

# Research Data Management

CAS intervisiegroep dataverzameling

**Neha Moopen**

RDM Support

Research Data Manager

23-12-2021

# Agenda

## 09:15 - 09:45

Introduction & DMPs

Metadata & Documentation

EXERCISE

## 10:00 - 10:45

Data Storage

EXERCISE

Data Archiving

(Meta)data Publication

EXERCISE

## 11:00 - 11:45

Data Access Protocol

Data Sharing

EXERCISE

## 12:00 - 12:30

buffer time :P

Q&A / discussion

continue exercises

# Introduction

# Definitions

**RESEARCH DATA MANAGEMENT**

**FAIR DATA**

**OPEN DATA**

# Research Data Management

Research Data Management (RDM) refers to the **active organization** and **maintenance** of data created during a research project.

It is an **ongoing activity throughout the data lifecycle**, from initial planning to suitable archiving of the data at the project's completion.

# FAIR Data

The FAIR Data Principles are **a set of guiding principles** to improve scientific data management and stewardship.

**FINDABILITY** makes it possible for others to discover your data (metadata, Persistent Identifiers, etc.).

**ACCESSIBILITY** makes it possible for humans and machines to gain access to your data, under specific conditions or restrictions where appropriate.

**INTEROPERABILITY** ensures data and metadata conform to recognized formats and standards which allows them to be combined and exchanged.

**REUSABILITY** requires lots of documentation, which is needed to support data and interpretation and reuse.

# Open Data

Open Data is data that can be **freely used, re-used, and redistributed by anyone** - subject only, at most, to the requirement to attribute and share-alike ([Open Data Handbook](#)).



**Your data does not have to be 'open' to be FAIR!**

Make your data... **'as open as possible, as closed as necessary'** ([European Commission](#)).

# Definitions

*in context*

## RESEARCH DATA MANAGEMENT

*RDM = an activity / practice*

## FAIR DATA

*FAIR = principles that guide RDM activities / practices*

## OPEN DATA

*data does not have to be 'open' to be FAIR!*



# Data Management Plans (DMPs)

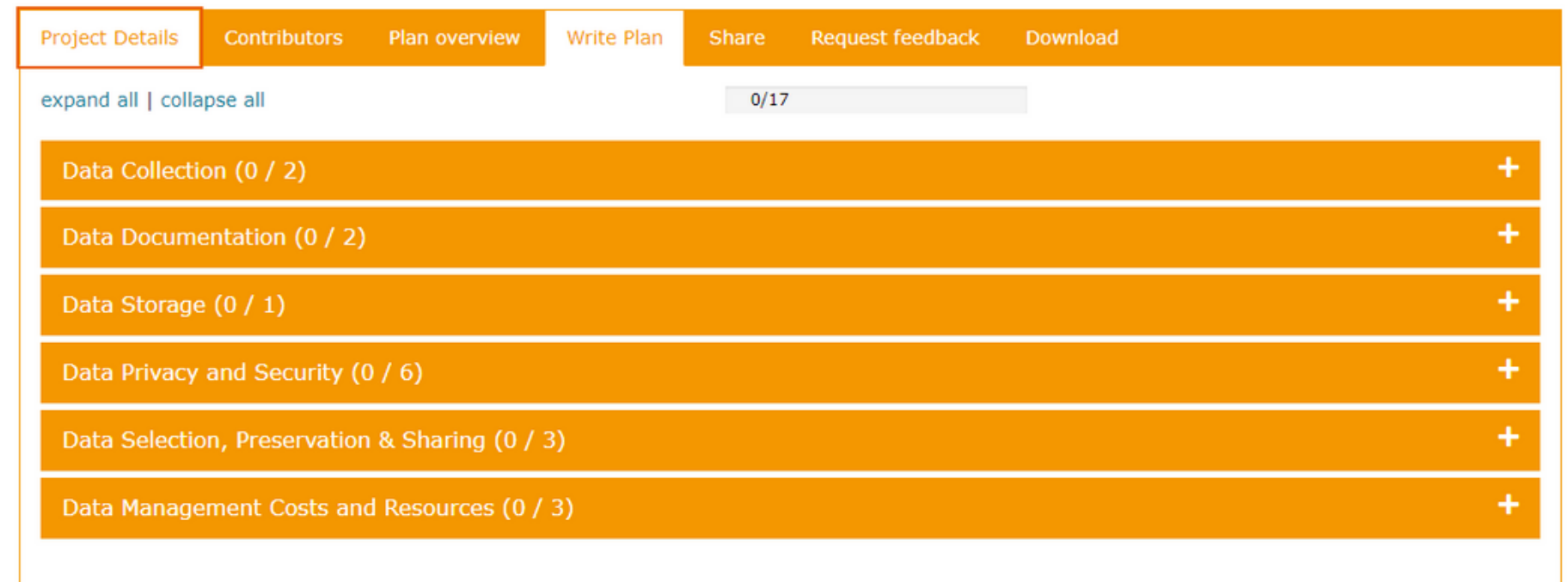
# Data Management Plans

A Data Management Plan (DMP) is a formal document that:

- describes your data, and
- outlines all aspects of managing your data, both during and after your project.

It is also a *living* document, it can (and should) be continually edited and updated.

A DMP helps make your RDM activities more concrete and actionable. It will save you time, work, and potentially money too.

A screenshot of the UU DMP template interface. The top navigation bar is orange and contains the following tabs: 'Project Details' (highlighted with a red border), 'Contributors', 'Plan overview', 'Write Plan', 'Share', 'Request feedback', and 'Download'. Below the navigation bar, there is a section with a blue header 'expand all | collapse all' and a progress indicator '0/17'. The main content area consists of several orange bars, each representing a section of the DMP. Each bar has a title, a progress indicator in parentheses, and a plus sign on the right. The sections are: 'Data Collection (0 / 2)', 'Data Documentation (0 / 2)', 'Data Storage (0 / 1)', 'Data Privacy and Security (0 / 6)', 'Data Selection, Preservation & Sharing (0 / 3)', and 'Data Management Costs and Resources (0 / 3)'.

👉 A screenshot of the UU DMP template on [DMPonline](https://dmponline.uu.nl/).

*Did you know?* When you complete a draft of your DMP, you can submit it to RDM Support for review and feedback!

# Metadata & Documentation

# Metadata

Metadata is **structured information** that describes one or more aspects of your research data.

In other words, metadata = 'data about data' 😁

Metadata is **machine-readable** and helps make your data findable and citable.

Metadata exists at different levels:

## PROJECT-LEVEL METADATA

This type of metadata describes higher-order aspects of your dataset: the “who, what, where, when, how and why” ...

It provides context for understanding why the data were collected and how they were used.

## DATA-LEVEL METADATA

This type of metadata is more granular and describes the data (variables) and dataset in detail.

# Metadata

## PROJECT-LEVEL METADATA

- Name of the project
- Dataset title
- Project description
- Dataset abstract
- Principal investigator and collaborators
- Contact information
- Dataset handle (DOI or URL)
- Dataset citation
- Data publication date
- Geographic description
- Time period of data collection
- Subject/keywords
- Project sponsor
- Dataset usage rights

## DATA-LEVEL METADATA

- Data origin: experimental, observational, raw or derived, physical collections, models, images, etc.
- Data type: integer, Boolean, character, floating point, etc.
- Instrument(s) used
- Data acquisition details: sensor deployment methods, experimental design, sensor calibration methods, etc.
- File type: CSV, mat, xlsx, tiff, HDF, NetCDF, etc.
- Data processing methods, software used
- Data processing scripts or codes
- Dataset parameter list, including
  - Variable names
  - Description of each variable
  - Units

# 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

- Grant / Study Proposals
- Study Protocol / Methodology
- Data Management Plan (DMP)
- Codebooks
- README files
- Lab Notebooks
- Legal / Policy / Administrative Documents

## DOCUMENTATION CHECKLIST!

Here is a starter checklist (v0) that you can use to make an inventory of your documentation:

<https://tinyurl.com/documentation-checklist>

**EXERCISES!**

**EXERCISES!**

# Exercises

## DATA MANAGEMENT PLAN

- Sign into [DMPonline](#) with your institutional credentials and create a DMP with the UU template.
- Complete the first page on *Project Details*.

If you already have a DMP, move to the next exercise for now. You can (and should) later review your DMP to see if it's up to date.

## DOCUMENTATION CHECKLIST

- Download / make a copy of the documentation checklist available via this link: <https://tinyurl.com/documentation-checklist>
- Complete the checklist as far as possible.
- Based on your checklist, go back to your DMP and update the *Data Documentation* section as far as possible.



# Data Storage

# Data Storage

WHAT DO YOU CONSIDER WHEN CHOOSING A STORAGE SOLUTION?

**STORAGE SPACE?**

**INTERNAL COLLABORATION?**

**PRICE?**

**EXTERNAL COLLABORATION?**

**BACKUPS?**


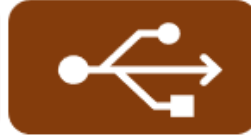




**SENSITIVE INFORMATION?**

**UU-MANAGED?**

**REMOTE ACCESS?**

# Data Storage

## OVERVIEW AND COMPARISON OF STORAGE SOLUTIONS

Storage Option					YODA		
Storage size	Varied	Varied	Varied	Varied	Varied	1TB	250GB
Price	NA	NA	Faculty	Faculty	TB €4/m	UU	UU
Back-up	✗	✗	✓	✓	✓	✓	✓
Controlled by UU	✗	✗	✓	✓	✓	✓	✓
Internal collaboration	✗	✗	✗	✓	✓	✓	✓
External Collaboration	✗	✗	✗	✗	✓	✓	✓
<b>Sensitive Information</b>	✗	✗	✓	✓	✓	✓	✓
Remote Access	✗	✗	✓	✓	✓	✓	✓

# Data Storage

## BEST PRACTICES IN STORING DATA

**I. Choose storage media wisely**



**II. Manage versions and copies of your data carefully**



**III. Structure names and folders**



**IV. Find and understand your data by assigning metadata**



**V. Use standard file formats**



**VI. Secure your data files**



For more information, see the RDM Support guide on [Storing & Preserving Data](#).

**EXERCISE!**

EXERCISE

# Exercises

## DATA STORAGE FINDER

- Go to the UU [Data Storage Finder](#) and see which storage tool might be most suitable for your project.
- Share the tool recommended for you in the chat!

## DATA MANAGEMENT PLAN

- Based on the recommendations of the Data Storage Finder, go back to your DMP and update the *Data Storage* section as far as possible.

# Data Archiving

# Data Archiving

## UNIVERSITY POLICY FRAMEWORK FOR RESEARCH DATA (UU)

Raw data from research must be kept available for a **minimum of ten years** for verification purposes...

...commencing from the date that the research results are published!

[Research policy NWO](#) › [Themes](#) › [Open Science](#) › [Research data management](#)

## Research data management

NWO expects researchers to:

- Preserve these data for **at least ten years**, unless legal provisions or discipline-specific guidelines dictate otherwise;



# Data Archiving

Data Archiving refers to the **long-term preservation of research data**.

Unlike (Meta)data Publication, Data Archiving is not *directly* related to the FAIR Data Principles.

It is typically done for verification purposes / to check & maintain the integrity of the original research, whereas the FAIR Data Principles are about sharing and reusing the data.

Nonetheless, similar RDM practices are involved in both activities!  
So archiving can always provide a basis for FAIRification...

VII. Specify what data to preserve



VIII. Prepare a data package



IX. Protect against data misuse or manipulation



X. Protect against data loss



XI. Specify access conditions



XII. Assign responsibilities



For more information, see the RDM Support guide on [Storing & Preserving Data](#).

**(Meta)Data Publication**

# (Meta)data Publication

## DATA ARCHIVING

You **archive** data for future reference.

- You create a data archiving package
- You make sure you can read and access the (raw) data later on.
- You ensure access to (specific) others for verification purposes.

**V**

**S**

## (META)DATA PUBLICATION

You **publish** data for findability and reusability.

- You create a data publication package
- You share data (and information about the data) that can be used by others for their own purposes.
- You specify the terms and conditions for access and reuse.

**REFLECT ON WHAT YOU NEED TO PRESERVE vs. WHAT YOU WANT TO PUBLISH**

# (Meta)data Publication

## Data Publication Package

Part of documentation is also determining the contents and organization of the Research Project Package:

- Raw data
- Processed data
- Documentation and Metadata
  - Variable codebook
  - Lab journals
  - Metadata standard (At the data and descriptive level; if any)
  - Protocols and Standard operating procedures (SOP)
- Scripts & Software
  - Analysis
  - Processing
- Legal Documents
  - Data protection impact assessment (DPIA)
  - Licenses
  - Informed Consent form (template only)
  - Data transfer agreement (DTA)
- Administration
  - Ethical review
  - DEC approval (animal research)
  - Grant & Consortium: applications and agreements
  - Data management plan (DMP)

### 👉 EXAMPLE

- If there are privacy-sensitive data involved, the processed data should be anonymized.
- Alternatively, the data files themselves can be placed under restricted access while the metadata and documentation are openly published.
- Remember, your data need not be 'open' to be FAIR!

# (Meta)data Publication

WHAT DO YOU CONSIDER WHEN CHOOSING A DATA REPOSITORY?



**PERSISTENT IDENTIFIER (DOI)?**

**SUSTAINABILITY?**

**QUALITY CONTROL?**

**SERVER LOCATION?**

**COSTS?**

# (Meta)data Publication

## HOW TO CHOOSE A DATA REPOSITORY?

### 1 DOMAIN-SPECIFIC REPOSITORY

Use a trusted repository already established for your research domain.

### 2 INSTITUTIONAL OR RECOMMENDED DATA REPOSITORY

If a trusted domain-specific repository is not available, use an institutional research data repository.  
The UU supports **Yoda** & **DataverseNL**.

### 3 GENERAL PURPOSE REPOSITORY

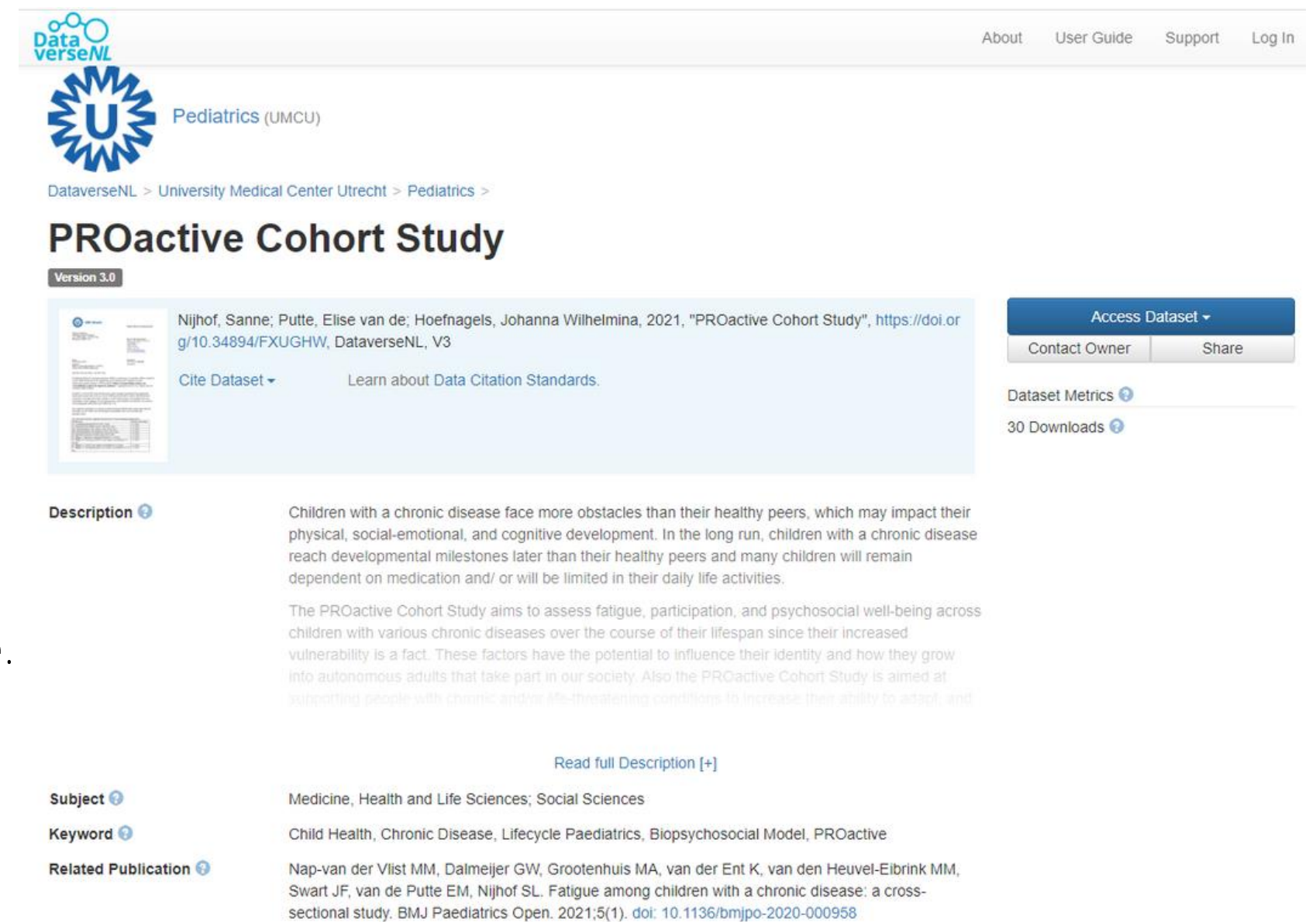
If neither of the above is available, use a general purpose repository like DANS EASY, Zenodo, 4TU.Center, Figshare.

# (Meta)data Publication

## AN EXAMPLE OF A PUBLISHED COHORT STUDY!

Nijhof, Sanne; Putte, Elise van de; Hoefnagels, Johanna Wilhelmina, 2021, "PROactive Cohort Study", <https://doi.org/10.34894/FXUGHW>, DataverseNL, V3

You'll see **the data itself has not been published**, but there is a **DOI** that makes the data **Findable and citeable**. There is also extensive **metadata and documentation** that outlines the **Accessibility** and ensures **Reusability**.



The screenshot shows the DataverseNL interface for the 'PROactive Cohort Study' dataset. The page header includes the DataverseNL logo and navigation links: About, User Guide, Support, and Log In. The dataset is associated with 'Pediatrics (UMCU)' and is part of the 'University Medical Center Utrecht' collection. The dataset title is 'PROactive Cohort Study' (Version 3.0). The description states: 'Children with a chronic disease face more obstacles than their healthy peers, which may impact their physical, social-emotional, and cognitive development. In the long run, children with a chronic disease reach developmental milestones later than their healthy peers and many children will remain dependent on medication and/ or will be limited in their daily life activities. The PROactive Cohort Study aims to assess fatigue, participation, and psychosocial well-being across children with various chronic diseases over the course of their lifespan since their increased vulnerability is a fact. These factors have the potential to influence their identity and how they grow into autonomous adults that take part in our society. Also the PROactive Cohort Study is aimed at supporting people with chronic and/or life-threatening conditions to increase their ability to adapt, and...'. The subject is 'Medicine, Health and Life Sciences; Social Sciences'. The keyword is 'Child Health, Chronic Disease, Lifecycle Paediatrics, Biopsychosocial Model, PROactive'. The related publication is 'Nap-van der Vliet MM, Dalmeijer GW, Grootenhuys MA, van der Ent K, van den Heuvel-Eibrink MM, Swart JF, van de Putte EM, Nijhof SL. Fatigue among children with a chronic disease: a cross-sectional study. BMJ Paediatrics Open. 2021;5(1). doi: 10.1136/bmjpo-2020-000958'. The dataset has 30 downloads.

**PROactive Cohort Study**  
Version 3.0

Nijhof, Sanne; Putte, Elise van de; Hoefnagels, Johanna Wilhelmina, 2021, "PROactive Cohort Study", <https://doi.org/10.34894/FXUGHW>, DataverseNL, V3

[Cite Dataset](#) [Learn about Data Citation Standards.](#)

**Description**

Children with a chronic disease face more obstacles than their healthy peers, which may impact their physical, social-emotional, and cognitive development. In the long run, children with a chronic disease reach developmental milestones later than their healthy peers and many children will remain dependent on medication and/ or will be limited in their daily life activities.

The PROactive Cohort Study aims to assess fatigue, participation, and psychosocial well-being across children with various chronic diseases over the course of their lifespan since their increased vulnerability is a fact. These factors have the potential to influence their identity and how they grow into autonomous adults that take part in our society. Also the PROactive Cohort Study is aimed at supporting people with chronic and/or life-threatening conditions to increase their ability to adapt, and...

[Read full Description \[+\]](#)

**Subject**

Medicine, Health and Life Sciences; Social Sciences

**Keyword**

Child Health, Chronic Disease, Lifecycle Paediatrics, Biopsychosocial Model, PROactive

**Related Publication**

Nap-van der Vliet MM, Dalmeijer GW, Grootenhuys MA, van der Ent K, van den Heuvel-Eibrink MM, Swart JF, van de Putte EM, Nijhof SL. Fatigue among children with a chronic disease: a cross-sectional study. BMJ Paediatrics Open. 2021;5(1). doi: 10.1136/bmjpo-2020-000958

[Access Dataset](#)  
[Contact Owner](#) [Share](#)

**Dataset Metrics**  
30 Downloads

**EXERCISES!**



# Exercises

## DATA REPOSITORY FINDER

- Go to the UU [Data Repository Finder](#) and see which data repository might be most suitable for publishing your project.
- Share the tool recommended for you in the chat!

## DATA SELECTION

- Reflect on what data & documentation you would like to publish for eventual citation and reuse.
- Share your thoughts in the chat!
- *Hint:* don't forget to check your project documentation checklist!

## DATA MANAGEMENT PLAN

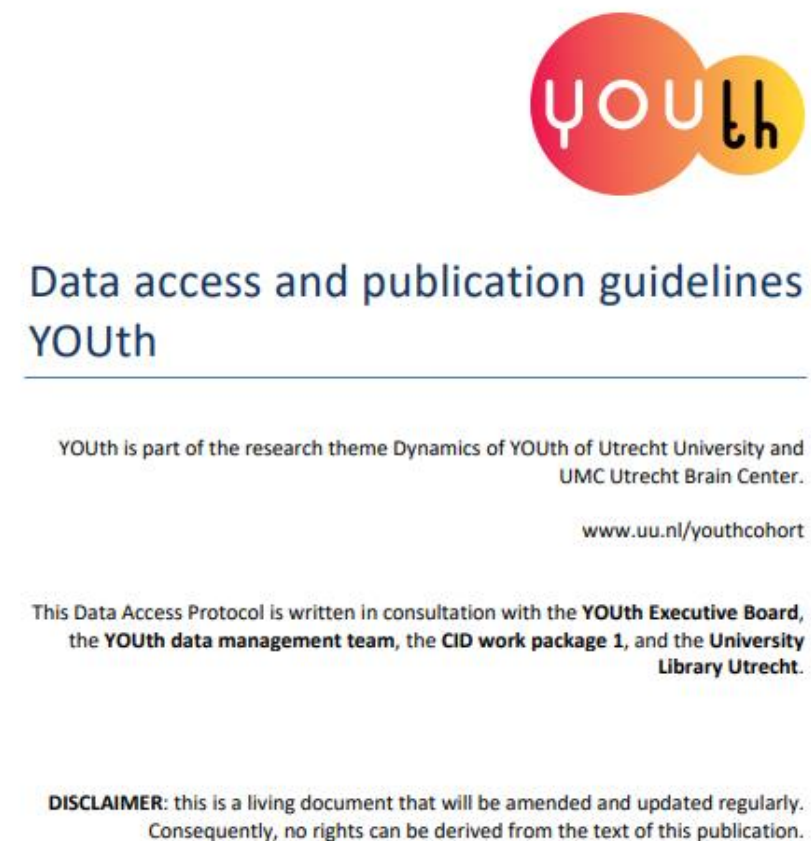
- Based on the above exercises, go back to your DMP and update the *Data Selection, Preservation & Sharing* question (5.2) as far as possible.

# 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 data sharing procedure for yourself, your research team, and potential re-users**. This DAP will ideally be public and findable in your chosen repository.

See the YOUTh Cohort Study's DAP here: [https://www.uu.nl/sites/default/files/dataaccessprotocol\\_youth\\_191029.pdf](https://www.uu.nl/sites/default/files/dataaccessprotocol_youth_191029.pdf)



## Table of Contents

1	General .....	3
2	Definitions .....	4
3	Data ownership .....	8
4	Conditions of data use .....	8
5	Contribution in return for access to YOUTH data .....	9
6	Data request procedure .....	9
7	Data access for verification purposes .....	11
8	Publication guidelines .....	11
9	Additional researcher responsibilities .....	14
10	Appendices .....	15
10.1	Data Request form YOUTH (version 4.0, October 28, 2019) .....	15
Annex 1: Form contributions to YOUTH .....		18
10.2	Allocation of PhD students to YOUTH .....	20
10.3	YOUTH Internship Protocol .....	23
Annex: Model Agreement Internships YOUTH .....		26
10.4	Pre-publication checklist YOUTH (version 1.0, September 3, 2019) .....	33

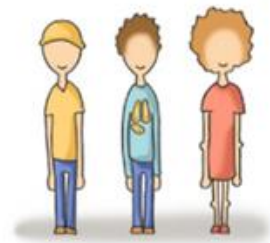
# Data Access Protocol

There are many topics within a DAP, it will require you (and/or the project team to come together) to decide on what is relevant and best for your data. This can include, for example, the terms & conditions for data reuse and the governance procedure.

See the PROactive Cohort Study's DAP here: <https://dataverse.nl/file.xhtml?fileId=141206&version=3.0>

## 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

1. General .....	3
2. Definitions .....	3
3. Data ownership .....	5
4. Conditions of data use .....	5
5. Contribution in return for access to PROactive data .....	6
6. Data request procedure .....	6
7. Collecting new data .....	9
8. Data access for verification purposes .....	10
9. Publication guidelines .....	11
10. Additional researcher responsibilities .....	12
11. PROactive members and contact details .....	13
12. Appendices .....	15
Appendix 1 .....	15
Appendix 2 .....	20

# Data Access Protocol

DAPs do not have to be 20-30 pages long like YOUTh's & PROactive's! 😊


You can make it as simple and elaborate as you like, it all depends on you and your project team.

## Reflect on:

- **What would you like to get out of sharing the data?** For example, citations/acknowledgments, co-authorship, collaboration? This should be specified in the DAP so the end-user knows their obligations.
- **What kind of time and effort can you and/or your team invest in the data governance?** For example, assessing incoming requests, preparing a datafile for sharing, maintaining a data sharing logbook.

**Note:** If there is privacy-sensitive data involved, even the simplest DAPs have to take some legal considerations into account!

# Data Access Protocol



AboutUser GuideSupportLog In

FilesMetadataTerms**Versions**

Terms of Use

<b>Waiver</b>	<p>Our <a href="#">Community Norms</a> as well as good scientific practices expect that proper credit is given via citation. Please use the data citation above, generated by the DataVerse.</p> <p>No waiver has been selected for this dataset.</p>
<b>Terms of Use</b>	<p>The PROactive team grants data access for dedicated research questions only. The research question should be precisely specified in the data request form. Please refer to the Data Access and Publication Guidelines for all information related to the data request protocol and reuse.</p>
<b>Confidentiality Declaration</b>	<p>Depending on who the recipient is and where they are located, a Data Use Agreement or Data Transfer Agreement in line with GDPR regulations will have to be signed. Consortium Agreements are also a possibility in the case of collaborations.</p>
<b>Special Permissions</b>	<p>The data request form has to be completed when requesting the data. Furthermore, the pre-publication checklist has to be followed prior to submission of a manuscript. Both of these can be found in the Data Access and Publication Guidelines. Note that PhD candidates or other students cannot request the data independently, but only in consultation with their supervisor(s).</p>
<b>Restrictions</b>	<p>It is important to note that the PROactive team does not consider overlap in research questions when evaluating data requests. The data cannot be used for purposes other than what it was specifically requested for and it cannot be further shared or distributed. The researchers will be responsible for ensuring the safety and security of the data. The restrictions that apply to the dataset will be outlined in the Data Use Agreement or Data Transfer Agreement.</p>
<b>Citation Requirements</b>	<p>We encourage researchers to include "PROactive Cohort Study" either as keyword or as corporate author on publications. We also ask researchers to include an acknowledgment on behalf of PROactive Cohort Study, this text can be found in the Data Access and Publication Guidelines. If possible, we also ask the PROactive Cohort Study logo in combination with the WKZ logo to be used in all communication materials.</p>
<b>Conditions</b>	<p>All projects requesting PROactive Cohort Study data should be appropriately funded and sustainable.</p>

Check under the *Terms* tab of the PROactive Cohort Study's publication:  
<https://doi.org/10.34894/FXUGHW>

**You'll see the terms and conditions are specified more briefly.**

This is an in-built form within DataVerseNL and you'll find it in all repositories! It can be a start, in case you don't want to get into a 20-30 page document.

Restricted Files + Terms of Access

<b>Restricted Files</b>	There are 0 restricted files in this dataset.
<b>Terms of Access</b>	The data will not be available for download directly via DataVerseNL. If a data request is approved, the data will be delivered in a safe and secure manner.
<b>Data Access Place</b>	Accessing the data may require utilizing the Digital Research Environment at the University Medical Center Utrecht or SURFfilesender for a secure data transfer.
<b>Availability Status</b>	Documentation and information regarding the dataset can be found on DataVerseNL, but the dataset itself is not publicly available for download, as it contains sensitive personal data. The data will only be delivered when a data request is approved and all the necessary agreements are signed.
<b>Study Completion</b>	This is an ongoing study, so there isn't a complete dataset. Twice a year we carry out a 'data lock' where all data collected in the previous 6 months are added to the larger dataset.

Guestbook

<b>Guestbook</b>	No guestbook is assigned to this dataset, you will not be prompted to provide any information on file download.
------------------	---



**EXERCISE**

# Exercises

## TERMS & CONDITIONS

- Imagine what a data sharing procedure could look like for your project.
- What are some terms and conditions you would like to specify for the end user?
- What investments (time & effort) would be required from you and your team for efficient data sharing
- If you like, share in the chat or discuss with one another!

## LINKS

- YOUTh Cohort Study's DAP:  
[https://www.uu.nl/sites/default/files/dataaccessprotocol\\_youth\\_191029.pdf](https://www.uu.nl/sites/default/files/dataaccessprotocol_youth_191029.pdf)
- PROactive Cohort Study's DAP here:  
<https://dataverse.nl/file.xhtml?fileId=141206&version=3.0>
- PROactive Cohort Study's publication:  
<https://doi.org/10.34894/FXUGHW> (for the Terms tab)



# Data Sharing

# Data Sharing

## TOOLS FOR DATA SHARING



**SURFFileSender** is a reliable tool to send data to another user.

**Note:** You should only transfer privacy-sensitive data once all the legal requirements are checked and met!



Instead of sending data to another user, you can use **Yoda** to add an (external) user to your research folder. **This does not always work in practice**, since the end-user could download the data to their computer and misuse it.

The IT teams are working on a **Virtual Research Environment** linked to Yoda, keep your eyes open for when it's available!

# Data Sharing

## IMPORTANT CONSIDERATIONS + DOCUMENTS / PROCEDURES (IN BRIEF)

When privacy-sensitive data is involved:

- The **ICF** should have clearly informed participants about data sharing and reuse + they should agree to it
- A **DPIA** may have to be carried out, this will reveal to what extent it is safe to share data (or not) and how that can be put into practice (for example, pseudonymization techniques)
- Any transfer of data outside the UU will require a **Data Transfer Agreement** in line with the GDPR, the complexity of the DTA will vary depending on the nature of the transfer (for example, transfer outside the EU).

**INFORMED CONSENT FORMS**

**DMP**

**DATA PROTECTION IMPACT  
ASSESSMENT (DPIA)**

**DATA TRANSFER AGREEMENT  
(DTA)**

**DATA PROCESSING  
AGREEMENT (DPA)**

Q&A!

# Questions so far...

- What are the benefits of open science?
- What will a person do if he or she finds that his or her research idea has been done during pre registration?
- What if a person finds out that someone has done the same data collection as him or her?
- Right now I have the excel sheet for codebook. I've heard about programs like DublinCore to create codebooks, would you recommend using a system like this?
- What repository/storage would you recommend for research concerning data from three waves including interviews, questionnaires and audio files? What are considerations to keep in mind when choosing a repository?
- I'm new to the topic of FAIRification. Would you recommend specific e-learnings or guides?

# Thanks!

**For questions or hands-on RDM support,  
you can reach me at: [n.moopen@uu.nl](mailto:n.moopen@uu.nl)**

I'd be happy to hear from you!

**Neha Moopen**

Research Data Manager