Managing Observational Data

Neha Moopen
Research Data Manager

AGENDA 🕑

+/- 30 minutes:

Managing Qualitative Data Through the Research Data Lifecycle

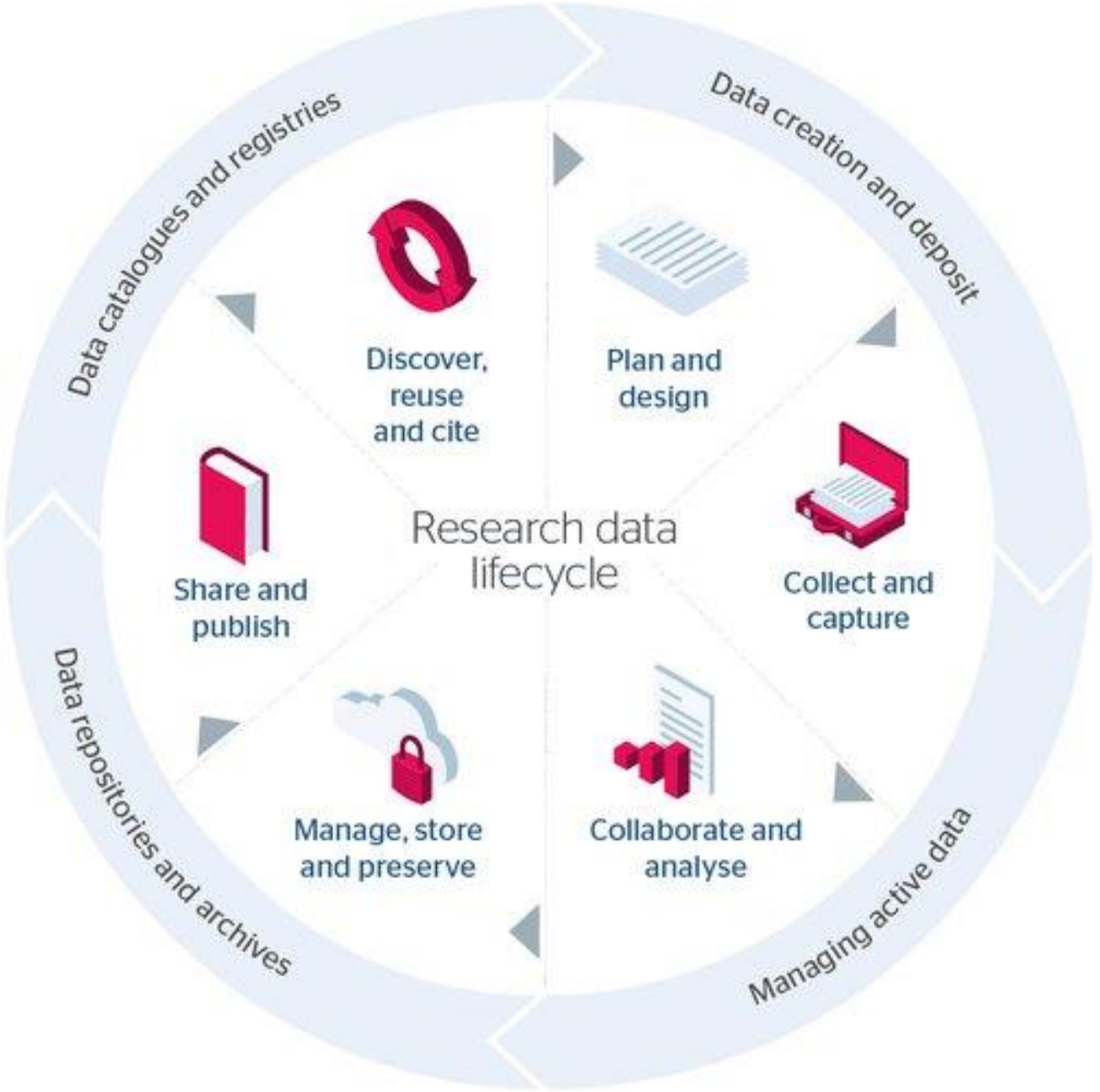
+/- 15 minutes:

Q&A

RESEARCH DATA LIFECYCLE

The *Research Data Lifecycle* describes the different stages research data go through before, during, and after a research project.

Each stage of the research data lifecycle entails various data management activities, and the choices made in one phase influence the next one.



First Things First…

FIRST THINGS FIRST…

You will need to:

- Obtain approval from your Faculty Ethics Review Board.
- Ensure you are complying with the GDPR.

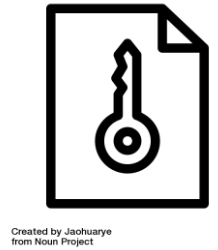
GDPR COMPLIANCE

DPIA

A Data Protection Impact Assessment systematically analyses, identifies and minimizes the data protection risks of a project or plan. It might be helpful (or required) to complete this.

INFORMED CONSENT

Informed Consent forms should be reviewed and updated to align with the GDPR. For example, information on the data subject's rights and the request to share data for further research should be more explicit.



PSEUDONYMIZATION

A pseudonymization protocol should be developed in advance, so it can be checked on whether it actually meets the criteria for pseudonymization according to the GDPR.

Plan & Design

PLAN & DESIGN (1/6)

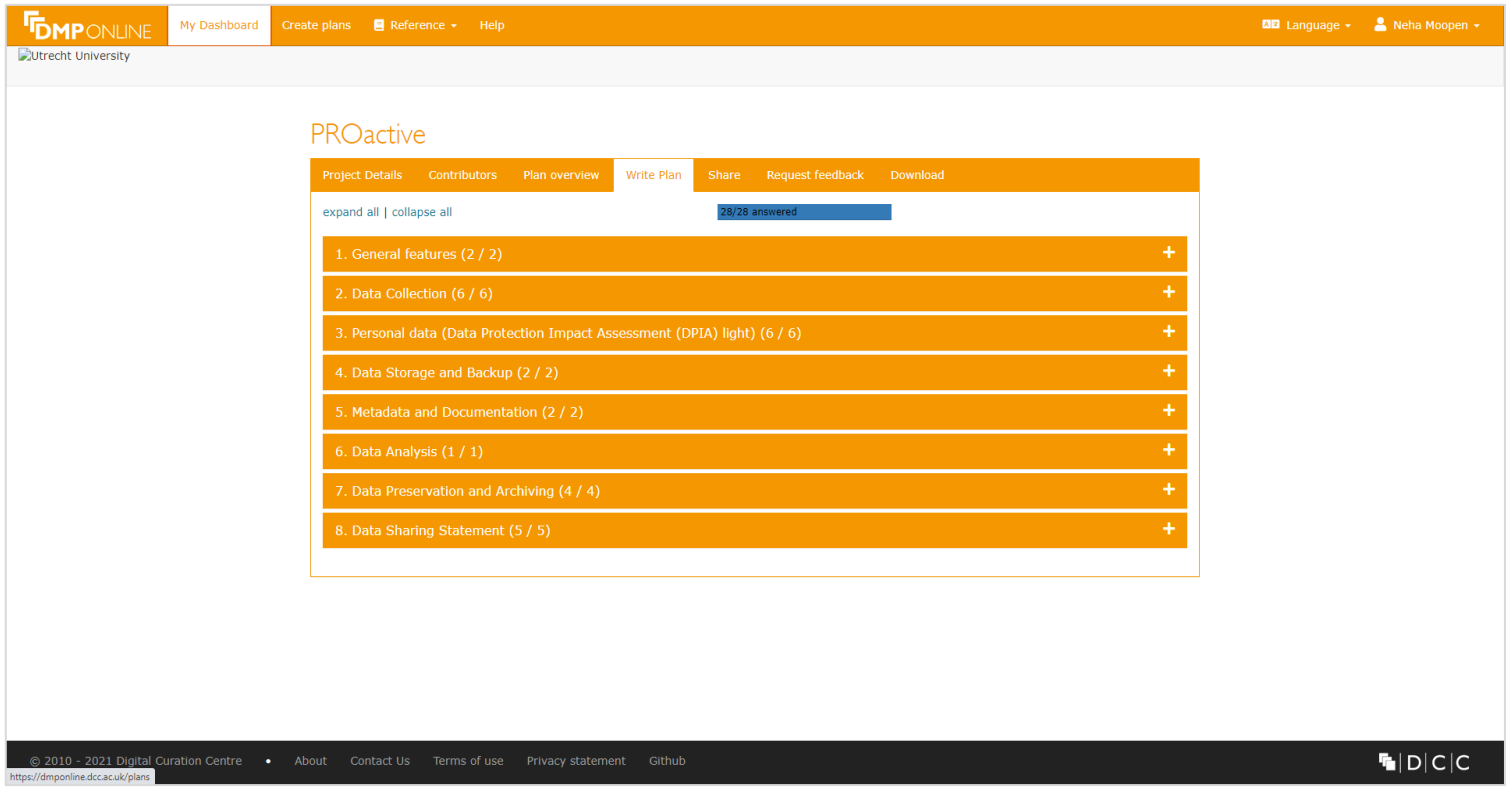
In this stage, you will need to:

- Write a Data Management Plan.
- Draw a Data Flow Diagram.

DATA MANAGEMENT PLAN

- A data management plan (DMP) is a digital + *living* document in which you describe your entire data lifecycle.

- It helps make RDM more concrete and actionable. Thus saving you time, work, and potentially money too.



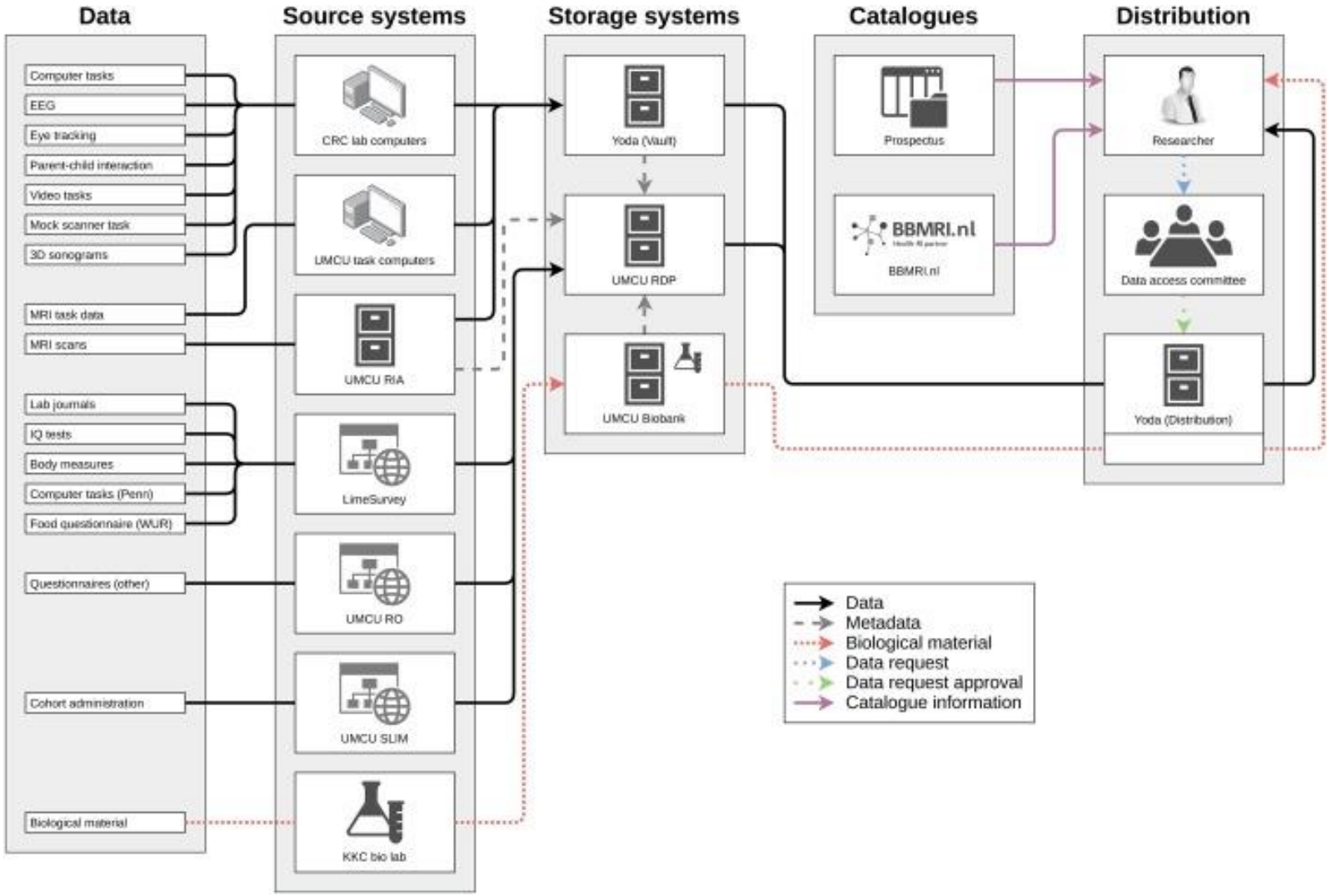
Sign into DMPonline with your institutional credentials and use the UMCU template.
Whenever you complete/update it, you can submit it to your division data manager for review.

see: https://utrechtuniversity.github.io/doy-data-handbook/data-management-plans.html

DATA FLOW DIAGRAM

A *data flow diagram* (*DFD*) maps out the flow of data through any process or system.

It will provide a bird's eye view of the whole process including steps, hardware, software, files, and people.

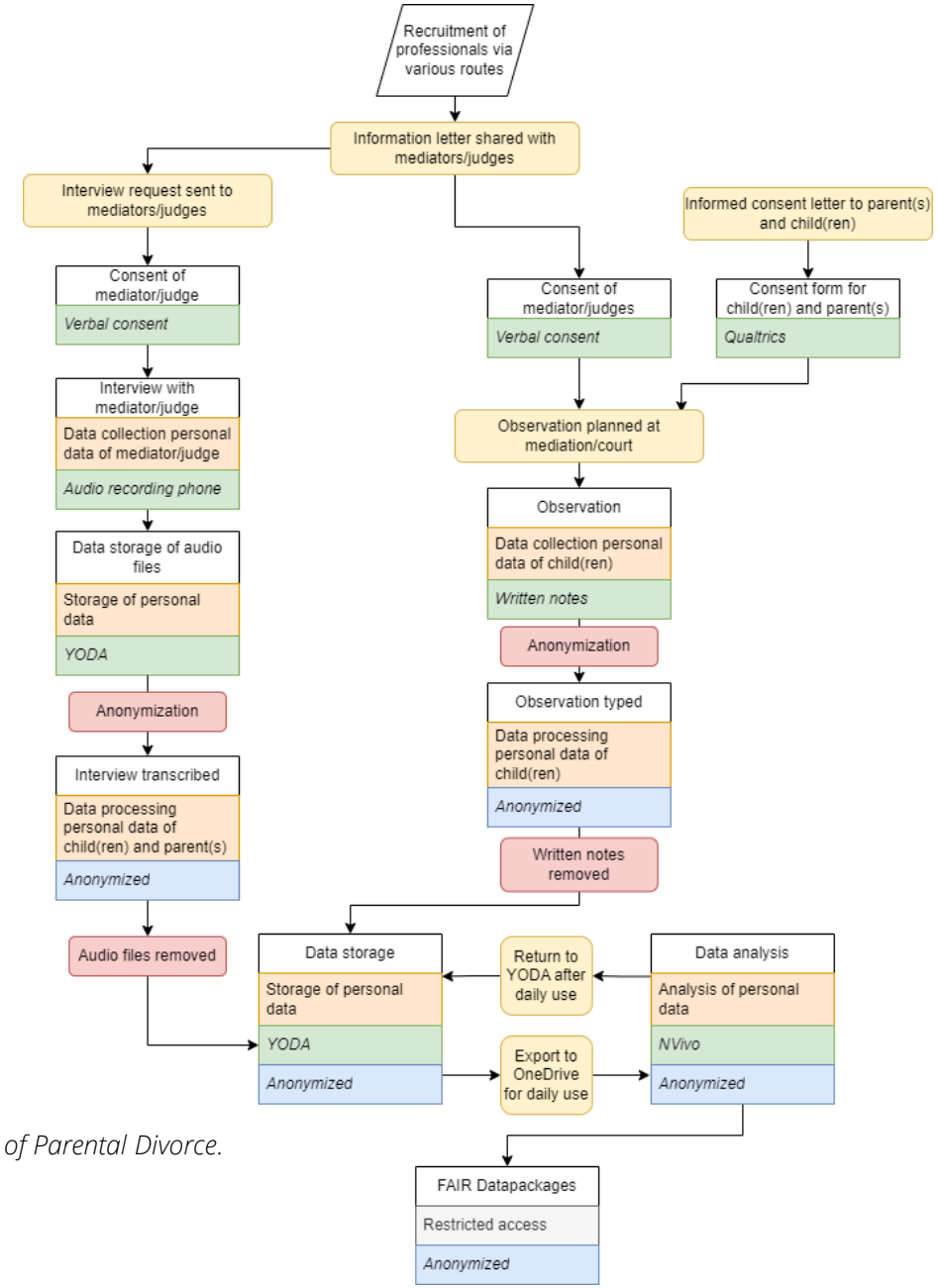


source: Zondergeld, J. J., Scholten, R. H., Vreede, B. M., Hessels, R. S., Pijl, A. G., Buizer-Voskamp, J. E., ... & Veldkamp, C. L. (2020). FAIR, safe and high-quality data: The data infrastructure and accessibility of the YOUth cohort study. *Developmental cognitive neuroscience*, *45*, 100834. https://doi.org/10.1016/j.dcn.2020.100834

DATA FLOW DIAGRAM

A *data flow diagram* (*DFD*) maps out the flow of data through any process or system.

It will provide a bird's eye view of the whole process including steps, hardware, software, files, and people.



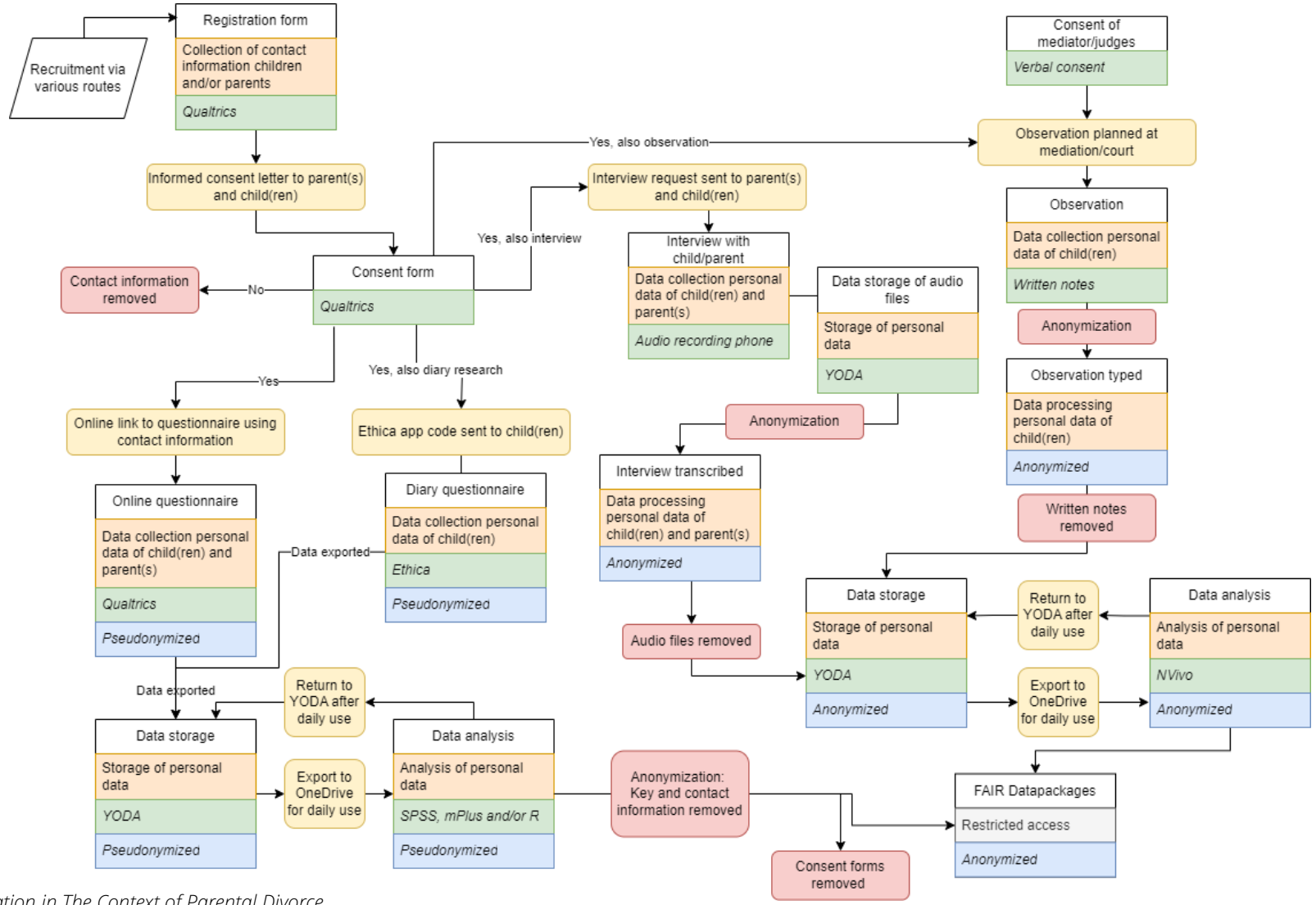
source: DPIA of *Hear, Hear! Nurturing Children's Self-Determination Through Participation in The Context of Parental Divorce.*

dr. Inge van der Valk (Ass. Prof., FSBS)
dr. Rianne van Dijk (Ass. Prof., FSBS)
dr. Charlotte Mol (Ass. Prof., LEG)
dr. Zoë Rejaän (Post-doc, FSBS)

DATA FLOW DIAGRAM

A *data flow diagram* (*DFD*) maps out the flow of data through any process or system.

It will provide a bird's eye view of the whole process including steps, hardware, software, files, and people.



source: DPIA of *Hear, Hear! Nurturing Children's Self-Determination Through Participation in The Context of Parental Divorce.*

dr. Inge van der Valk (Ass. Prof., FSBS)
dr. Rianne van Dijk (Ass. Prof., FSBS)
dr. Charlotte Mol (Ass. Prof., LEG)
dr. Zoë Rejaän (Post-doc, FSBS)

Collect & Capture

COLLECT & CAPTURE (2/6)

In this stage, you will need to:

- Write + Follow a Data Collection Protocol.
- Use tools & equipment safely + efficiently.
- Implement a naming convention.
- Utilize a clear folder structure.

DATA COLLECTION PROTOCOL

This should be a ‘single source of truth’ for those involved in data collection and for those who need to understand how the data were generated.

In other words, it is a manual for data collection and needs to be accessible to everyone.

TOOLS & EQUIPMENT

- Use institutionally-managed tools & devices as much as possible.

  This includes laptops/computers, audiovisual recording equipment.

- Transfer data from tools & devices as quickly as possible after data collection to a secure storage location. Consider automating this with scripts if you can.

- Store your data in two locations: one will be active storage, the other will be a backup.

NAMING CONVENTIONS

A naming convention is a set of rules for naming things. You can apply it to things like folders, files, and variables.

Names that are informative and useful for machines and humans are a step toward efficient data management and reproducible research.

The more consistent and meaningful the name, the easier it will be to locate and identify things, understand what they contain, and (re)use them.

Instead of developing a naming convention from scratch, you can start with one that is already being used in programming and software development communities:

Naming Covention	Example	Description
original name	`an awesome name`	N/A
snake_case	`an_awesome_name`	All words are lowercase and separated by an underscore (`_`)
kebab-case	`an-awesome-name`	All words are lowercase and separated by a hyphen (`-`)
PascalCase	`AnAwesomeName`	All words are capitalized. Spaces are not used.
camelCase	`anAwesomeName`	The first word is lowercase, the remaining words are capitalized. Spaces are not used.

see: https://utrechtuniversity.github.io/doy-data-handbook/naming-conventions.html

NAMING CONVENTIONS

Human-Readable Names

You can tailor naming conventions like `snake_case` and `PascalCase` to suit your project and workflow. Determine what information is relevant (or not) to create meaningful names and how you can string this information together. Don't forget to document this in your DMP!

Elements for Human-Readable Names

Names should be =<25 characters long and can include:

- Date of creation/update (`YYYY-MM-DD` or `YYYYMMDD`)
- Description of content, like type of data
- Initials of creator/reviewer
- Project number or acronym
- Location/coordinates
- Version number (like `v2` or v2.2`)

Machine-Readable Names

When names are machine-readable, they can be efficiently processed by computers and software. This makes it easier to search for files and run operations that involve programming like extracting information from file names or working with regular expressions.

Avoid

- Spaces
- Special characters like `$`, `@`, `%`, `#`, `&`, `*`, `!`, `/`, `\`
- Punction characters like `,`, `:`, `;`, `?`, `'`, `"`
- Accented characters

see: https://utrechtuniversity.github.io/doy-data-handbook/naming-conventions.html

FOLDER STRUCTURE

Good project organization & documentation improves efficiency + reusability of project materials for yourself and others.

OPTION 1: One main study

18-321_MYPRO
- A_ShortDescription
- B_Documentation
- C_PersonalData
- D_DataPreparation
- E_ResearchData
- F_DataAnalysis
- G_Output

A_ShortDescription

B_Documentation
- _OLD
- 1_METC
- 2_Monitoring
- 3_Finance
- 4_Questionnaires
- 5_Meetings
- 6_Contracts
- 7_QualityCheck

C_PersonalData
- 1_SourceData
- 2_CodeLists
- 3_NAW
- 4_ParticipantCorrespondence
- 5_RandomizationKey
- 6_DatamanagementTools

D_DataPreparation
- 1_SourceData
- 2_DataRelations
- 3_Scripts_and_Keys
- 4_DataDictionaries

E_ResearchData
- 1_MetaData
- 2_ResearchData

F_DataAnalysis
- 1_Publication
  - 1a_InfoUsedData
  - 1b_DataManagementScripts
  - 1c_Decisionlog
  - 1d_UsedData
  - 1e_DataDictionary
  - 1f_DataAnalysisScripts

G_Output
- 1_Publication
- 2_Data
- 3_OtherOutput

source: Guidance for UMCU Research Folder Structure

Collaborate & Analyze

COLLABORATE & ANALYZE (3/6)

In this stage, you will need to:

- Write a data processing & analysis protocol. Including quality checks.
- Use tools effectively, for example: *transcription services / optical character recognition / scanning / deidentification tools*.
- Ensure all data processing are done on institutional devices, for students – no local laptops
- Have rules about versioning within the team (related to naming conventions).
- If using CAQDAS tools like Nvivo, Atlas.ti – preserve all the scripts or project files or history files. The same goes for SPSS syntax & R scripts etc.
- For other types of processing/analysis, such as Excel sheets, preserve all the templates & material.

Manage, Store, Preserve

MANAGE, STORE, PRESERVE (4/6)

In this stage, you will need to:

- Document your data thoroughly.
- Prepare a Data Package.
- Provide accessible file formats.

DOCUMENTATION

Documentation refers to **contextual information** pertaining to your research data.

Documentation is meant to be **human-readable** and it is a crucial aspect of interoperability and reusability. It accompanies (structured) metadata and guides users to understand and interpret your data and reuse it effectively.

Examples include: *Grant / Study Proposals, Study, Protocol / Methodology, Data Management Plan (DMP), Codebooks, README files, Lab Notebooks, Legal / Policy / Administrative Documents.*

Here is a starter checklist (v0) that you can use to make an inventory of your documentation:
https://tinyurl.com/documentation-checklist

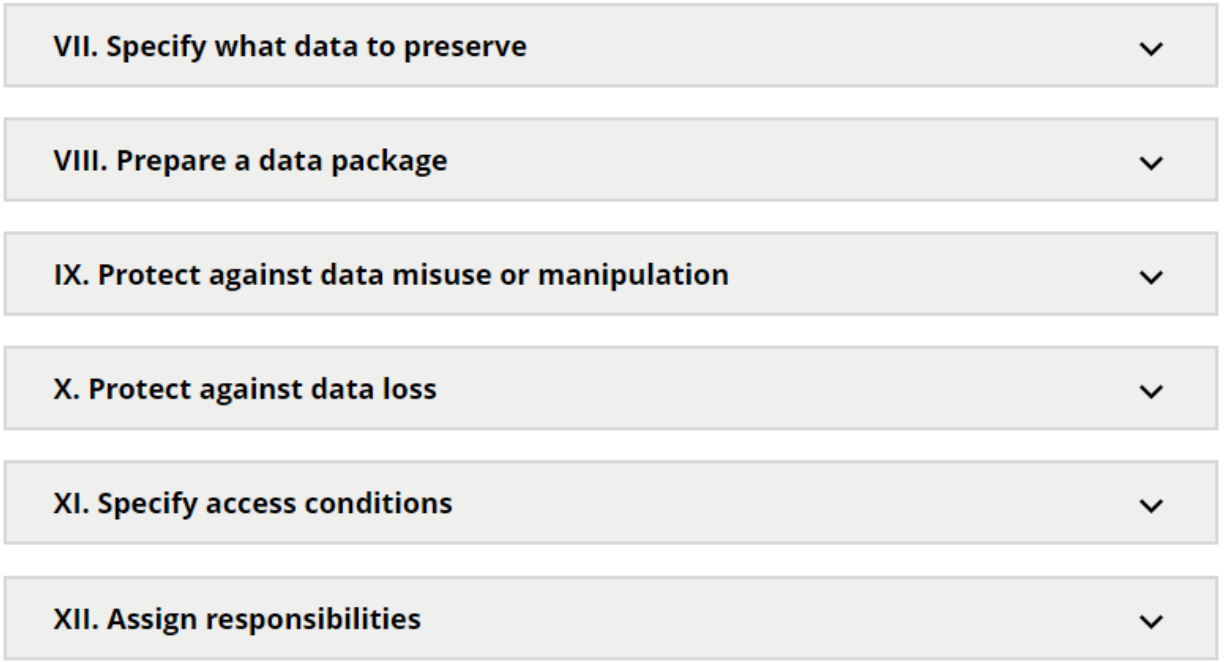
DATA PACKAGING

Creating a Data Package is primarily for long-term preservation of research data. For example, you may be required to store data for 10 years according to policy.

This is typically done for verification purposes / to check & maintain the integrity of the original research.

When you have prepared a Data Package, it will be easier to publish & share data.

For more information, see the RDM Support guide on Storing & Preserving Data.



FILE FORMATS

Ensure file formats are accessible for everyone in the long-term.

Create a copy of data and documentation in open-source formats where possible.

For example, *csv* instead of *xlsx* & REFI-QDA instead of regular Nvivo & Atlas.ti files

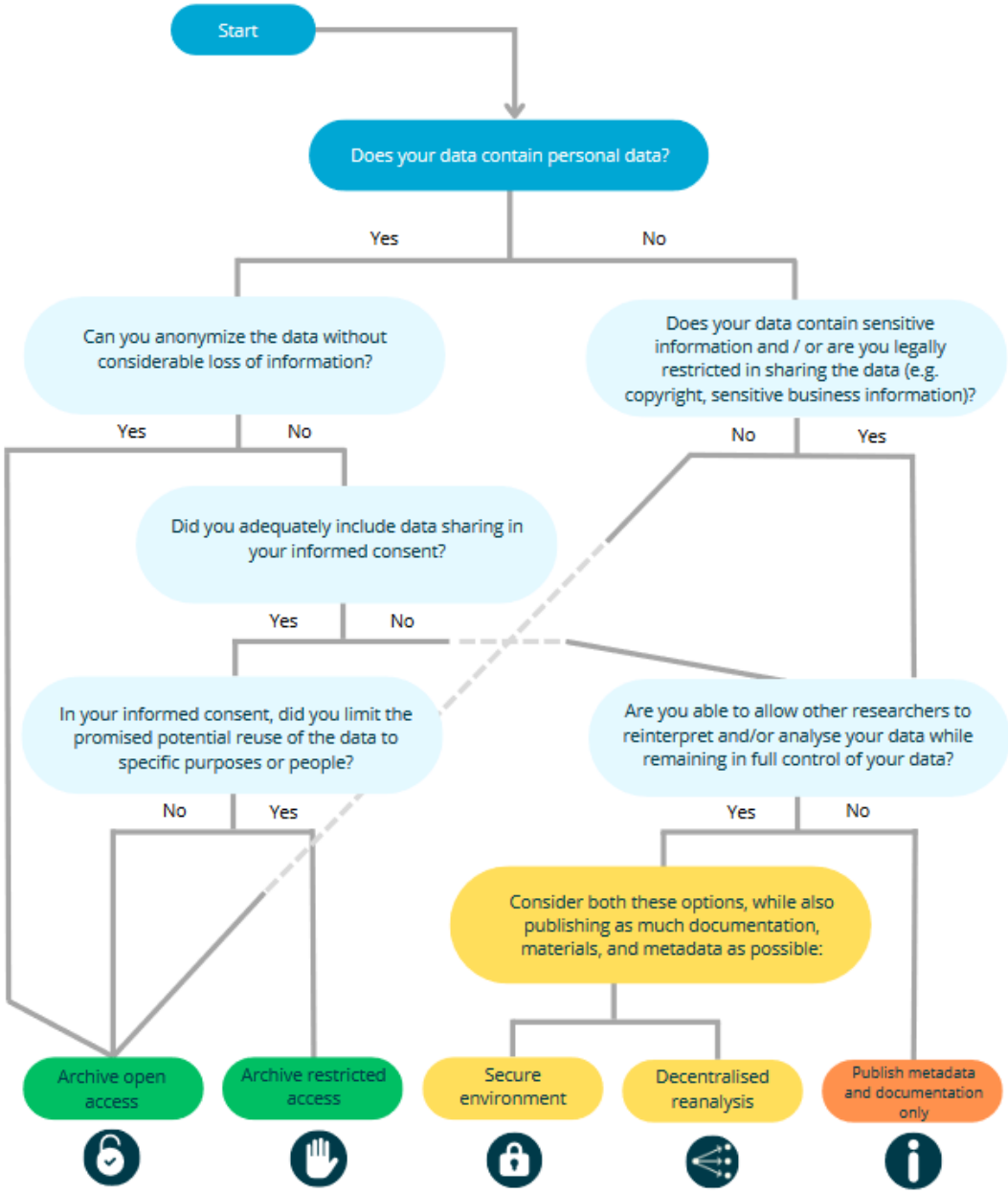
Share & Publish

SHARE & PUBLISH (5/6)

- To share or not to share…
- Possibilities for data sharing are determined by what participants consented to.
- Use pseudonymization & de-identification tools where possible. But remember this may not count as true anonymization.
- For publishing, follow the rule of: *as open as possible, as closed as necessary.*

Qualitative data reuse decision tree

as open as possible, as closed as necessary



source: Verburg, M., Braukmann, R., & Mahabier, W. (2023). Decision Tree - Making Qualitative Data Reusable (Version 2). Zenodo. https://doi.org/10.5281/zenodo.8160890

Discover, Reuse, Cite

DISCOVER, REUSE, CITE (6/6)

In this stage, you will need to:

- Write a Data Access Protocol.

DATA ACCESS PROTOCOL

When you're ready to start sharing your data, you can set up a detailed Data Access Protocol (DAP) that outlines the procedure for yourself and potential end users. This DAP will ideally be public and findable in your chosen repository.

There are many topics within the DAP, it will require the project team to come together and decide on what is best. For example, the terms & conditions for data reuse and the governance procedure.

PROactive

Data access and publication guidelines

This Data Access Protocol is written in consultation with the **PROactive team**

DISCLAIMER: this is a living document that will be amended and updated regularly. Consequently, no rights can be derived from the text of this publication.

Version 1, 1 February 2021

Table of contents

***source*:** Data Access Protocol of the *PROactive Cohort Study*.

PROactive Cohort Study: Data
Nijhof, Sanne; Putte, Elise van de; Hoefnagels, Johanna Wilhelmina, 2021, "PROactive Cohort Study: Data", https://doi.org/10.34894/FXUGHW, DataverseNL, V8

Resources

BONUS RESOURCES

- Making Qualitative Data Reusable - A Short Guidebook For Researchers And Data Stewards Working With Qualitative Data: https://doi.org/10.5281/zenodo.8160880
- Decision Tree - Making Qualitative Data Reusable: https://zenodo.org/doi/10.5281/zenodo.7777548
- Qualitative Data Repository: https://qdr.syr.edu/
- QualaLab: https://qualalab.org/
- UK Data Service & CESSDA

Q&A

Thanks!

Neha Moopen
n.moopen@uu.nl

Utrecht University

Sharing science,
shaping tomorrow