Introduction to Research Data Management

Linde Tuybens

ADS Doctoral Day 2024

Research Data Management Team

- Part of the **Research Affairs Office** at **RIVA – Research Innovation & Valorisation Antwerp**

- Offer advice on all things related research data management, including data management plans, Nagoya, and Open Science and Open Data

- rdm-support@uantwerpen.be

- o **Siham Benramdane**
- o **Jef Peeters**
- o **Linde Tuybens**
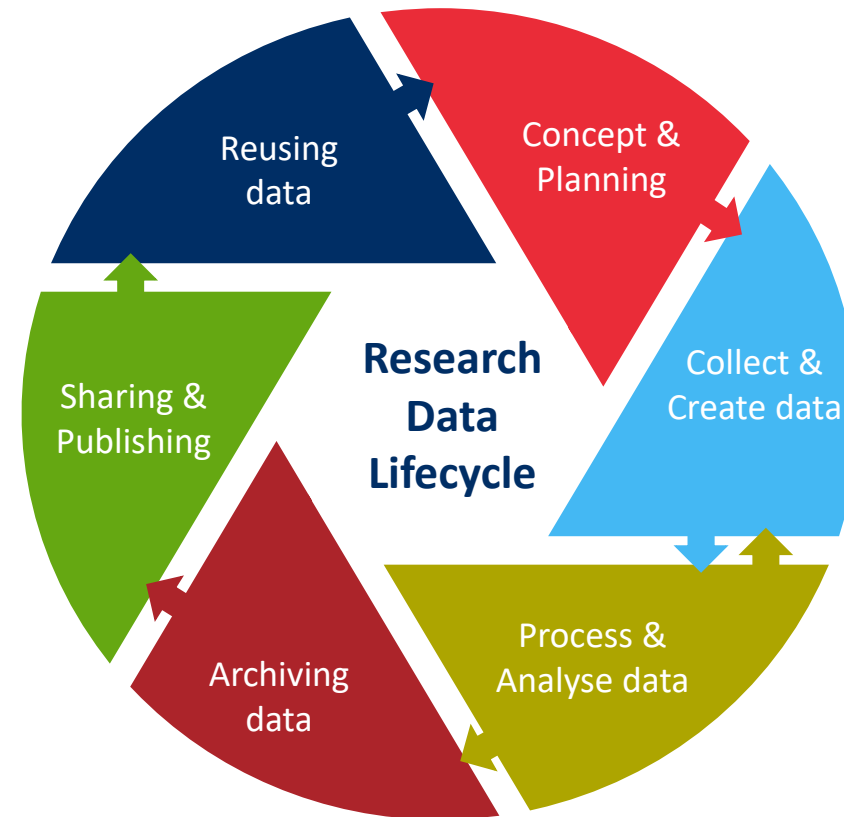- o **Dunya Nasser**

Research Data

*"Research data are all **digital or physical data** - regardless of the manner in which these data are collected or stored - used or analyzed to support research findings, validate research results or underlie a scientific reasoning, discussion or calculation in the study. Research data cover the entire spectrum of **raw data to processed and analyzed data** included or discussed in a publication. These data can be **generated data, derived or composite data,** as well as **self-generated data and data provided by third parties**."*
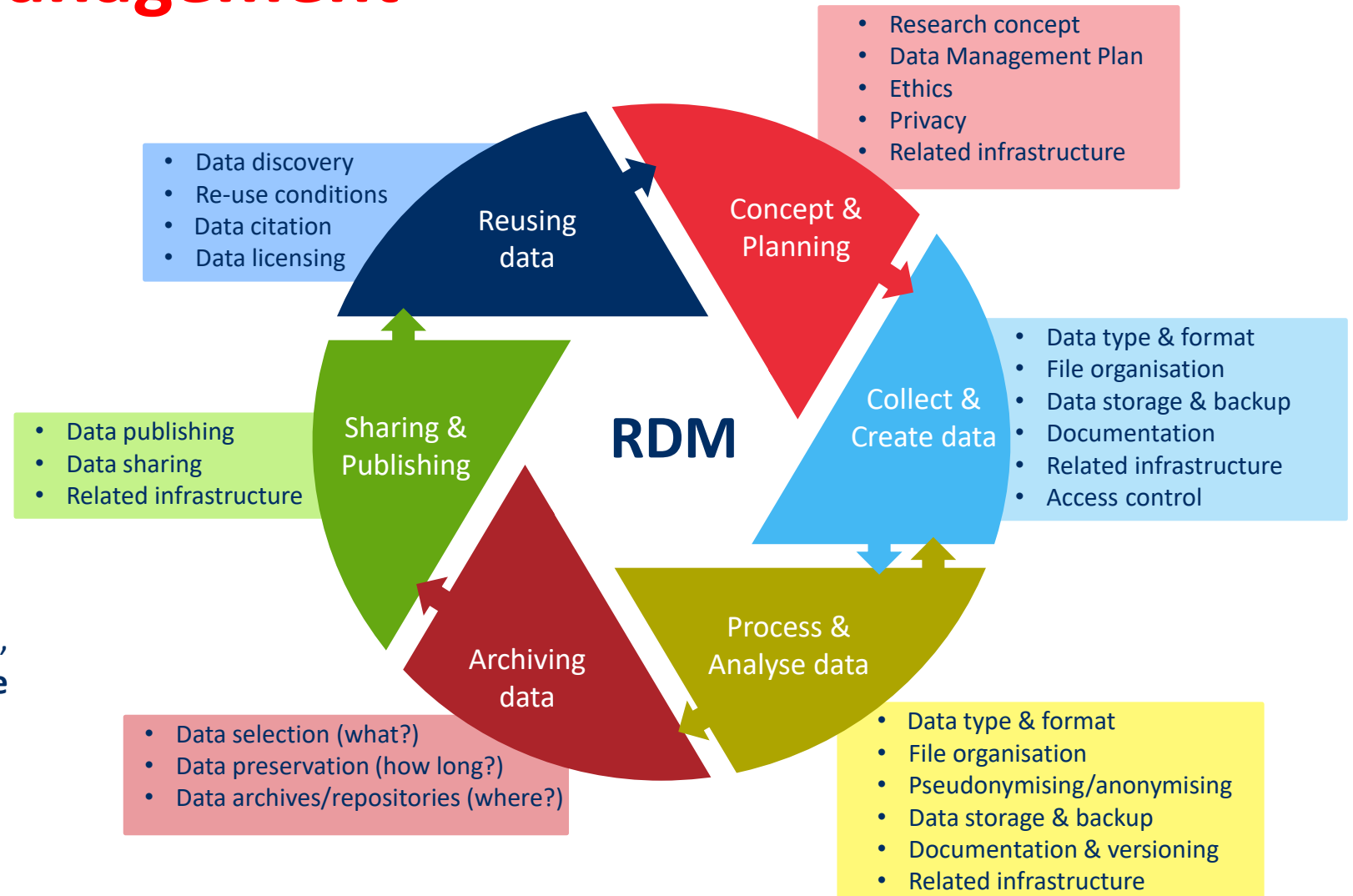
RDM Policy of University Antwerp

Research Data Management

Research Data Management

> The compilation of many small practices that make your data **easier to understand**, **less likely to be lost**, and **more likely to be usable** during a project or ten years later

Concept & Planning
- Research concept
- Data Management Plan
- Ethics
- Privacy
- Related infrastructure

Collect & Create data
- Data type & format
- File organisation
- Data storage & backup
- Documentation
- Related infrastructure
- Access control

Process & Analyse data
- Data type & format
- File organisation
- Pseudonymising/anonymising
- Data storage & backup
- Documentation & versioning
- Related infrastructure

Archiving data
- Data selection (what?)
- Data preservation (how long?)
- Data archives/repositories (where?)

Sharing & Publishing
- Data publishing
- Data sharing
- Related infrastructure

Reusing data
- Data discovery
- Re-use conditions
- Data citation
- Data licensing

RDM

What's in it for you? The benefits of good RDM

You want to protect the valuable data you spent so long collecting/processing

Oh no, my laptop crashed and I have no backup!

You may want to work with the data later in your career

What is the difference between file feb2018v1 and feb2018v2?

You want to work efficiently

What was the name of the file containing my notes?

You care about scientific integrity, guaranteeing reproducibility of the research

A reviewer remarks that my conclusion is not supported by the data

or the sources I've used. Can I retrace my steps?

You want to share your data with others

I see a great chance for collaboration! But my data are a mess...

Research data lifecycle

Research Data Management
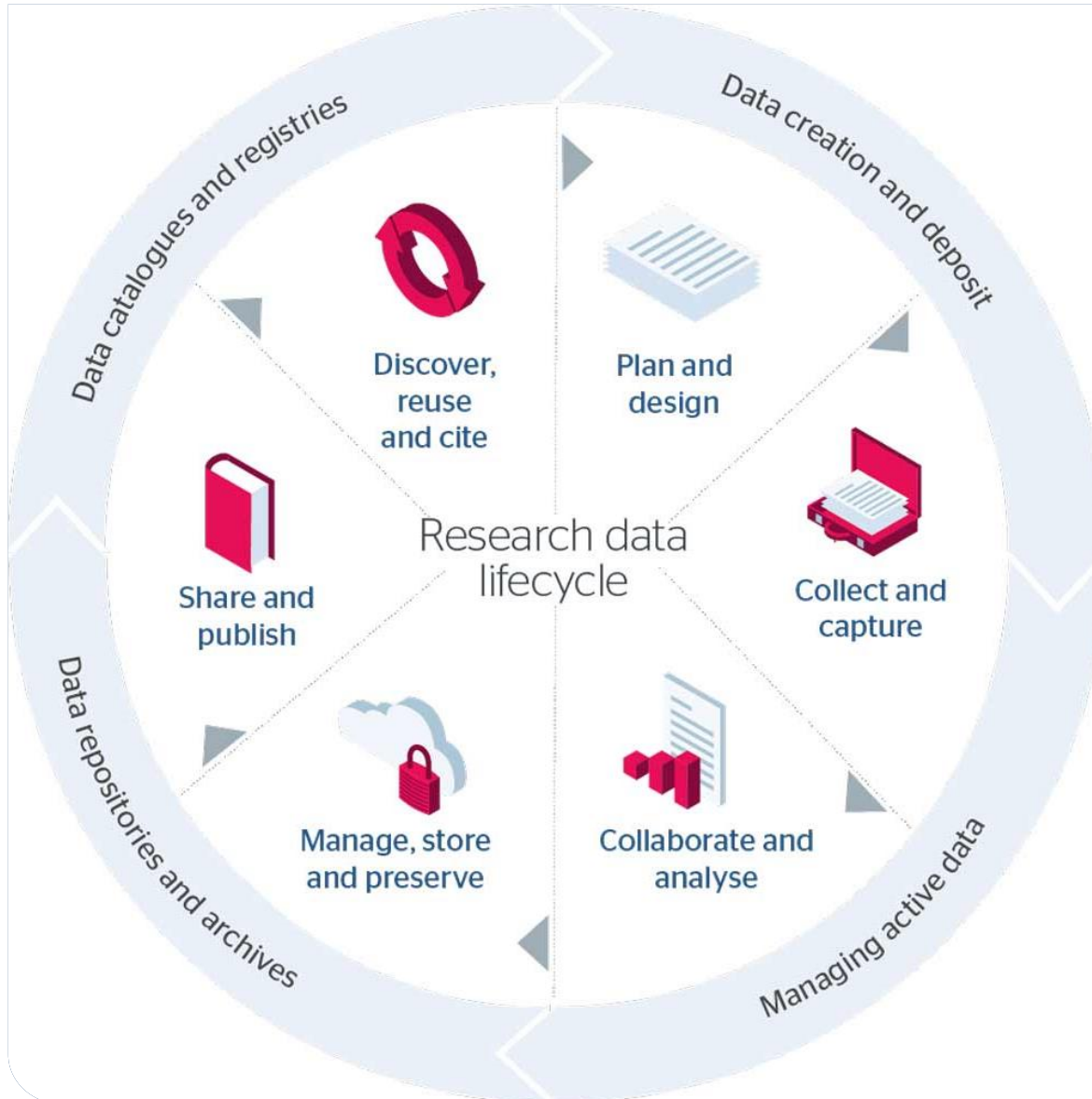
- Manage **risks** (legal, financial, technical, ethical, …)
- Ensure **data quality**
- Safeguard **confidentiality**, **integrity** and **availability** of data
- Ensure (re-)**usability**
- **Transparency**

Sidenote: Nagoya Protocol & ABS regulation

- Related to the **access and use of non-human genetic resources** and related traditional knowledge **collected abroad**

- https://nagoya.vlir.be/

- If there is a **chance** your research is subject to the **Nagoya Protocol,** contact us **as soon as possible** (rdm-support@uantwerpen.be)

University of Antwerp

Data Management Plan

Formal document describing how research data will be processed during the research lifecycle

University of Antwerp

DMP Contents

- **Contextual information**

- **Data description**

- **Ethical and legal issues**

- **Documentation and Metadata**

- **Data storage and backup during project**

- **Data preservation after the project**

- **Data sharing and reuse**

- **Responsibilities**

Data Management Plan

- Mandatory for most funders (EU, FWO, BOF, IOF, VLAIO, Belspo, …)
- Initial version usually required within 6 months, with exceptions
- Submitted to:
  - rdm-support@uantwerpen.be (BOF, FWO, IOF, VLAIO)
  - Funder (EC, Belspo)
- Living document, some funders require updates at evaluation stages
- Final version to be submitted along with the final report

University of Antwerp

Overview of funder RDM requirements from proposal to end of project

Funder	Proposal	Initial DMP (month 6)	Final DMP (end of project)	Data preservation	Data publishing
FWO	5 RDM questions in FWO application form	Submit to rdm-support@uantwerpen.be	Submit to FWO (e-portal) with final report	Min. 5 years	Advice: data linked to publications
BOF/IOF	5 RDM questions in application form	Submit to rdm-support@uantwerpen.be	Submit to rdm-support@uantwerpen.be	Min. 10 years	Advice: data linked to publications
VLAIO	5 RDM questions	Submit to rdm-support@uantwerpen.be	Submit to rdm-support@uantwerpen.be	Min. 5 years	
BELSPO	5 RDM questions (provisional DMP) in grant application	Submit to BELSPO	Submit to BELSPO	Long-term in repository	All data and metadata in certified and trusted repository, as open as possible
Horizon Europe	RDM for FAIR data and Open Science practices	Submit to EC portal	Submit to EC portal	Long-term in repository	All data and metadata in certified and trusted repository, as open and early as possible

University of Antwerp

Based on funder requirement table of KU Leuven RDM website: https://www.kuleuven.be/rdm/en/guidance/funder

Guiding Principles

University of Antwerp

As open as possible, as closed as necessary

Why have Open Research data?

- **Reproducibility, replicability and integrity**
- **Efficiency**
    - Collaboration
    - Reuse
- **Public funding, public results**
- **Increase public trust in science**

University of Antwerp
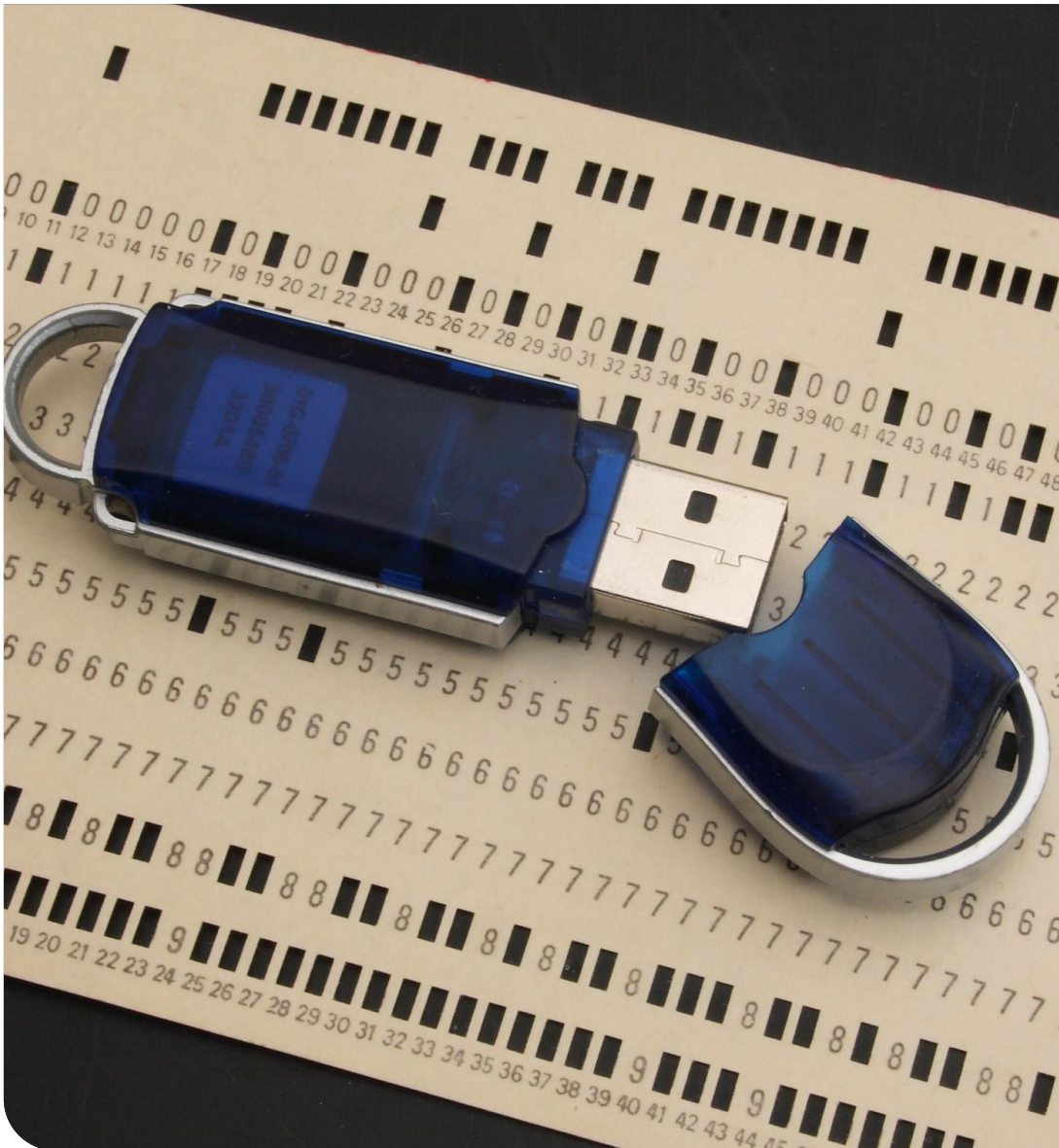
When not to open data?

- **Personal Data (GDPR)**

- **Ethical concerns**

- **Legal or regulatory restrictions (e.g. Dual Use)**

- **Third Party restrictions**

  - Intellectual property rights

  - Contractual obligations (NDA, licensing, …)

- **Legitimate commercial interests, including those of the university**

  - E.g.: patent filing

F indable   A ccessible   I nteroperable   R eusable
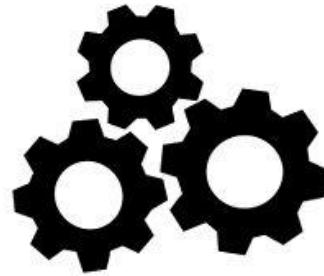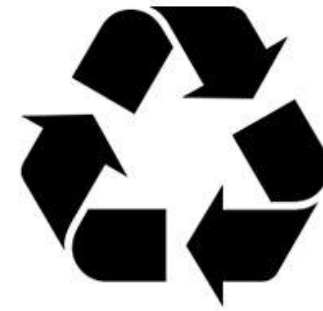
FAIR principles	Repository	Researcher
F1. (Meta)data are assigned a globally unique and persistent identifier	X	
F2. Data are described with rich metadata	X	X
F3. Metadata clearly and explicitly include the identifier of the data they describe	X	
F4. (Meta)data are registered or indexed in a searchable resource	X	

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
  A1.1 The protocol is open, free, and universally implementable
  A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
	X	

A2. Metadata are accessible, even when the data are no longer available	X	
I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	X	X
I2. (Meta)data use vocabularies that follow FAIR principles	X	X
I3. (Meta)data include qualified references to other (meta)data		X

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
  R1.1. (Meta)data are released with a clear and accessible data usage license
  R1.2. (Meta)data are associated with detailed provenance
  R1.3. (Meta)data meet domain-relevant community standards
	X	X
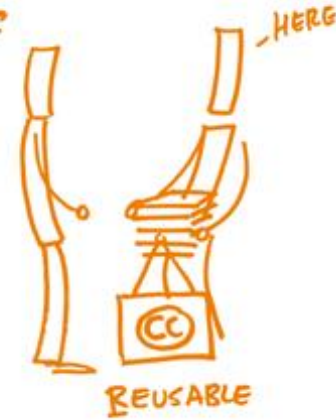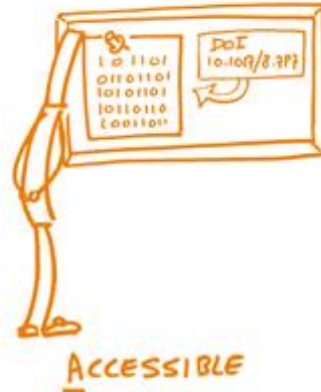
FINDABLE

- **Persistent Identifiers (PIDs)**
  - **DOI**
  - **ORCID**
  - **ROR**
- **Rich metadata**
  - **For search and filtering**
  - **Machine readable**

ACCESSIBLE

- **Access conditions are clear and understandable**
- **Access is not necessarily open, but communicated via a standard protocol**
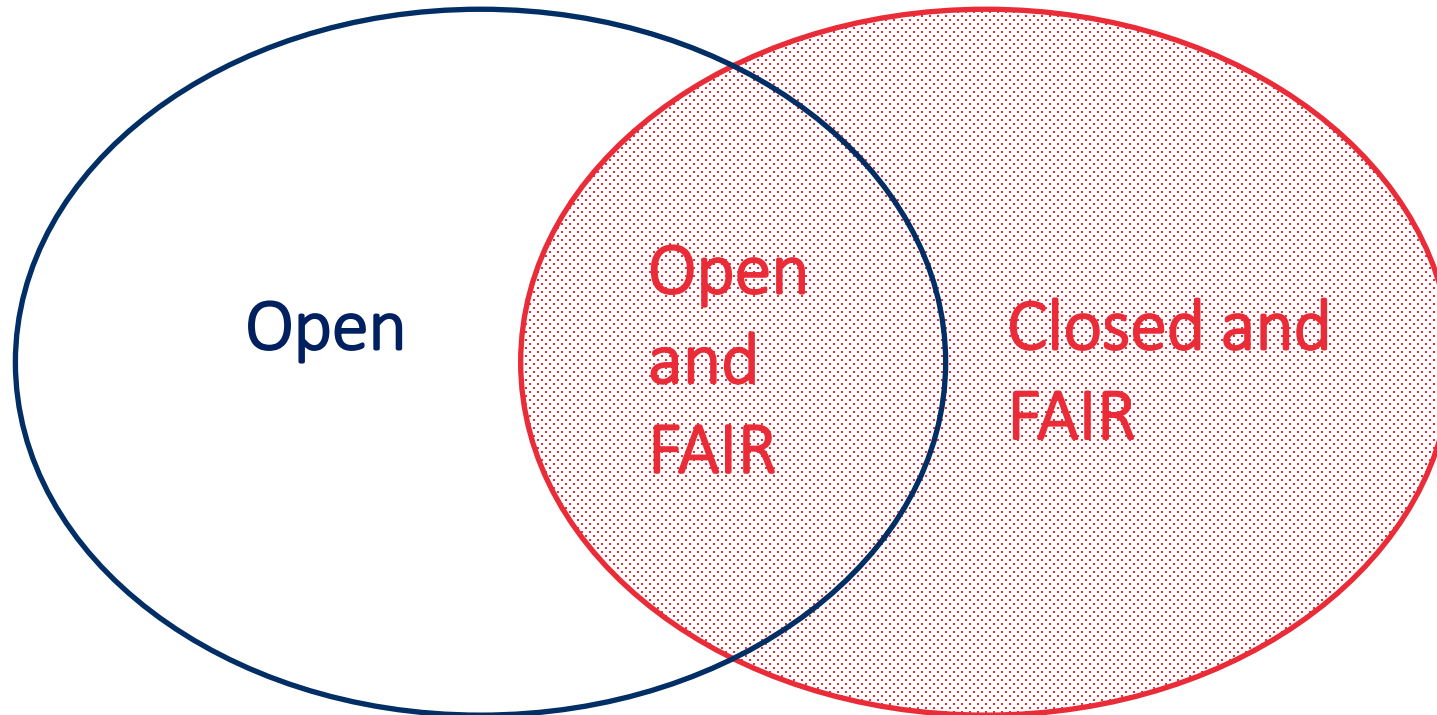- **Authentication, were necessary**
- **Metadata always available**

INTEROPERABLE

- **Standardised language (vocabularies)**
- **(Meta)data standards**
- **Open file formats**
- **Linked metadata**

REUSABLE

- **Documentation and metadata**
- **Usage license**
- **Community standards**

FAIR ≠ Open

Data sharing today?

All data in the publication

Data as supplementary material

Data made available on request

Data on project/personal websites

Through cloud storage providers

As an article in a data journal

In a disciplinary repository or infrastructure

In a general repository

In an institutional repository

Data sharing today?

All data in the publication ~~(crossed out)~~

Data as supplementary material ~~(crossed out)~~

Data made available on request ~~(crossed out)~~

Data on project/personal websites **?**

Through cloud storage providers ~~(crossed out)~~

As an article in a data journal

In a disciplinary repository or infrastructure

In a general repository

In an institutional repository
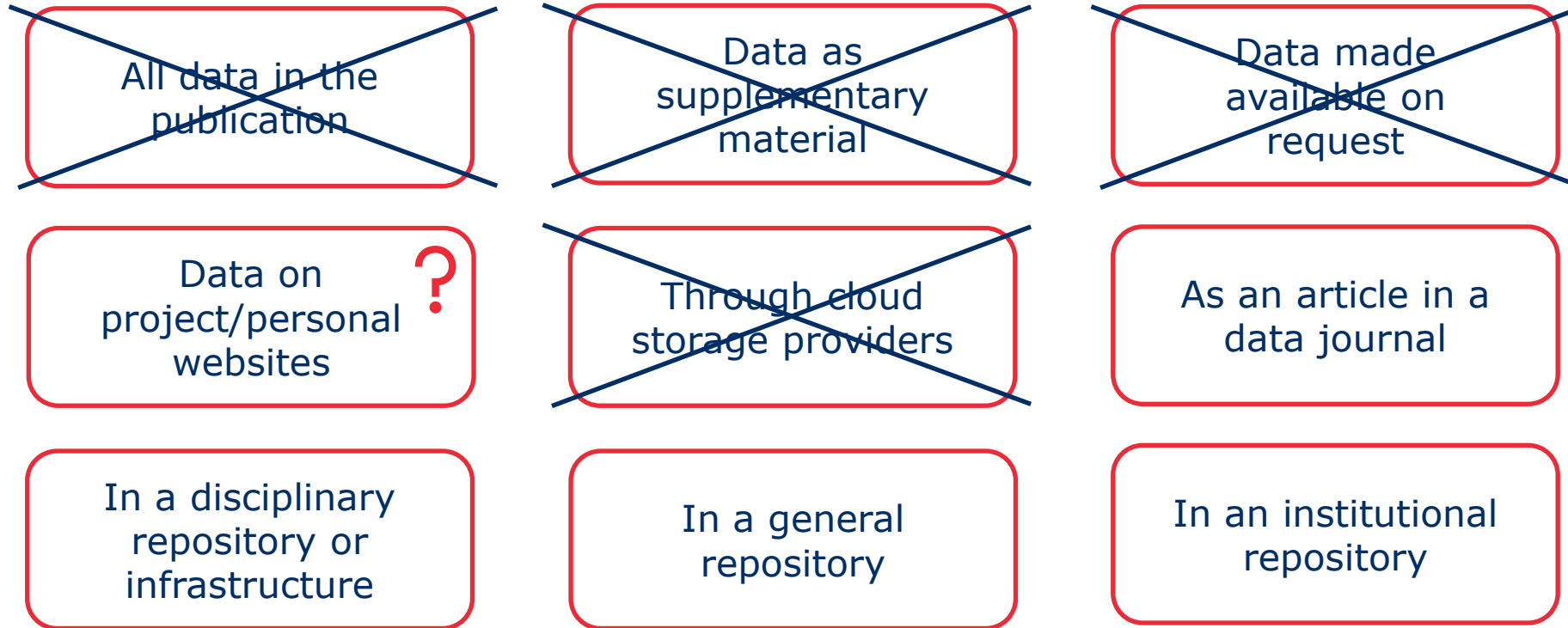
Preferred method for depositing data

- Publish in a domain specific or a generic repository or database

- Make sure the dataset has its own permanent identifier (doi, handle, …)

- Use repositories that link with ORCID

- Create at least a metadata record with study level metadata and add domain specific metadata if applicable

- Use standard formats and vocabularies

- Include documentation (README.txt)

- Include a license (CC-0 or CC-By)

- Include a data availability statement in your publications

- Register the dataset in IRUA

University of Antwerp

Data documentation

Data documentation	
WHY?	Needed for correct interpretation by any user (including yourself)
WHAT?	Data collection methodology, analytical and procedural information, data manipulations, measurement units, data characteristics, code definitions, terminology, data quality control measures, … Structure/organisation of data files, file naming conventions, any comments, …
HOW?	Use README.txt file, codebook, ELN or any other form of documentation. Keep these documents alongside your datasets.

How is your data analysis going?

Can't understand the data

… and the data collector does not answer my emails or my phone calls

That is terrible and so cruel ! Who is it, who collected the data ?

I did… 3 years ago
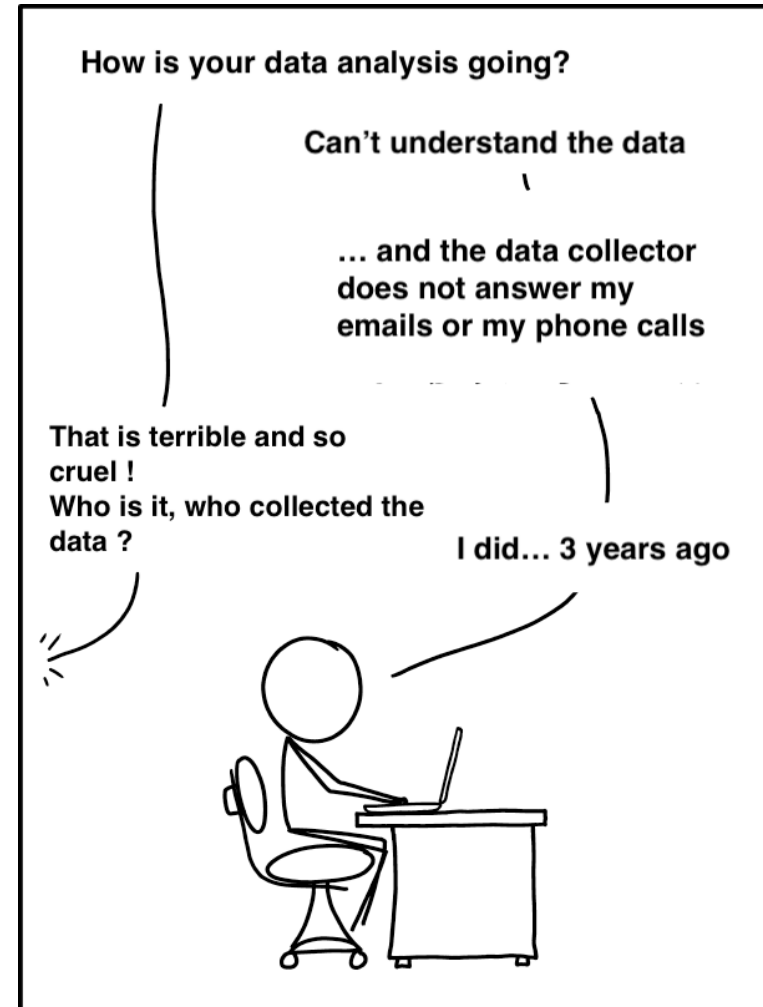
Your first collaborators are your future selves, be nice to them !

Data documentation and Metadata

	Data documentation	Metadata
WHY?	Needed for correct interpretation by any user (including yourself)	Needed to find your datasets quickly in repositories, databases, project websites, …
WHAT?	Data collection methodology, analytical and procedural information, data manipulations, measurement units, data characteristics, code definitions, terminology, data quality control measures, … Structure/organisation of data files, file naming conventions, any comments, …	Creator, title, PID (e.g. DOI), year, publisher, date, abstract, location, file format, page number, version, language, rights, …
HOW?	Use README.txt file, codebook, ELN or any other form of documentation. Keep these documents alongside your datasets.	Use metadata standards, that are usually determined by the tool or repository.

README

- A **README file** is highly recommended as an entry point
- Contains all documentation or references other documentation
- Should be the first file created when collecting or creating data

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

RDM: How to get started

- **Discuss RDM practices with your promotor**

- **Set up your ORCID if you do not have one (https://orcid.org/)**

- **Add to it to your account on our institutional repository (instructions)**

- **Check out our support pages on RDM**

- **Make use of our DMP authoring tool (https://dmponline.be)**

- **Check out our ADS course on DMP**

- **Reach out if you have any RDM related questions (rdm-support@uantwerpen.be)**

University of Antwerp

DMPonline.be

- **Web-based tool** for data management planning
    - Originally developed by the DCC – modified by DMPbelgium consortium
    - Owned by Belnet
- **DMP templates + guidance**
- Does not check or validate your answers

- **Tutorial**: DMPonline knowledge clip by Ugent

How to get started

- **Discuss RDM practices with your promotor**

- **Set up your ORCID if you do not have one (https://orcid.org/)**

- **Add to it to your account on our institutional repository (instructions)**

- **Check out our support pages on RDM**

- **Make use of our DMP authoring tool (https://dmponline.be)**

- **Check out our ADS course on DMP**

- **Reach out if you have any RDM related questions (rdm-support@uantwerpen.be)**

University of Antwerp

Unless stated otherwise, content in these slides is distributed under a

University of Antwerp