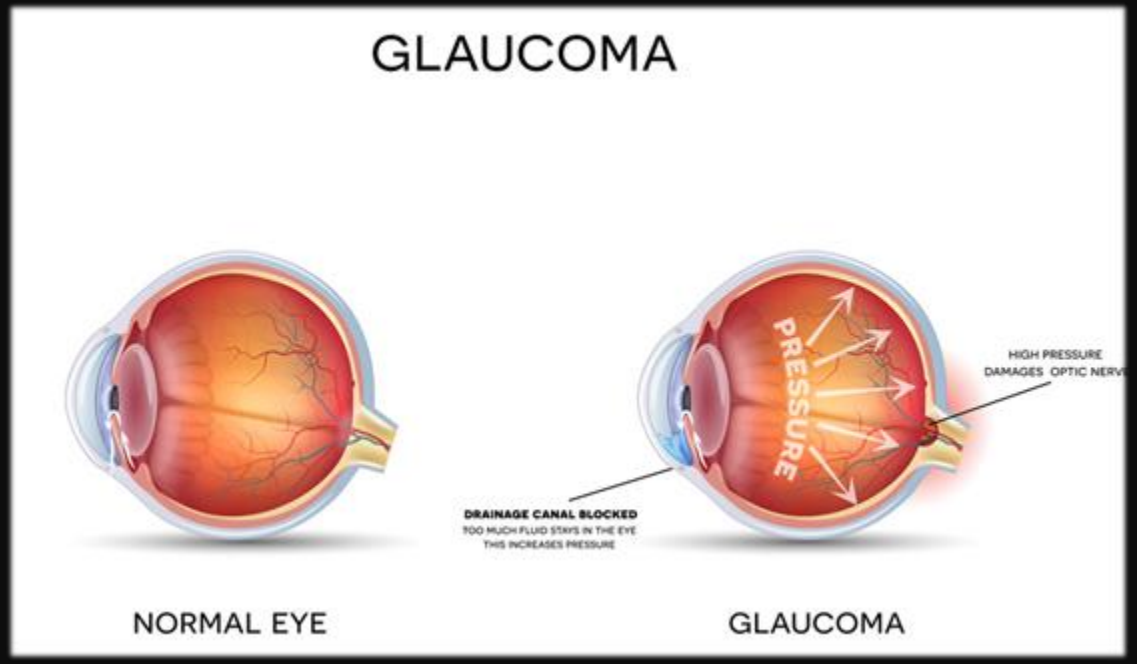
Researcher collaborate with an industry partner & a hospital, exchanging datasets and code



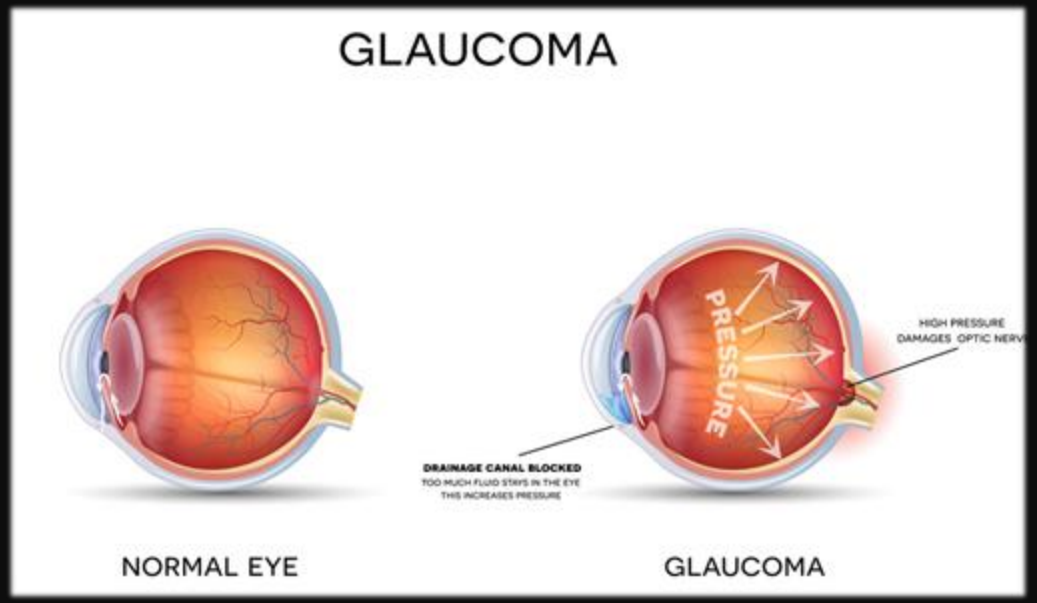
Questions

What are research data management topics you need to consider in:

1) This Glaucoma example
2) Your own research project



What are research data management topics you need to consider?

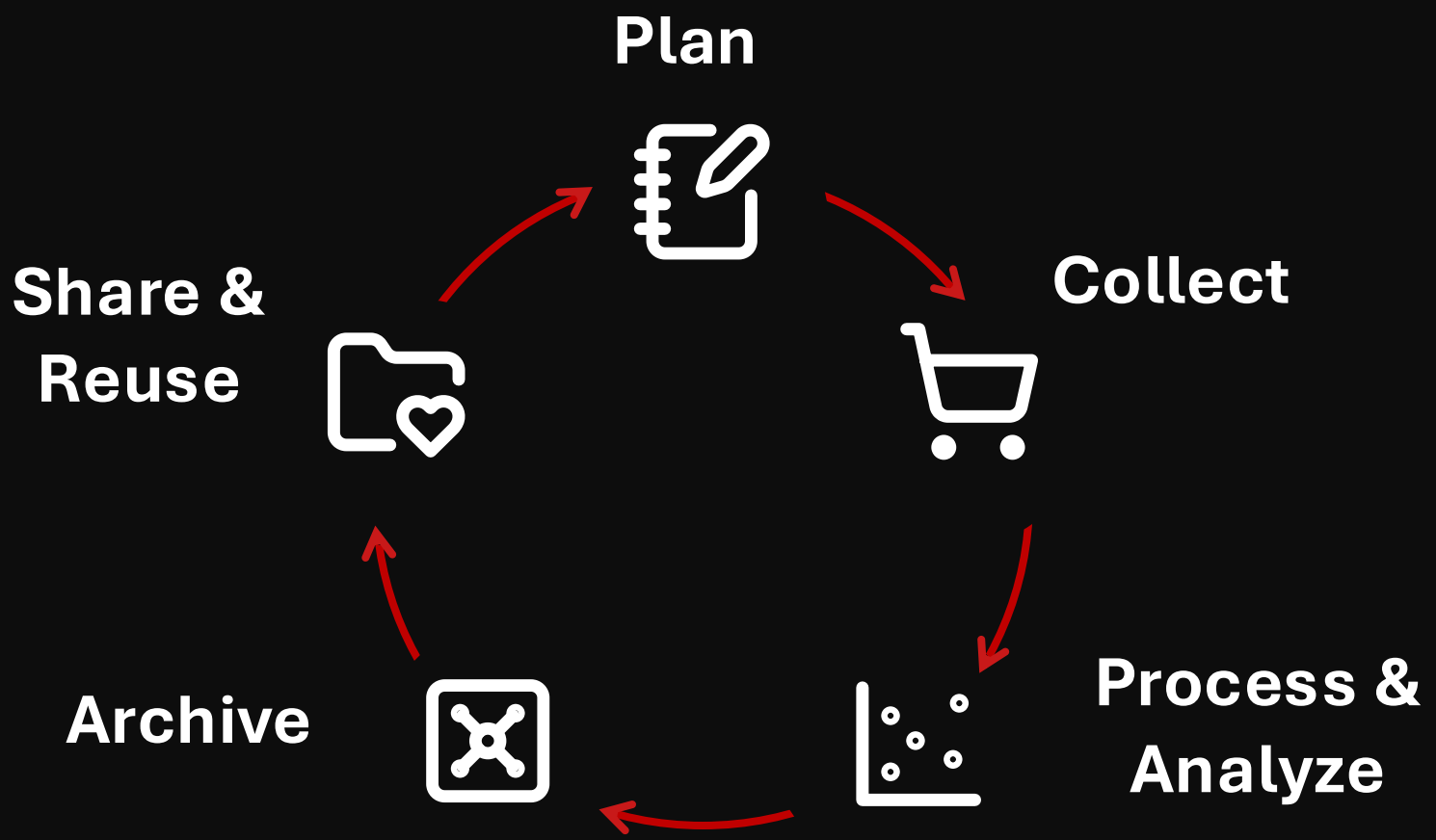


- Ethics
- Privacy
- Intellectual property
- Agreements
- Collection
- Sharing
- Storage

- DMP
- FAIR
- Documentation
- Folder structure
- Filenames
- Encryption
- Anonymisation

- Versioning
- Metadata
- Software
- Archiving
- Publishing
- Repositories
- Licensing

Research Data Lifecycle



Adopted from RDMkit: https://rdmkit.elixir-europe.org/data_life_cycle

Plan

Topics to consider when planning

RDM Costs

Ethics

Privacy

RDM costs

Good RDM planning means lower RDM costs later (both time and money)

Our advice is to reserve part of the funding for research data management-related costs

Ethical review process

Mandatory for all research or educational projects with human data

The purpose is to protect human participants and their data from harm

Low- and medium-risk
Assessed by departmental ERB.

2 working weeks to process

High-risk:
Assessed by the central ERB.

6 working weeks to process

Ethical Review | TU/e : https://tuenl.sharepoint.com/sites/intranet-ethical-review

Personal Data

Pla

Any information...

relating to an identified and identifiable...

natural person.

Any data that can be traced back to a living person

Privacy Handbook | UU: https://utrechtuniversity.github.io/dataprivacyhandbook/definitions.html

Special Categories of Personal Data

Further care is needed for special categories of personal data

Racial or ethnic origin

Political opinions

Religious or philosophical beliefs

Trade union membership

Genetic and biometric data to identify someone

Sex life or sexual orientation

Physical and mental health conditions

Best practices to protect privacy of participants

Minimize personal data collected

Assess the privacy and security risks — Data Processing Impact Assessment (DPIA) may be needed

Inform participants, obtain a consent

Anonymize or pseudonymize personal data

Ethics application process

1. Complete ERB form and consent form. A data steward can help you (rdmsupport@tue.nl)
2. Email the documents to the ERB ethics@tue.nl

New ERB process is being piloted at Research Cockpit.
Contact your Data Steward.

If you are getting data from an external party, you may need an agreement

E.g., data sharing agreement

Agreements | TU/e: https://tuenl.sharepoint.com/sites/intranet-privacy-security/SitePages/agreements.aspx

Data management plan

What type of data you will collect, and how much

How you will address privacy and intellectual property challenges

How you will archive and publish your data

How you are going to achieve FAIR principles

NWO and European Commission require a DMP

UU | Cost of Data management: uu.nl/en/research/research-data-management/guides/costs-of-data-management
NWO DMP: nwo.nl/en/research-data-management

Create your Data Management Plan on

Research Cockpit

https://tue.atlassian.net/helpcenter/research/

Planning may take time, but the effort will pay off later to make your data Open & FAIR

What do ”Open” & “FAIR” mean?

Open data is data anyone can freely access, use, modify, and share for any purpose

Not all data can be shared freely

FAIR Data

FAIR Data is data that satisfies FAIR Principles

Findable

FAIR Data is data that satisfies FAIR Principles

Findable

Accessible

FAIR Data is data that satisfies FAIR Principles

Findable

Accessible

Interoperable

FAIR Data is data that satisfies FAIR Principles

Findable

Accessible

Interoperable

Reusable

Findable

Accessible

Interoperable

Reusable

?

Open Data

Open ≠ FAIR

Open

Yes

No

FAIR & Open

No

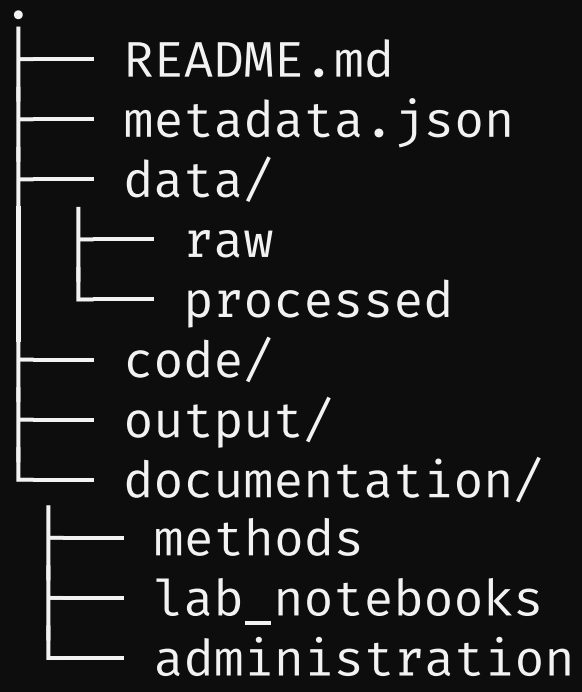
Yes

FAIR

Collect

Set your folder structure

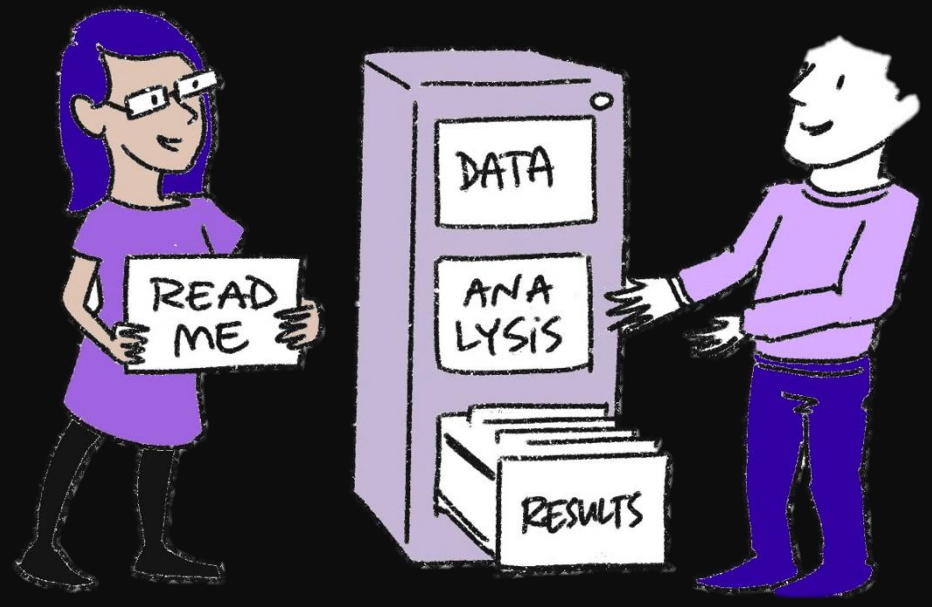
Document the rationale on README



A README file makes your data human-friendly

A README file is a human-readable text file document that explains the project.

No one can read your mind—not even you in the future.



Make a README: makeareadme.com/

Guide to writing "readme" style metadata | Cornell: data.research.cornell.edu/data-management/sharing/readme/

A metadata file makes your data machine-readable

On 4TU.ResearchData or Zenodo, you can start a dataset to prepare and download a metadata file, even without data.

You can manually create metadata file using DataCite Metadata Generator

Photo by AltumCode on Unsplash

DataCite Metadata Generator: dhvlab.gwi.uni-muenchen.de/datacite-generator/

Name files meaningfully, distinctly, and consistently—and document it

No special characters. Whitespaces are discouraged.

Case style: Do you want a `camelCase`, `PascalCase`, `snake_case`, **or** `kebab-case`**?**

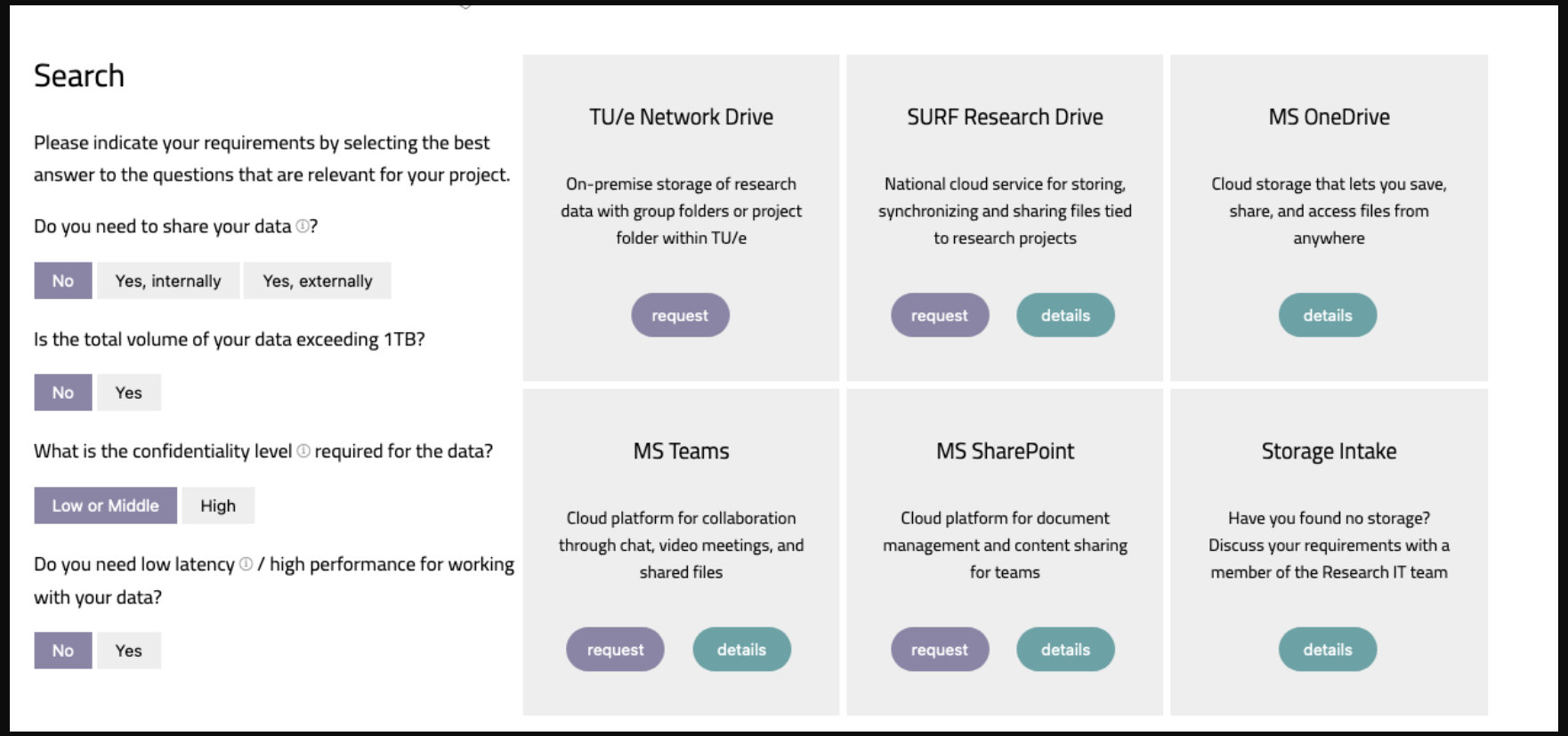
ISO date format is good to use: “2024-07-03”

Turing Way: https://book.the-turing-way.org/project-design/filenaming.html

Store your data in a secured space with backups, where two or more people retain access.

Storage options at TU/e

Storage Finder at Solution Searcher



https://openpar.pages.tue.nl/solution-searcher/category-storage.html

	Personal/ Sensitive Data	Collaboration
TU/e Network Drive	✓✓	TU/e only
SURF Research Drive	✓	✓
Teams/SharePoint (up to 2TB)	✓	✓
OneDrive (up to 1TB)	✓	File/folder-based sharing only

You can encrypt your data using Cryptomator

You can do client-side encryption to store your data in password-protected Vaults

Cryptomator Hub | Resarch Cockpit : https://tue.atlassian.net/helpcenter/research/

Only sending data one-time?
Use SURF File Sender

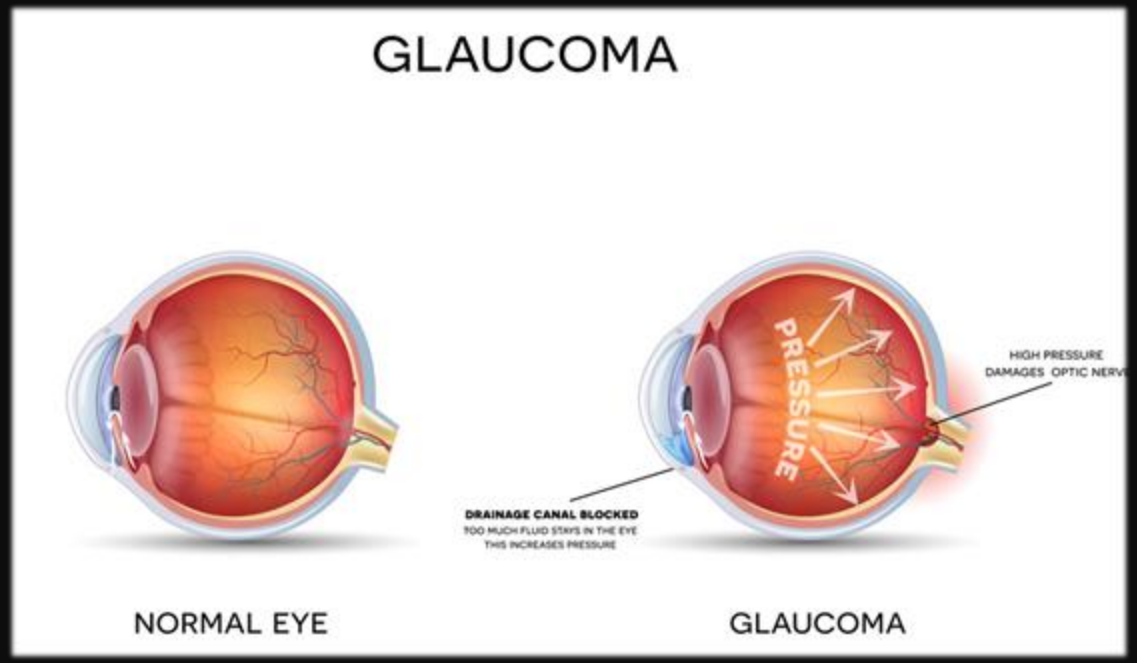
Remember to turn on encryption

SURFfilesender: https://filesender.surf.nl/

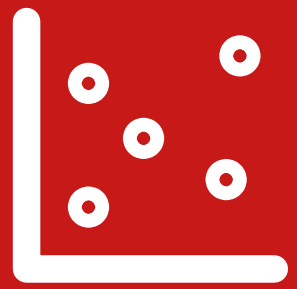
Storage for the Glaucoma example

Where would you store the glaucoma data at TU/e?

A) Dropbox
B) OneDrive
C) ResearchDrive
D) Network drive

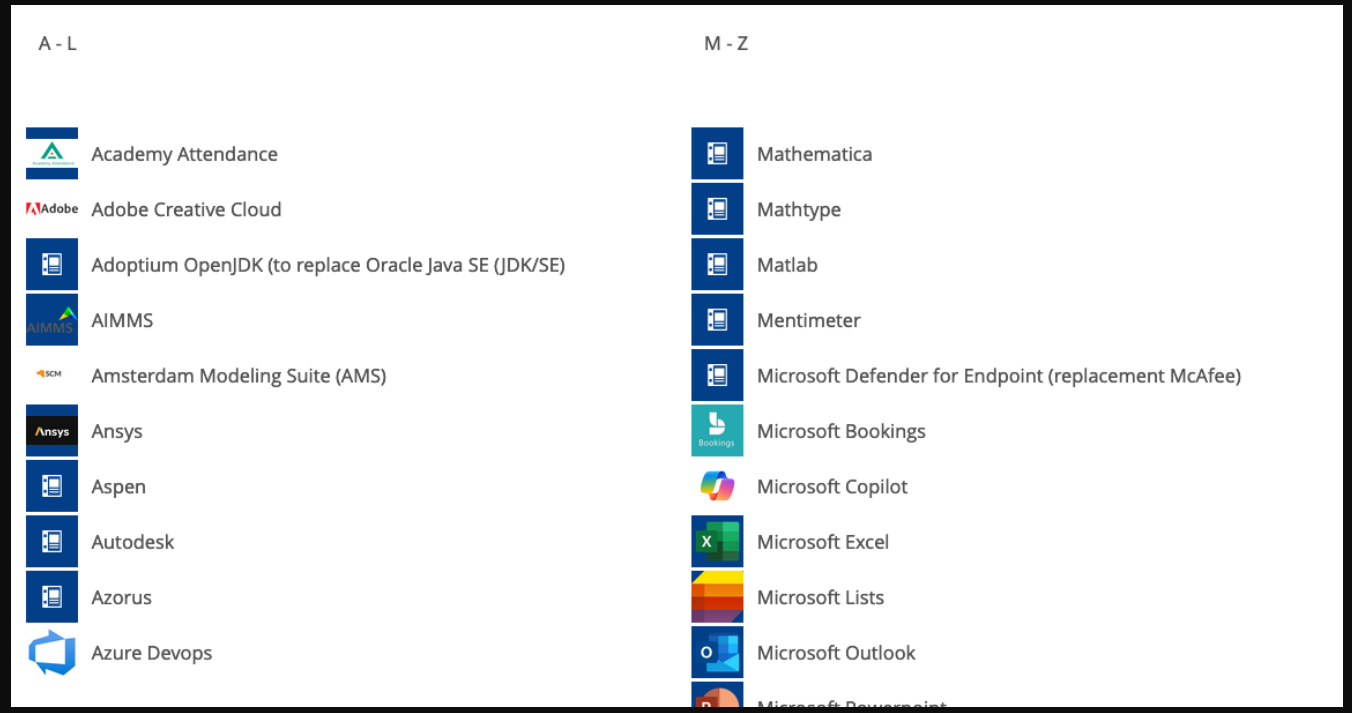


Process & Analyze



Find TU/e Tools on Topdesk

Overview of software available at TU/e: TU/e Software | Topdesk

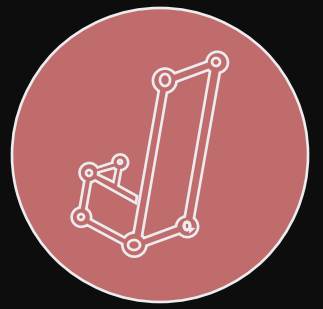


Use open-source tools,

whenever possible.

Python

R



JASP

Jamovi

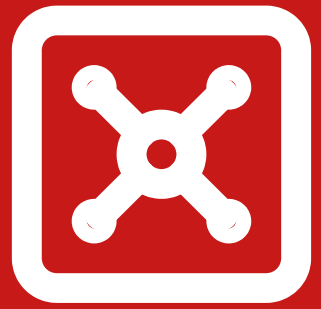
If the original dataset is in a proprietary format, convert it to an open file format.

Questions

Do you need to use proprietary tools?

Any opportunities for using open tools?

Archive



Archiving data means keeping data to demonstrate the integrity of research

Typically, each publication requires an archival package

Preserving Data | UU: uu.nl/en/research/research-data-management/guides/after-research/preserving-data

Access to archival package is limited

Accessed only when it’s need (for example, an integrity audit)

Usually stored in “colder storage” (cheaper storage but expensive transfers)

Photo by Jason Dent on Unsplash

Share & Reuse

Creating a publication data package to share your data

The goal is to share data with others and future data users.

Anyone can access a publication package by default

	Publication Package	Archival Package
Main Goal	To share 	To demonstrate integrity 
Access	Public  By default	Restricted 

You can deposit data in a data repository & connect your publication.

Your discipline may have specialized data repositories. Use re3data.org to find out.

You can use generic repositories, such as 4TU.ResearchData & Zenodo.

Remember to assign an open license when depositing your data.

Data, documentation, non-code materials

CC BY 4.0

Code

MIT

Share your data as openly as possible as closed as necessary

Privacy, intellectual property, or collaboration agreements may pose challenge

FAIR Data Management in Horizon 2020: ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Photo by Mehmet Uzut on Unsplash

Questions

What would be your challenges for sharing and re-using data?

Thank you

More info on RDM

Data Stewards | Intranet

TU/e Research Data Management

Search Storage | Storage Searcher

Create a DMP/ ERB

Research Cockpit

Contact Data Steward

rdmsupport@tue.nl

- Icons in this slide deck are from Lucide, open-source icon library (https://lucide.dev/icons/rainbow)
- Image are from:
  - Unsplash: https://unsplash.com/
  - Turing Way: https://book.the-turing-way.org