Research Data Management

Research Data Management: making it possible to retrieve data from last year. *The Turing Way* project illustration by Scriberia. Original version on Zenodo. *Community & Scriberia (2020)*.

Nami Sunami
Data Steward
n.sunami@tue.nl

orcid.org/0000-0001-5482-8370

DOI: 10.53962/gdkf-a36m

After this part, you'll be able to...

Evaluate your own research data management practices in the research data life cycle

Apply strategies to make your research data Findable, Accessible, Interoperable and Reusable (FAIR)

"Research Data"

Data itself

Nutrition Facts

4 servings per container

Serving size 1 1/2 cup (208g)

Amount per serving

Calories **240**

	% Daily Value*
Total Fat 4g	**5%**
Saturated Fat 1.5g	**8%**
Trans Fat 0g	
Cholesterol 5mg	**2%**
Sodium 430mg	**19%**
Total Carbohydrate 46g	**17%**

Metadata

“Research Data Management”

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

ORIGIN

```
        1 attaaaggtt tataccttcc caggtaacaa accaaccaac tttcgatctc ttgtagatct
       61 gttctctaaa cgaactttaa aatctgtgtg gctgtcactc ggctgcatgc ttagtgcact
      121 cacgcagtat aattaataac taattactgt cgttgacagg acacgagtaa ctcgtctatc
      181 ttctgcaggc tgcttacggt ttcgtccgtg ttgcagccga tcatcagcac atctaggttt
      241 cgtccgggtg tgaccgaaag gtaagatgga gagccttgtc cctggtttca acgagaaaac
      301 acacgtccaa ctcagtttgc ctgttttaca ggttcgcgac gtgctcgtac gtggctttgg
      361 agactccgtg gaggaggtct tatcagaggc acgtcaacat cttaaagatg gcacttgtgg
      421 cttagtagaa gttgaaaaag gcgttttgcc tcaacttgaa cagccctatg tgttcatcaa
      481 acgttcggat gctcgaactg cacctcatgg tcatgttatg gttgagctgg tagcagaact
      541 cgaaggcatt cagtacggtc gtagtggtga gacacttggt gtccttgtcc ctcatgtggg
      601 cgaaatacca gtggcttacc gcaaggttct tcttcgtaag aacggtaata aaggagctgg
      661 tggccatagt tacggcgccg atctaaagtc atttgactta ggcgacgagc ttggcactga
      721 tccttatgaa gattttcaag aaaactggaa cactaaacat agcagtggtg ttacccgtga
      781 actcatgcgt gagcttaacg gaggggcata cactcgctat gtcgataaca acttctgtgg
      841 ccctgatggc taccctcttg agtgcattaa agaccttcta gcacgtgctg gtaaagcttc
      901 atgcactttg tccgaacaac tggactttat tgacactaag aggggtgtat actgctgccg
      961 tgaacatgag catgaaattg cttggtacac ggaacgttct gaaaagagct atgaattgca
     1021 gacacctttt gaaattaaat tggcaaagaa atttgacacc ttcaatgggg aatgtccaaa
     1081 ttttgtattt cccttaaatt ccataatcaa gactattcaa ccaagggttg aaaagaaaaa
     1141 gcttgatggc tttatgggta gaattcgatc tgtctatcca gttgcgtcac caaatgaatg
```

Coronavirus genome sequence

GenBank

GenBank: https://www.ncbi.nlm.nih.gov/nuccore/MN908947

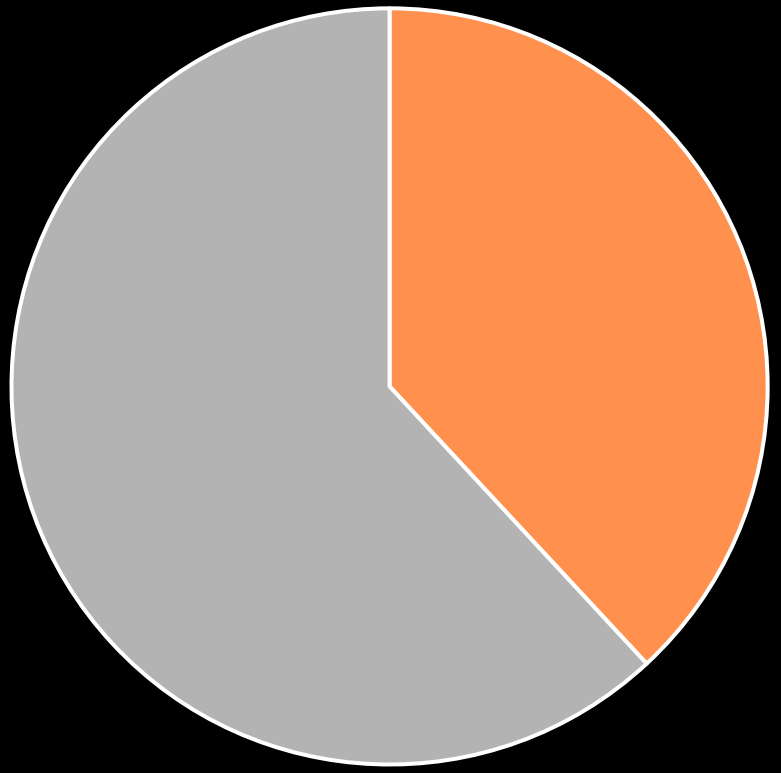
Open & FAIR Data drive innovation
(& save lives)

Research Data: public good

Photo by Will van Wingerden on Unsplash

unesco.org/en/open-solutions/open-data

Journals are asking for data



38% requiring data sharing

Based on 2017 data from 709 journals in life, health, and physical sciences.

Kim, J., Kim, S., Cho, H.-M., Chang, J. H., & Kim, S. Y. (2020). Data sharing policies of journals in life, health, and physical sciences indexed in Journal Citation Reports. *PeerJ*, *8*, e9924. https://doi.org/10.7717/peerj.9924

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

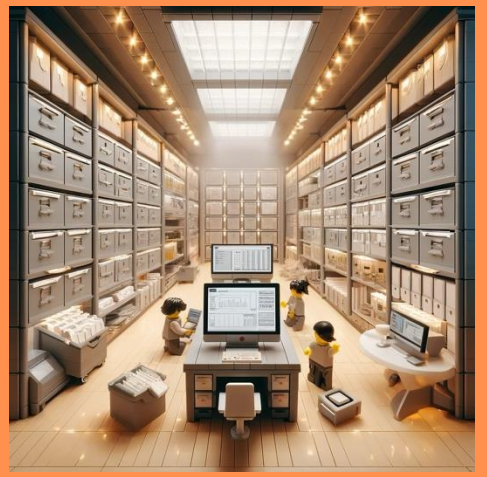
FAIR Data

Open ≠ FAIR

Open

Yes

No



FAIR & Open

No

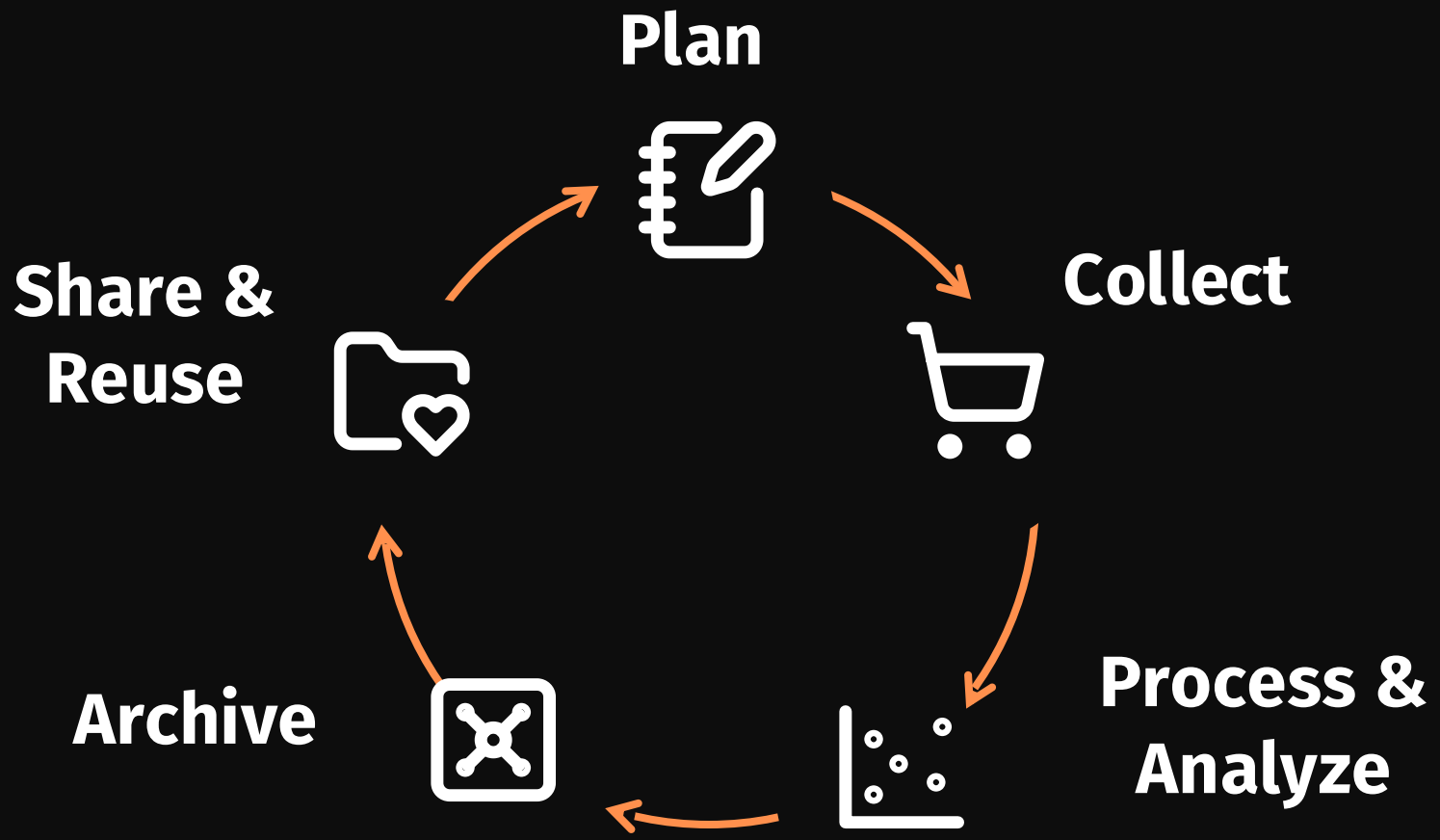
Yes

FAIR

What do you do with your data when you finish your project?

Research Data Management can go wrong

We can also do it better: efficient & sustainable



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

Community standards, funders, and laws require good RDM

Code of Conduct for Research Integrity

Funders

NWO

European Commission

GDPR

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

Good RDM benefits you and the research community

Less likely to lose data

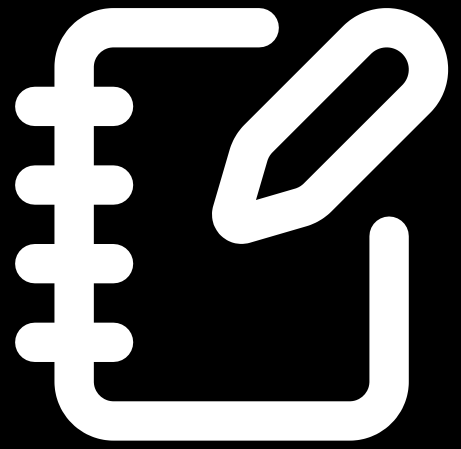
Recognition

Protection for integrity

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

How to plan for your research project

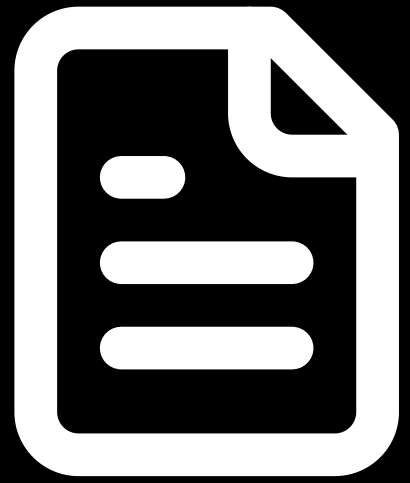


Data Management Plan

A living document for the research project

Funding agencies require data management plans, reviewed by a data steward

At TU/e, you can prepare a Data Management Plan on the Research Cockpit (more on this later)



What kind of data do you have?

A. Data from human participants

B. Sensitive Data

C. Data from external parties

Ethical review is required for all research involving human data at TU/e

*including anonymous / secondary data

ERB application can be submitted after creating a Data Management Plan on Research Cockpit

Research Cockpit: cockpit.research.tue.nl

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

Personal Data: Any data that can be traced back to a living person

Normal Personal Data

Special Category Personal Data

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

GDPR prohibits processing special categories of personal data except certain circumstances

- Racial or ethnic origin
- Political opinions
- Religious or philosophical beliefs
- Trade union membership
- Genetic and biometric data for identification
- Sex life or sexual orientation
- Health Data

Data Processing Impact Assessment (DPIA) may be needed

Contact your data steward if you plan to process special categories personal data

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

If you are getting data from an external party, it's best to have an agreement about data

Will your party agree to publish your data?

Permission to publish the Data	To meet the principles of findability, accessibility, interoperability, and reusability (FAIR), (part of) the Data may be published in anonymous form if requested by a scientific journal: ☐ yes ☐ no

Data Sharing Agreement Template TU/e Receiver

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

When writing a grant proposal, it's best to reserve funding for RDM-related costs

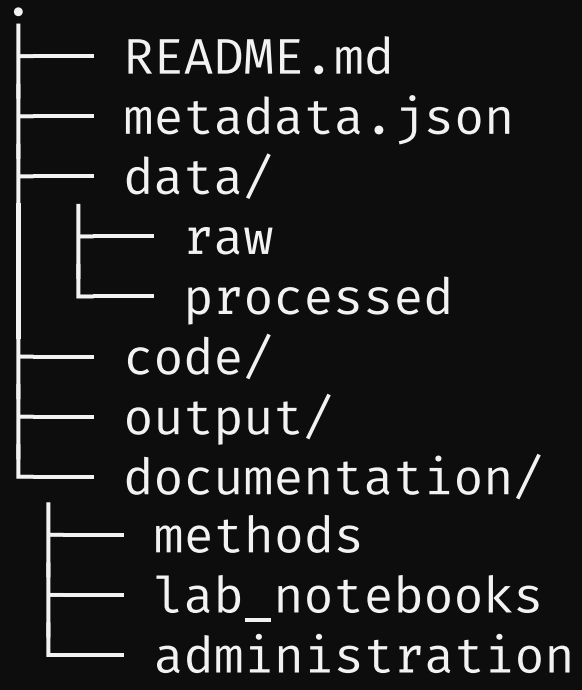
"Data management and making the data FAIR (assigning DOI, describing the data, assigning metadata to the data, etc.) is part of the ongoing research activities of involved researchers. Although we do not foresee any extra expenses, a budget of €5000 is reserved for RDM purposes."

Costs of data management | UU https://www.uu.nl/en/research/research-data-management/guides/costs-of-data-management

We’ll try out the Research Cockpit later today

Collecting Phase

Set your folder structure and document it in the README file



Make a README file to make your data human-friendly

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

You can complete a dataset form on 4TU.ResearchData or Zenodo, download the metadata file, and save it together with your data.

*You can also 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

No special characters. Whitespaces are discouraged.

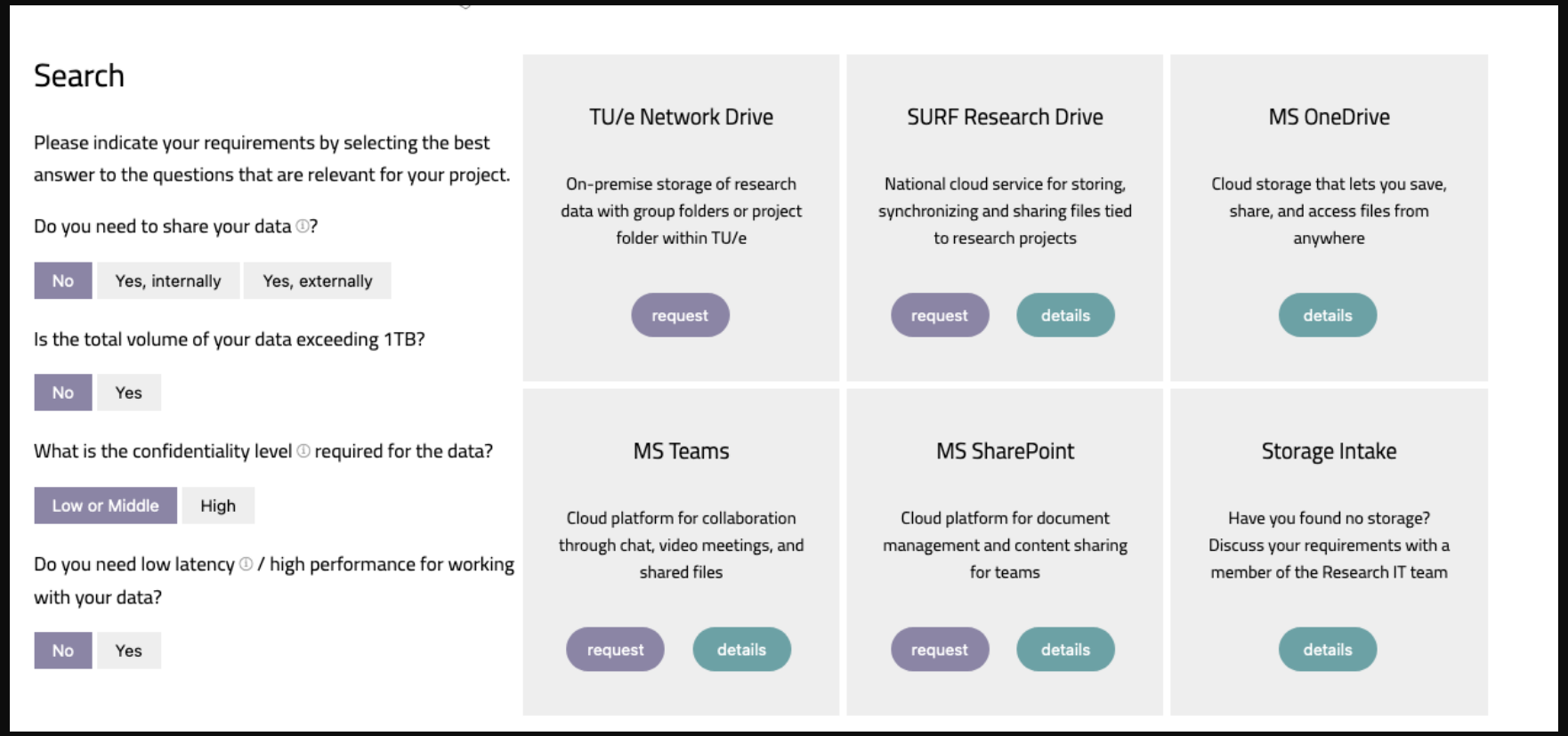
Case style: `camelCase`, `PascalCase`, `snake_case`, or `kebab-case`?

Use ISO date format: “2024-07-03”

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

Storage options at TU/e

Storage Finder at Solution Searcher



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

Where do you store your data?

A. TU/e Network Drive

B. SURF Research Drive

C. Teams / SharePoint

D. OneDrive

E. Other

TU/e Storage Options

	Security	Scalability
Network Drive (External sharing not possible)	Best	Best (~100 TB)
Research Drive	Good	Good (~10 TB)
Teams / SharePoint	Good	OK (~2.5 TB)

*OneDrive is recommended for student project only

You can encrypt your data using Cryptomator in a vault and store it in any storage solution

You can manage access using Cryptomator Hub

Learn more about Cryptomator Hub on the Research Cockpit:
https://cockpit.research.tue.nl/helpcenter/research/portal/11/article/2452390165

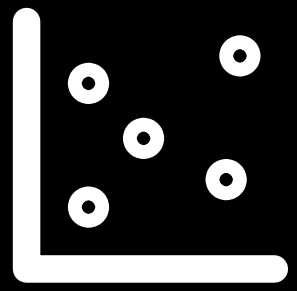
Only sending data one-time?
Then use SURFfilesender

Remember to turn on encryption

SURFfilesender: filesender.surf.nl/

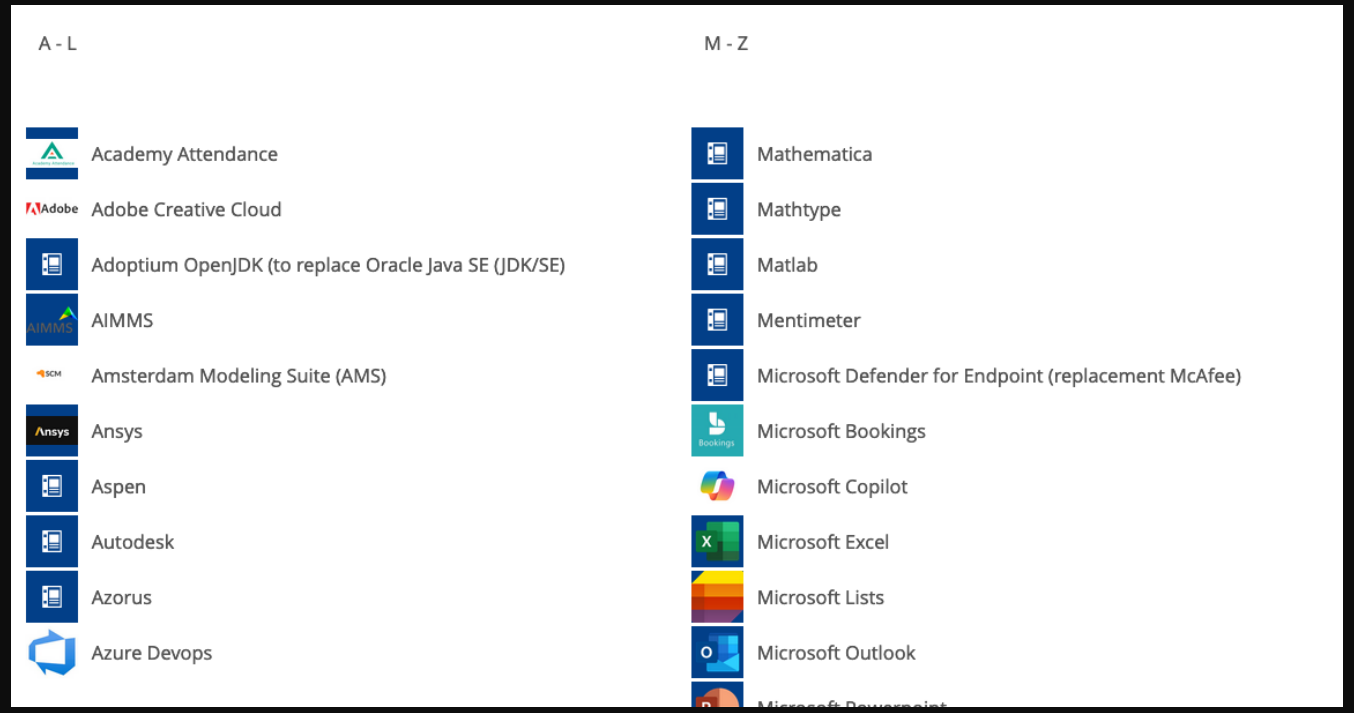
Do not send data over email!

Processing & Analyzing Data



You can find available tools on Topdesk

TU/e Software | Topdesk



The recommendation is to use open-source tools, whenever possible.

Python

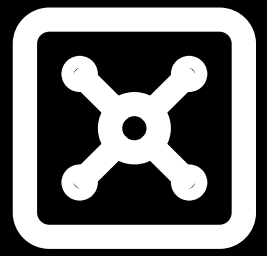
R

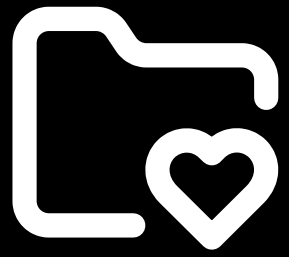
If the original dataset is in a proprietary format, convert it to an open file format.

e.g., text file, HDF5, NetCDF

Do you need to use proprietary tools?

Any opportunities for using open tools?

Archiving 

Share & Reuse 

Archiving: keeping data to demonstrate the integrity of research

Usually, each publication requires an archival package

Archival package is only accessed when it’s needed

Photo by Jason Dent on Unsplash

Sharing: Creating a publication data package to share your data

The goal is to share data with future data users (can be you)

	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.

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

Your discipline may have a specialized data repository. Visit re3data.org to find out.

When sharing your research, assign an open license, for example

Data, documentation, non-code materials

CC BY 4.0

Code

MIT

https://chooser-beta.creativecommons.org/
https://choosealicense.com/

Share your data as openly as possible as closed as necessary

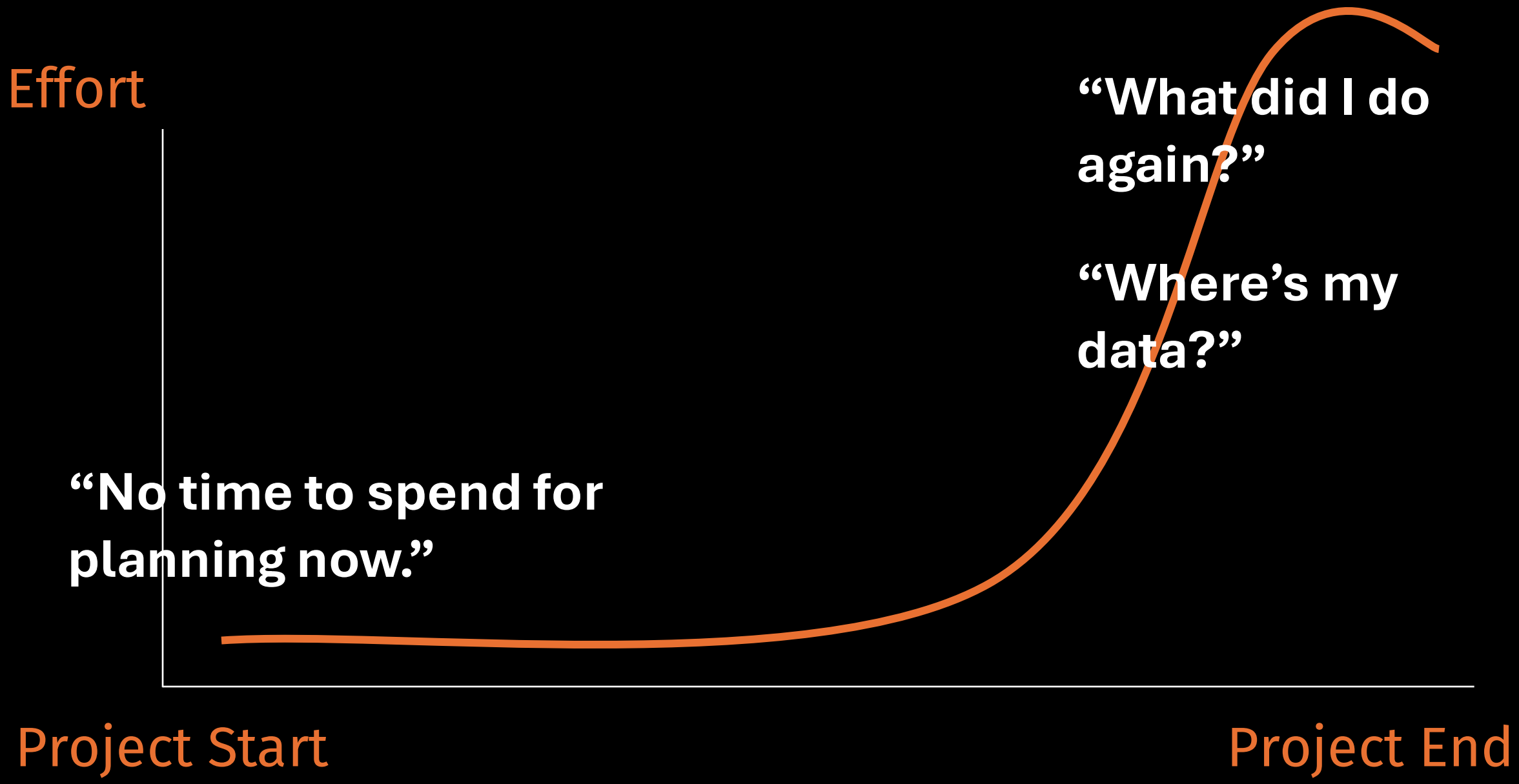
That said, look out for limitations related to privacy, intellectual property, or data agreements

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

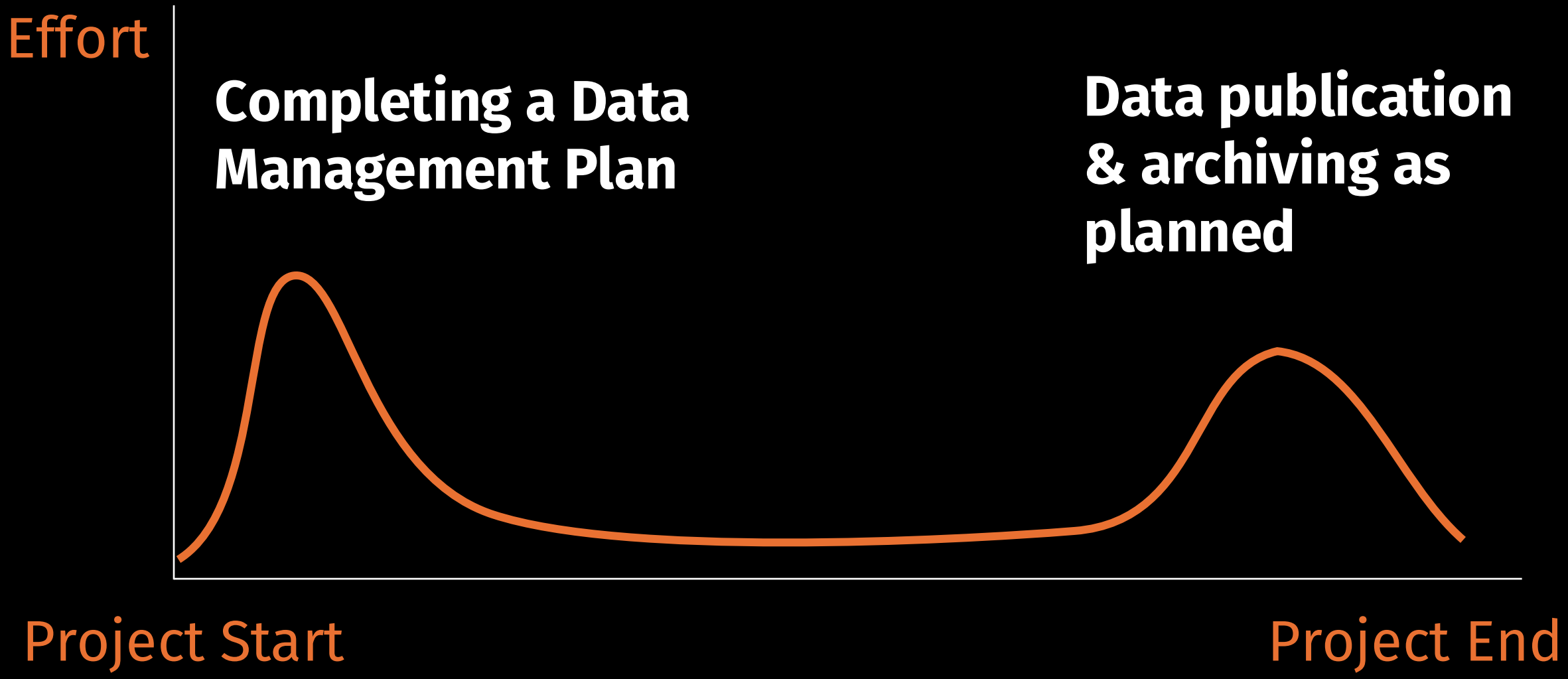
The future user of your data may be you

Managing Research Data can be overwhelming...

It’s best to plan ahead to avoid chaos at the end


Effort
“What did I do again?”
“Where’s my data?”
“No time to spend for planning now.”
Project Start
Project End

Ideal:
Start with the end in mind

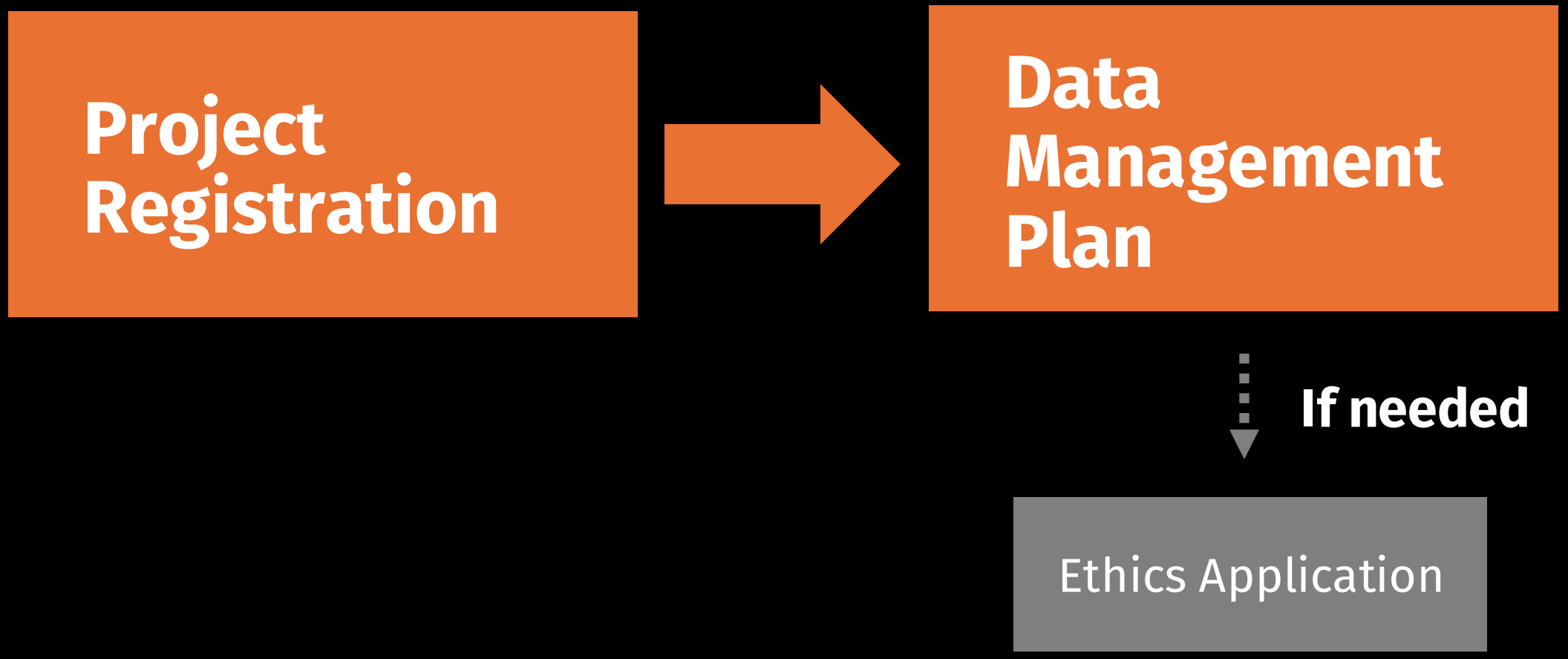

Effort
Completing a Data Management Plan
Data publication & archiving as planned
Project Start
Project End

Research Cockpit

https://cockpit.research.tue.nl

Let’s try out the Research Cockpit!

General workflow



Project Registration

1. Visit Research Cockpit (cockpit.research.tue.nl)
2. Find “Data Management Plan”
3. Find ”Create a Data Management Plan
4. Fill in Research Project Registration
5. Click “Send”

Research Project registration

Required fields are marked with an asterisk *

Project Name *

Use e.g. the name as defined in the research grant for the project.

Test Project Registration 2025-06-18

Department *

Specifying your department allows us to connect you to the right personal departmental support staff.

Applied Physics and Science Education (AP&SE) ×

What stage is your research project in now?

Please select one or more options

- [x] Plan my Research: Tasks like making a Data Management Plan (DMP), getting Ethical Review approvals, or setting up collaboration agreements.
- [ ] Do my Research: Activities like collecting, storing, analyzing or computing my data.
- [ ] Share my Research: Writing, publishing, and saving your research paper.

Do you already have an externally approved Data Management Plan (DMP) for this project? *

In the next step, you'll either receive a form to attach your externally approved DMP (this is not the pre-award data management paragraph), or you'll receive a blank DMP template to start with.

- ( ) Yes
- (•) No

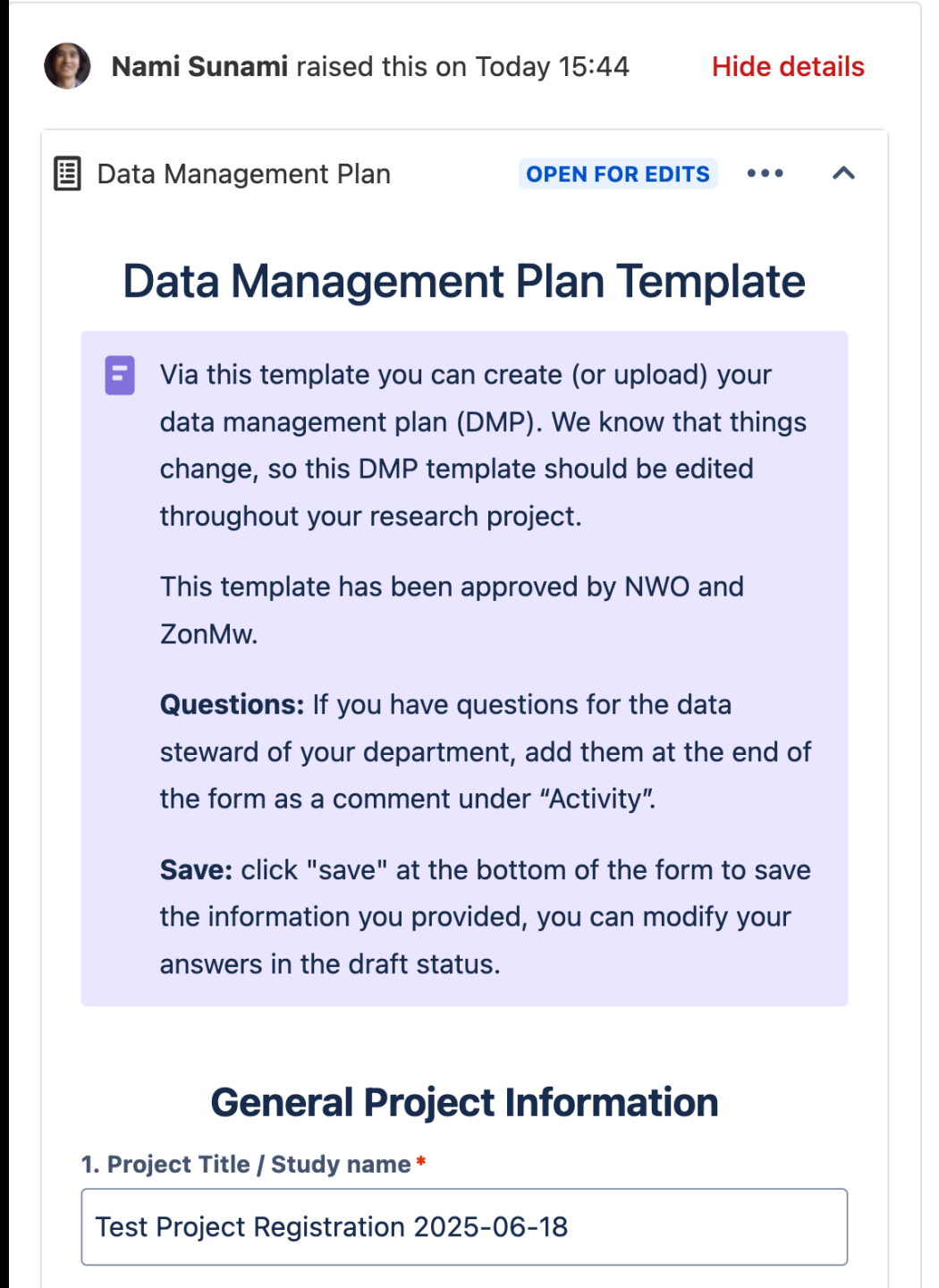
Data Management Plan (DMP)

1. Go to your avatar -> Requests
2. Find your Draft DMP
3. Fill in the information in the DMP

You can share your DMP with your collaborators

If you no longer need the DMP, please retract it.

Retract my DMP



How was your experience with the Research Cockpit?

Don’t hesitate to reach out to your Data Steward

Liz Guzman-Ramirez
Coordinator

Angela Aleksovska
EE

Davide Nardi
M&CS

Jay Nair
IE&IS

Jonathan Genora
ID

Lucia Forrová
BE

Neda Norouzi
BME, ME

Nami Sunami
CE&C, APSE

rdmsupport@tue.nl

rdm.tue.nl

Data Stewards | SharePoint
(login required)

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