Research Data Management towards FAIR

Nami Sunami (Data Steward) orcid.org/0000-0001-5482-8370
Library and Information Services (LIS), TU Eindhoven

Sunami, N. (2024, July 3). Research Data Management towards FAIR. ResesarchEquals. https://doi.org/10.53962/kr9t-rzak

Managing research data is challenging

Data Stewards can help you

A Data Steward can help you with ...

Responsible Science

Privacy & GDPR

Agreements

Ethical Review Board (ERB) applications

Open Science & RDM

FAIR Data

Sharing

Sustainable Workflow

Open Science & RDM

TU/e Data Steward Team

Floor Luub
M&CS

Neda Norouzi
BME, ME

Lucia Forrová
BE

Mariana Oshima Menegon
ID

Edyta Cios
IE&IS

Nami Sunami
CE&C, APSE

Angela Aleksovska
EE

Liz Guzman-Ramirez
Coordinator

What’s Research Data Management?

Research Data Management refers to the way you collect, analyze, store, share, archive and publish research data.

Why do we want to manage research data better?

Managing research data well makes your science sustainable.

Improving research data management is a key to Open Science.

Managing data well will

benefit the world

benefit you as a researcher

For the world:

You can share about your research better, maximizing your contribution to the world.

For you:

You can easily go back and build on to what you did in the past, helping yourself to do science better.

As a research community & society, we hold each other accountable to manage data well.

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

What do we mean by “doing a good job” managing data?

FAIR data principles can guide us to define a good RDM.

FAIR stands for:

Findable

Accessible

Interoperable

Re-usable

https://www.go-fair.org/fair-principles/

Findable

Findable data means that humans and machines can find your data easily.

Findable data has persistent identifier & metadata, and is indexed in a searchable resource.

Persistent identifier (PID) is a globally unique ID assigned to any object.

DOI: Digital Object Identifier

ORCID: Open Researcher and Contributor ID

ROR: Research Organization Registry

Metadata refers to

data about data.

Like a label on a can.

Such as, title, author, data collection date, and DOI

Image generated by Microsoft Co-Pilot

You can register your data to online registry (data repositories) & assign a DOI & metadata.

4TU.ResearchData

Zenodo

You can choose to share only metadata. Making data findable does not mean making the data accessible.

Accessible

Accessible data means that

there's a way to access the data.

Metadata is best to be kept

always accessible.

When needed, require

authentication or authorization

for accessing the data.

You can do so in major data repositories.

Authentication means verifying the identity of the person.

Authorization means assigning an access right.

You may have access to data.

But can you connect with other pieces of knowledge?

Interoperable

Interoperable data means that the data can be integrated with other data.

Use common & sustainable file formats, preferably open ones.

File Formats | How to FAIR: https://howtofair.dk/how-to-fair/file-formats/

Use a machine-readable schema to document your metadata & data

Data repositories can export metadata to different schema.

Common schemas include JSON-LD, XML, RDF

Use a structured vocabulary to link concepts.

Structured vocabulary is a machine-actionable, curated dictionary

Examples include: WikiData

Link your data with other data using a persistent identifier (such as DOI).

You now understand the data.

But can you use it to create new knowledge?

Re-usable

Re-usable data means that others can use the data for a new purpose and can replicate your data.

Assign an accessible license to allow re-use

CC-BY 4.0

MIT

https://www.go-fair.org/fair-principles/r1-1-metadata-released-clear-accessible-data-usage-license/

FAIR ≠ OPEN

Access-restricted data can be FAIR.

Open data is not always FAIR.

Now we know about FAIR principles.

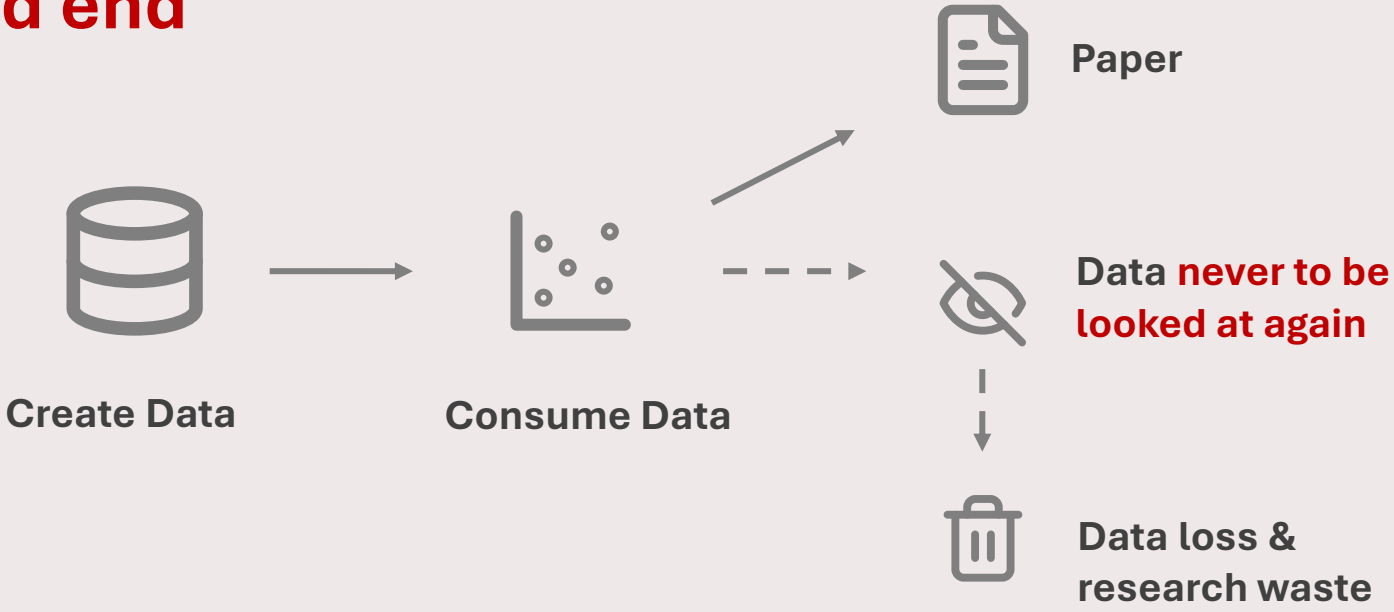
But practicing FAIR principles is a journey.

Where can we start?

Where to start doing RDM towards FAIR

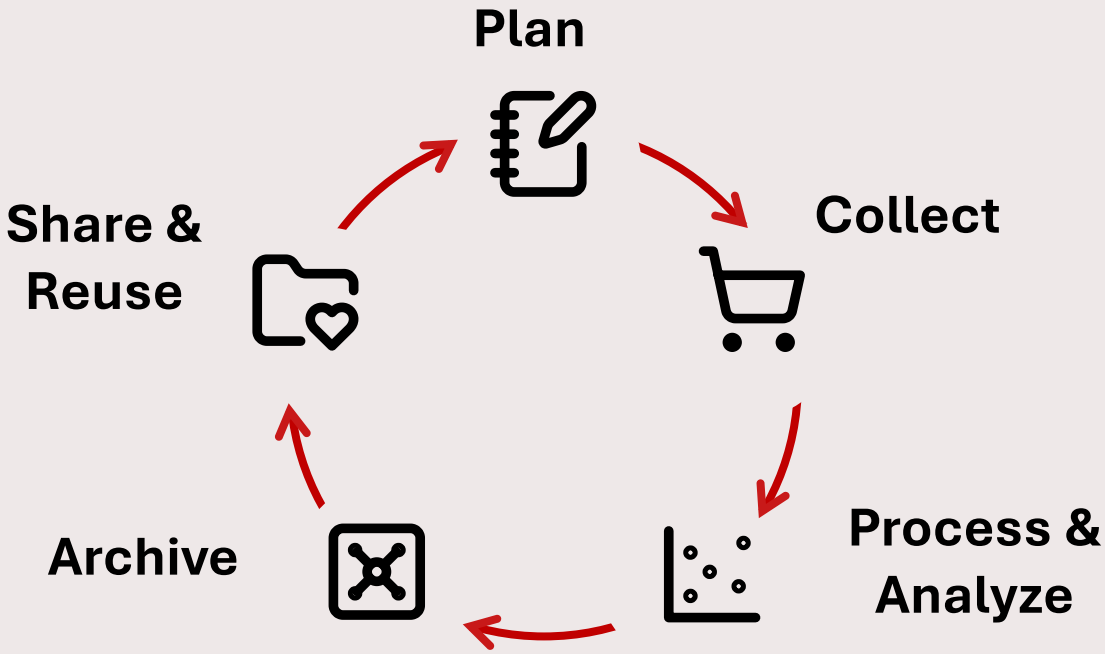
First step is to think of research data in a life cycle.

Traditionally, we thought research data as something to be consumed, with a dead end



Instead, we want to think of research data as something to be nurtured in a life cycle, to be re-used.

Research data lifecycle

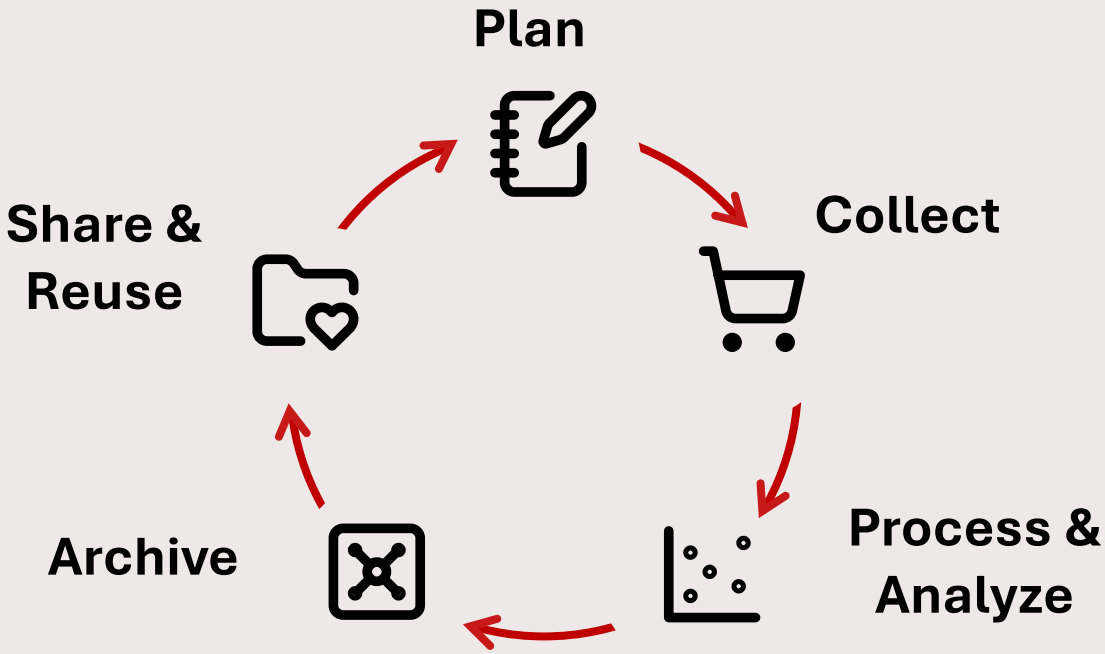


Adopted from RDMkit: https://rdmkit.elixir-europe.org/data_life_cycle

The end of one’s research becomes another’s beginning.

What are key action areas for each step?

Research data lifecycle



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

Plan

Completing a data management plan (DMP) can highlight what's needed for your data.

Writing a good plan takes time, but it will pay off in a long run.

It will save time and money.

A data management plan is a living document. Don’t let your perfectionism deter you.

A typical data management plan includes

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

How much is the RDM cost

Funders such as 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

If your data includes personal data,

you may need additional precautions.

Personal data is “any information relating to an identified or identifiable natural person.”

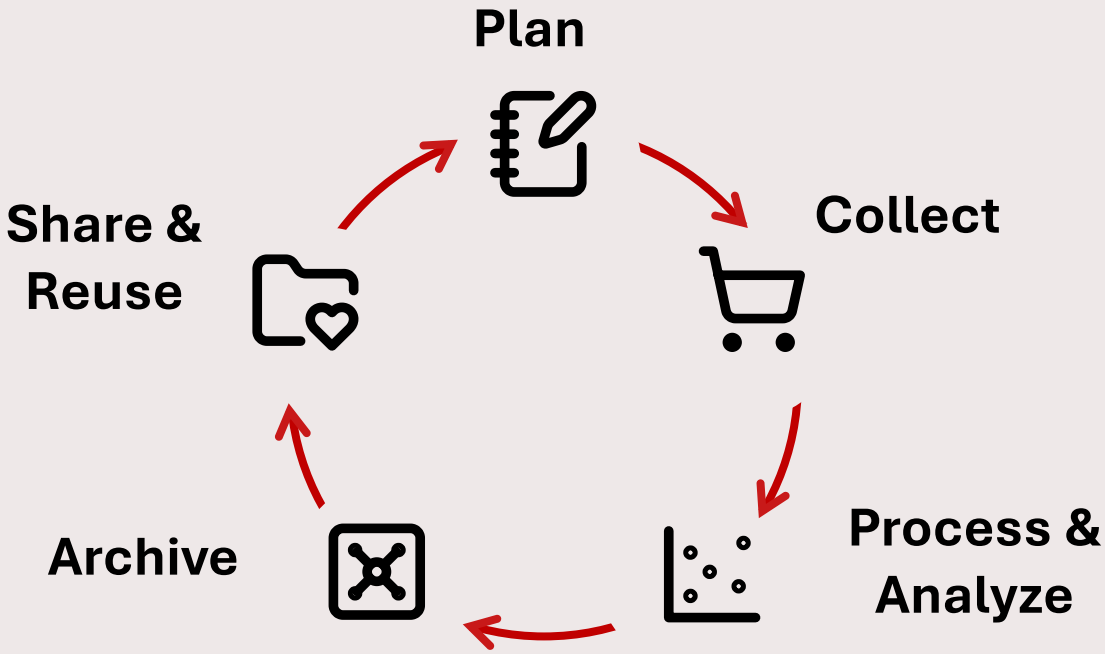
Apply techniques such as data minimization, pseudonymization, and anonymization

fpf.org/blog/a-visual-guide-to-practical-data-de-identification/

If you are getting data from a third-party, you may need an agreement

Such as data sharing agreement

Research data lifecycle

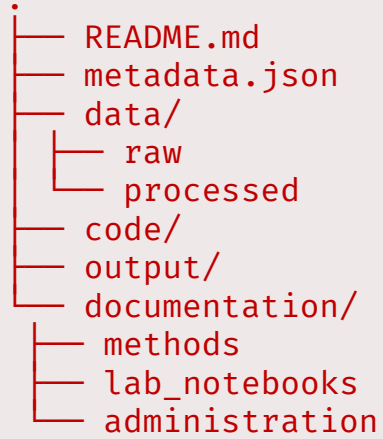


Adopted from RDMkit: https://rdmkit.elixir-europe.org/data_life_cycle

Collect

Start your research project strong,

with your folder organized.



A README file is always a good idea for us humans

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 is also good—but for machines.

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

You can also use metadata generators, such as DataCite Metadata Generator

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`?

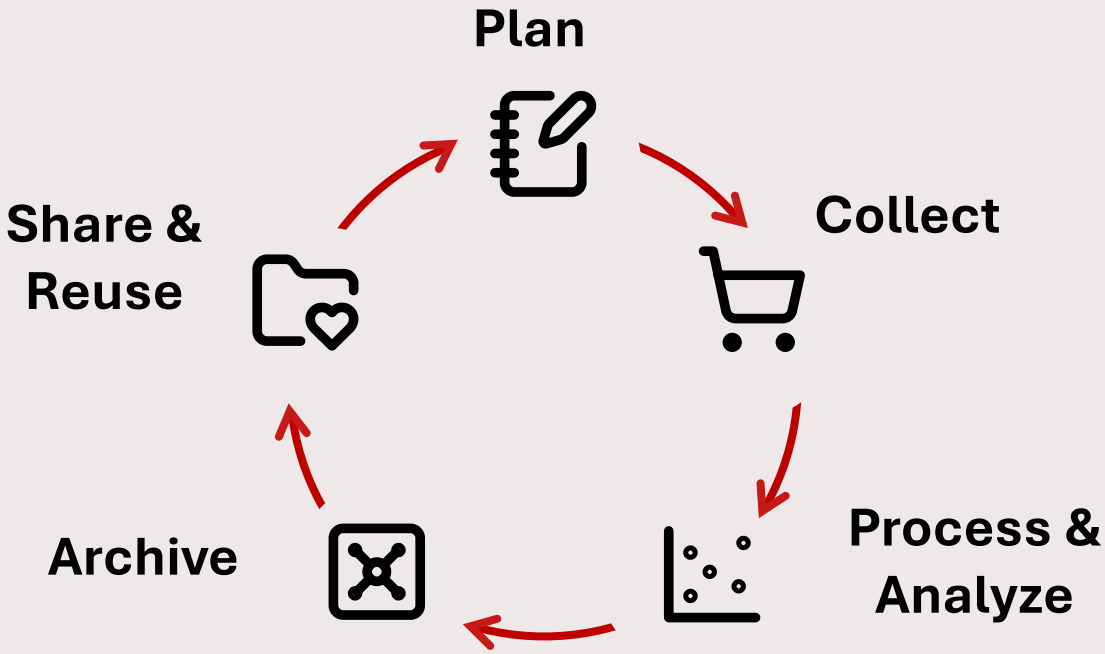
ISO date format is good to use: “2024-07-03”

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

Now you collected data.

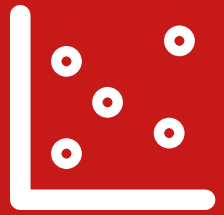
How will you process and analyze?

Research data lifecycle



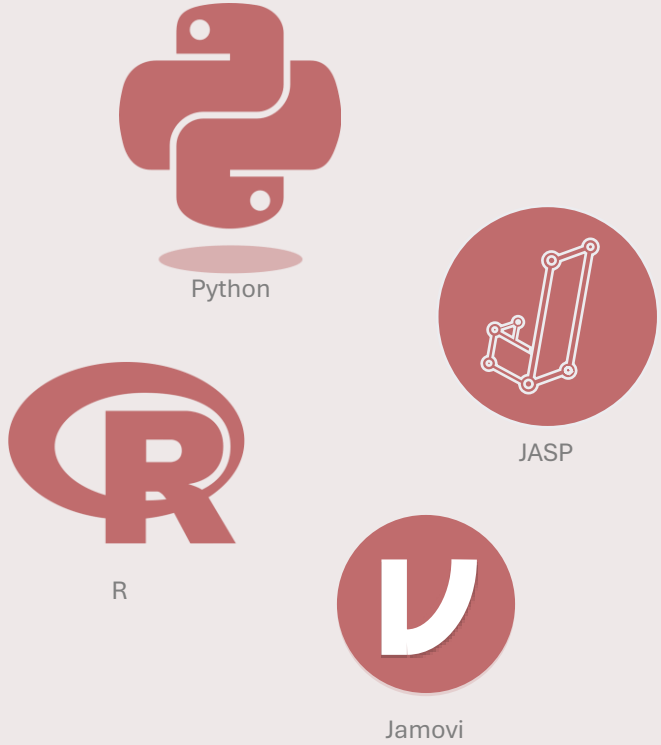
Adopted from RDMkit: https://rdmkit.elixir-europe.org/data_life_cycle

Process & Analyze



Use open-source tools,

whenever possible.



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

Archive

Archiving data means storing data to demonstrate the integrity of research

Typically, each publication requires an archival package

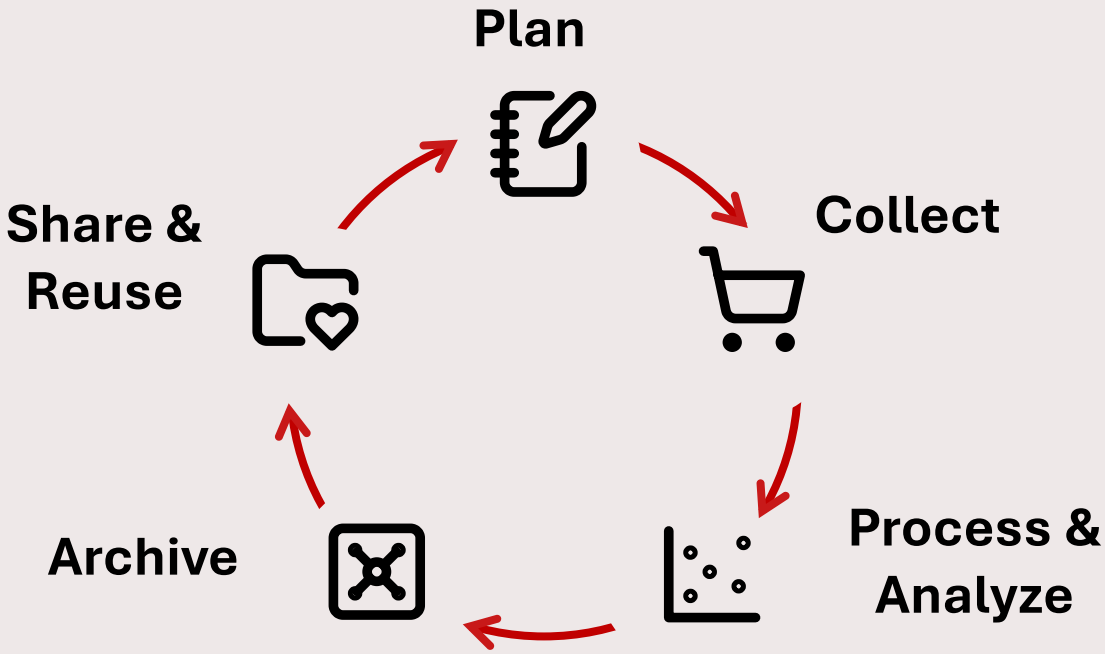
You can prepare an archival data package to include information needed to demonstrate the integrity.

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

An archival package is meant to be accessed only when it's needed—for example, an integrity audit.

When you want to share your data, you can prepare a publication data package.

Research data lifecycle



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

Share & Reuse

Often, publication package is different from archival package

The goal of a publication package is to share data with the world.

The goal of an archival package is to demonstrate integrity.

By default, anyone can access a publication package.

For an archival package, the default is restricted access.

	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

It’s best to share 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

Now that you shared the data, someone else may use the data.

And it could be you in the future.

What a ride!

FAIR stands for:

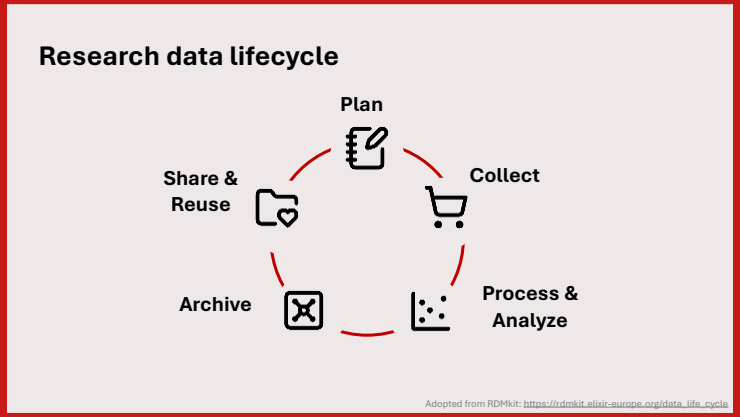
Findable

Accessible

Interoperable

Re-usable

https://www.go-fair.org/fair-principles/



Questions?

Thoughts to share?

rdmsupport@tue.nl

Footnotes

- Icons in this slide deck are from Lucide, open-source icon library (https://lucide.dev/icons/rainbow)
- This slide deck is licensed under CC BY 4.0