Utrecht University UMC Utrecht

29-10-2024

Research Data Management & FAIRification

Neha Moopen
Research Data Manager

Research Data Management & FAIRification

Neha Moopen
Research Data Manager

INTRODUCTION 👋

- Research Data Manager @ University Library
- We support researchers with everything RDM, in line with open science & FAIR principles.
- More information: RDM Support website

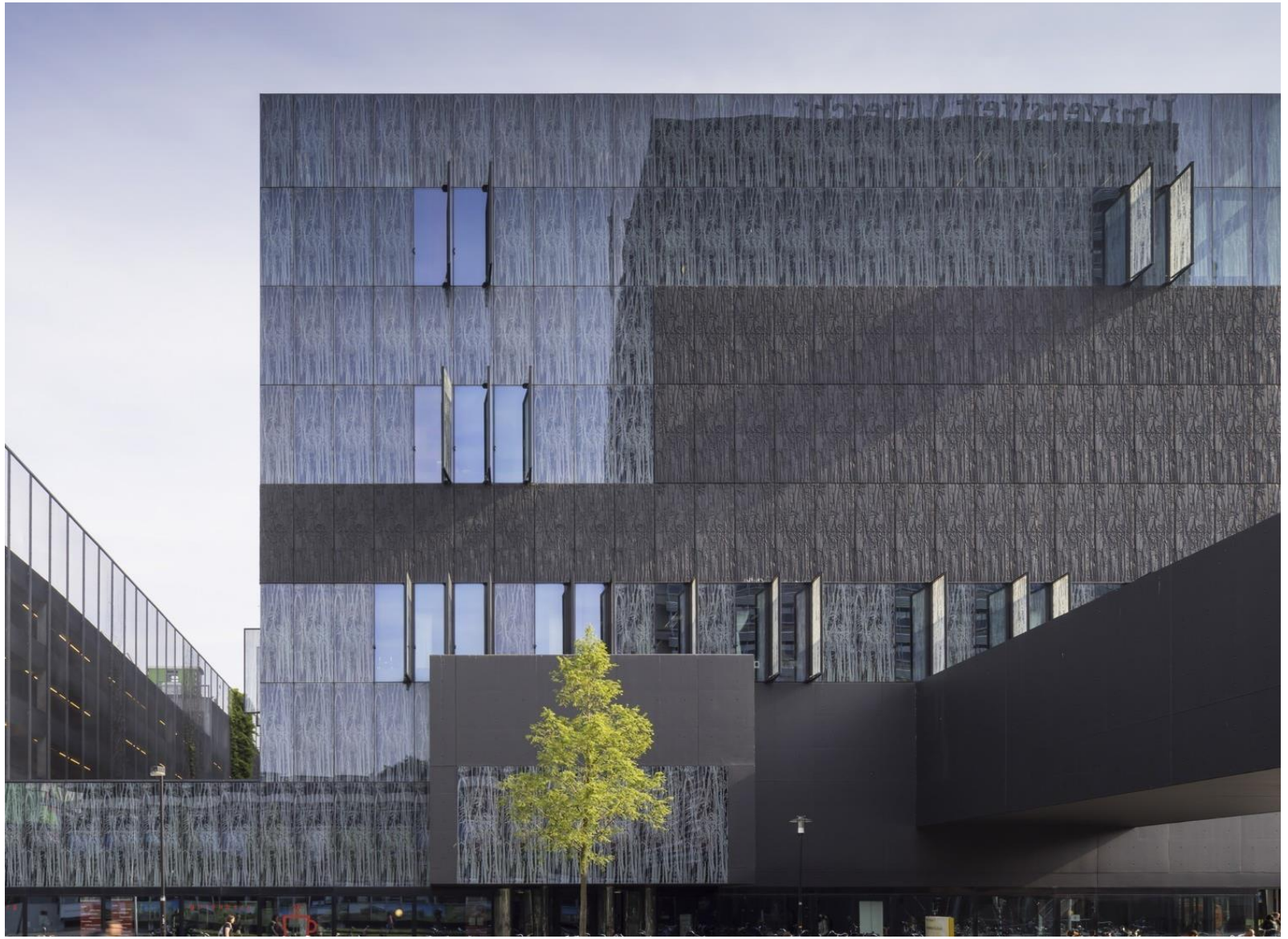
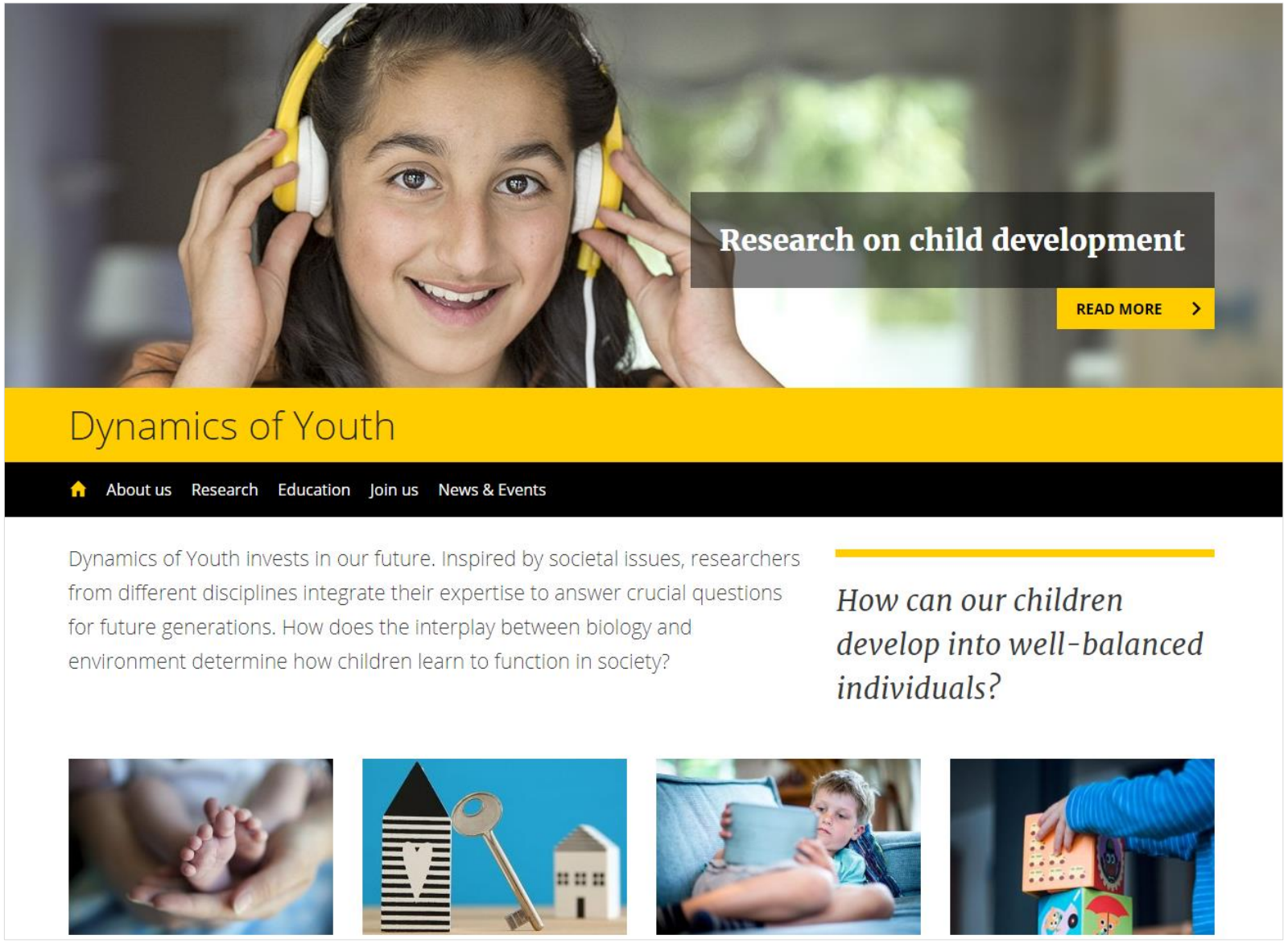


image: https://www.archilovers.com/projects/130396/utrecht-university-library.html

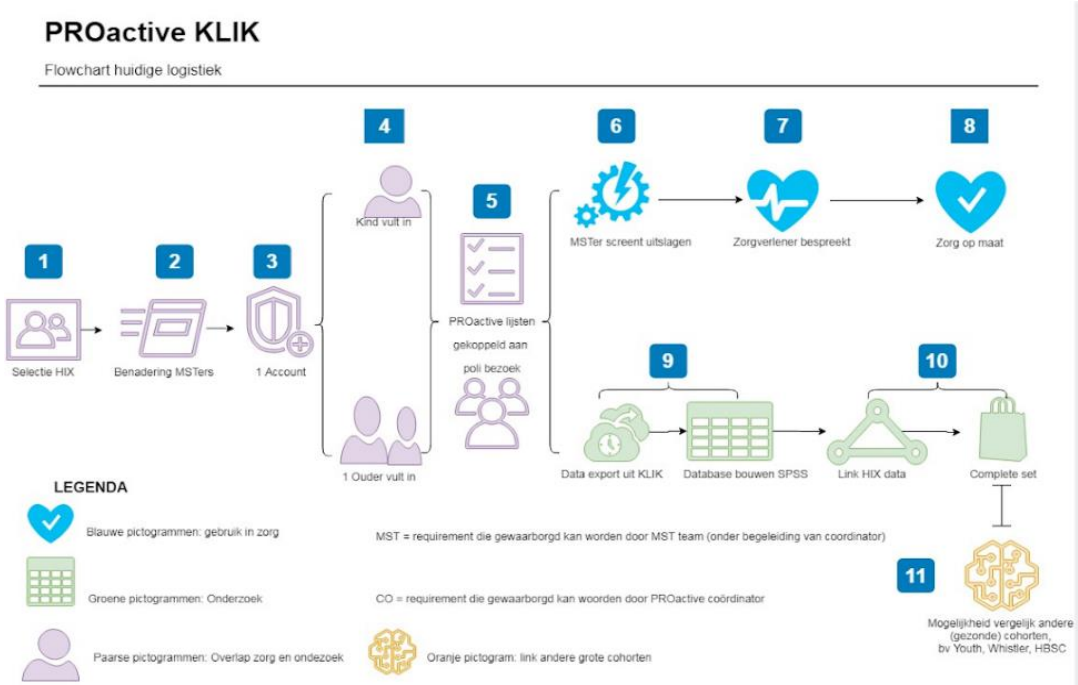
INTRODUCTION 👋



- Currently embedded in *Dynamics of Youth* to improve FAIRness of data within the strategic theme.
- We hope to develop some tools/templates/solutions to make FAIRification easier for youth researchers.

screenshot: https://www.uu.nl/en/research/dynamics-of-youth

Let's take a quick look at PROactive's RDM & FAIRification journey!



1 **Requirements:** Toegang alle HIX dossiers en eigen lijsten (MST&CO), poli's screenen (MST), verlengde armconstructie zorgverleners (MST&CO)

Verbeterpunten: duidelijk welke inclusiecriteria gelden voor inclusie zorg en onderzoek, afspraken over patiënten die in meerdere groepen vallen, afspraken overlap PMC/WKZ

2 **Requirements:** Mails versturen en bellen uit naam zorgverlener (MST), goed functionerend logistiek bestand (CO), up-to-date inwerkdocumenten (MST), bijhouden investigator site files (CO), begeleiding MSTers in uitvoeren logistiek (dagelijkse helpdesk en frequente overleggen (CO))

Verbeterpunten: Duidelijker onderscheid zorg/onderzoek, inbedding MSTers in zorgstructuur/financiering, checken met AVG requirements, nieuwe niet-WMO aanvraag met protocol en SOPs

3 **Requirements:** Ouders en kind hebben beiden een emailadres nodig, helpdeskfunctie voor patiënten (MST), contact met KLIK team (dubbele accounts, vragenlijsten die niet werken (CO))

Verbeterpunten: KLIK account en IC nu gebaseerd op zorgverlenende ouders ipv gezagsdragende ouders

4 **Requirements:** Nauwkeurige follow-up zodat volledig door beiden ingevuld wordt (MST), kwaliteitsbewaking op dataverzameling, follow-up en responspercentages (CO)

Verbeterpunten: Nauwkeurig monitoren follow up respons-percentages en tijdig bijsturen, evalueren vragenlijsten KLIK en komen tot een gereduceerd vragenlijst pakket

5 **Requirements vragenlijst systeem**: Een medium 1) waarin ouders en kinderen apart kunnen inloggen en niet elkaars antwoorden kunnen zien, 2) wat grafisch aantrekkelijk is, 3) gekoppeld aan Hix, 4) waarin data export gemakkelijk verloopt

Verbeterpunten: Eventueel ander vragenlijst systeem of databeheersysteem

6 **Requirements:** Consensus over wat te screen en terug te koppelen, tijd voor screenen en terugkoppelen (MST)

Verbeterpunten: Wat screenen/terugkoppelen, oa bij follow up los van poliklinische controle. Nadenken over ethische verplichting in kader van onderzoek/zorg, overlap met andere zorg/vragenlijsten

7 **Requirements:**Zorgverleners getraind in KLIK (CO), koppeling KLIK en HIX (MST), evaluatie met zorgverleners KLIK groepen (CO)

Verbeterpunten: Tweefactor authenticatie en dubbele accounts bij koppelen KLIK en HIX, Structuur indien niet gekoppeld aan polibezoek

8 **Requirements/verbeterpunten:** uitgewerkte zorgpaden met duidelijke rolverdelingen (arts, verpleegkundig specialist, PSA, fysiotherapie, eerste lijn), onderzoek naar effectieve interventies (bv PROfeel, spelinterventies), validering en normering uitgevraagde concepten met afkapwaardes en grafische terugkoppeling voor gebruik in zorg

9 **Requirements:** Afgebakende protocollen in KLIK, ontdubbelen Excel files (veel handmatig werk, CO), vragenlijsten samenvoegen (CO), datastructuur op G-schijf (CO)

Verbeterpunten: Efficiëntere manier van data extraheren uit KLIK/verwerken/databases bouwen, mogelijk automatische data-export (API/SFDP), huidige syntaxen evalueren en plan bedenken hoe follow-up data overzichtelijk kan worden verwerkt (nu heel veel handmatig werk)

10 **Requirements:** Extractie biologische data uit HIX (MST&CO), controleren toestemming (CO), syntaxen uitvoeren voor sumscores (CO)

Verbeterpunten: Datamanagement en automatische extractie data uit HIX (nu handmatig dubbel geëxtraheerd, veel handmatig werk. Uiteindelijk platform waarin data gemakkelijk beschikbaar kan worden gesteld voor verschillende onderzoeksvragen/evt commissie die zich daarover buigt en toezicht houdt

11 **Requirements:** Overzichtelijke datasets

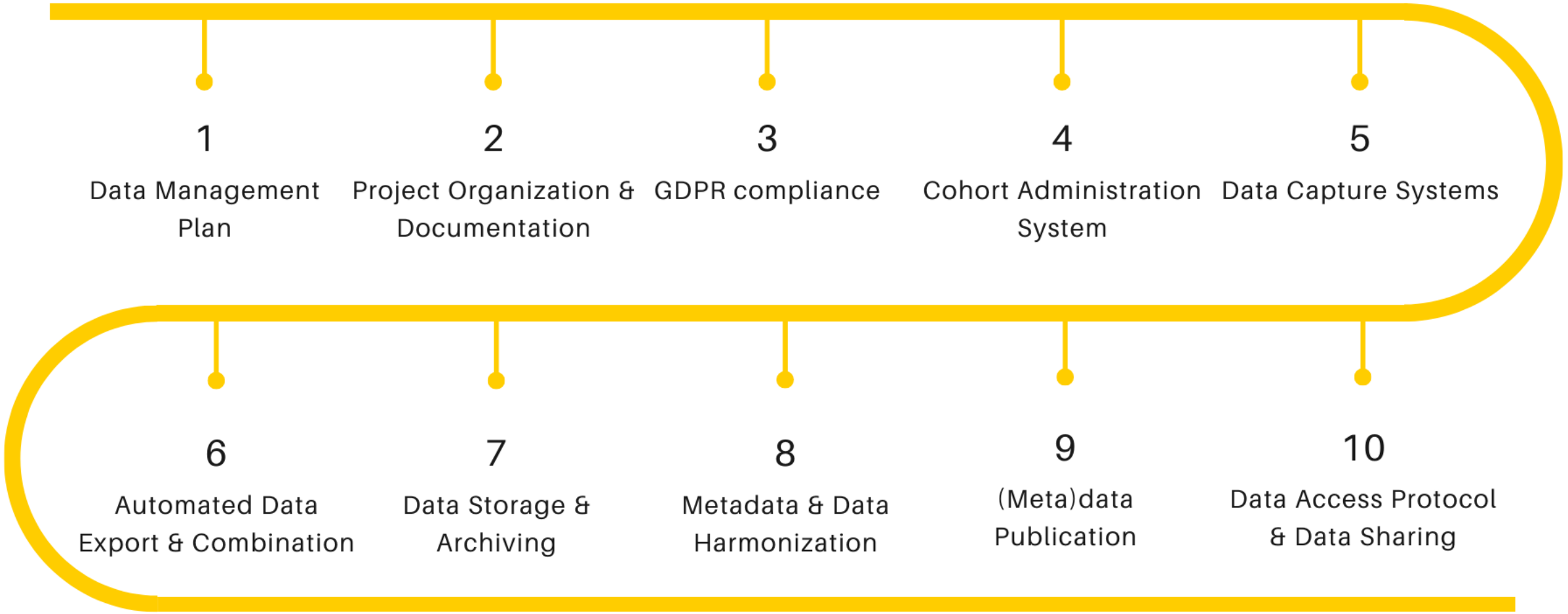
Verbeterpunten: Afstemming tussen cohorten, governance structuur data PROactive cohort

PROactive: NEEDS & REQUIREMENTS

- More efficiency in data management / data workflow.
- Automation wherever possible.
- Solutions for collaboration & data sharing: should be efficient, structured, and secure.
- Findability and visibility.
- Homogenization with other cohorts.
- Upgrade data management and storage systems.
- The patient remains central, that can't be compromised.

source: PROactive team

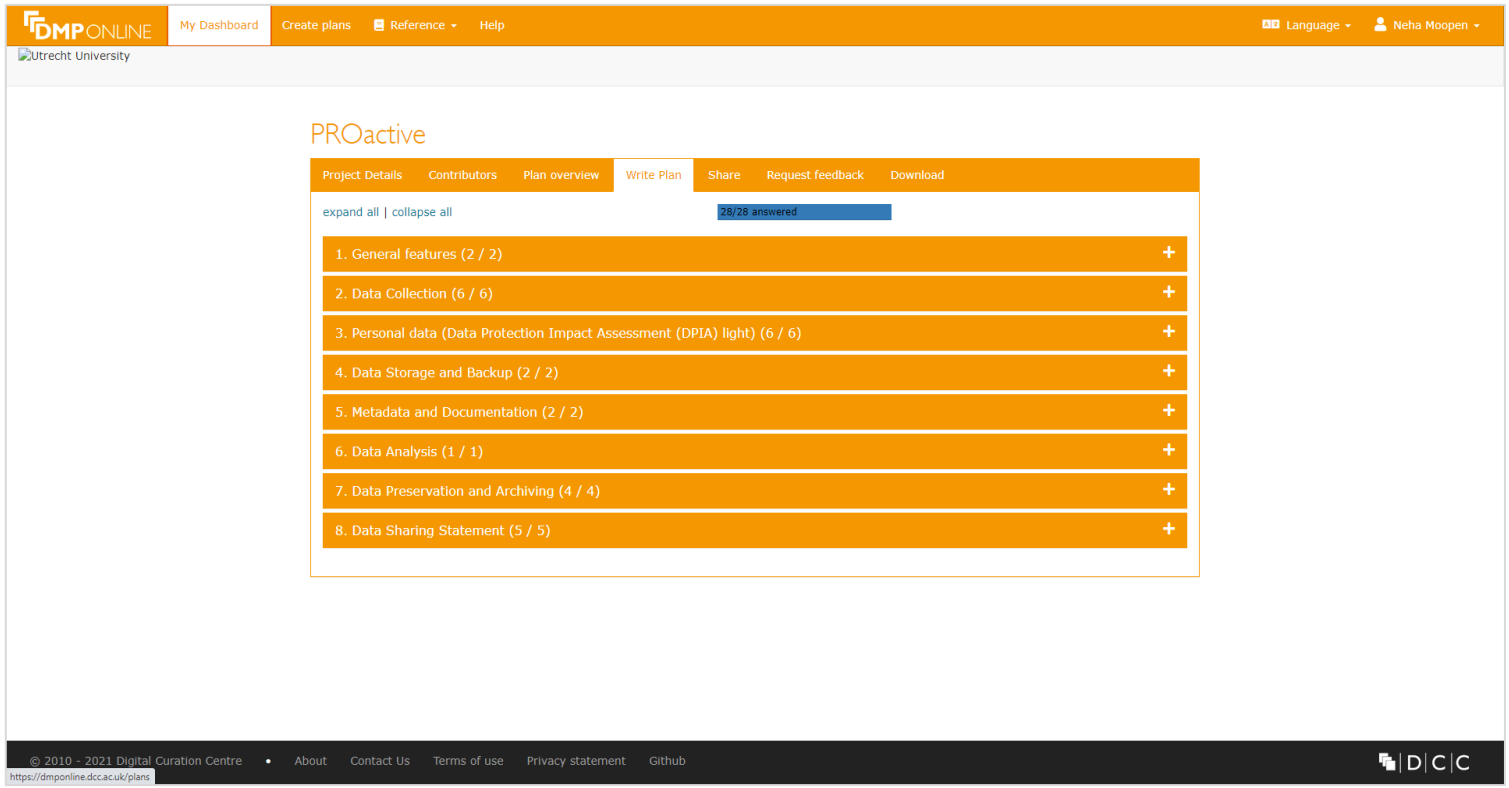
PROactive: THE FAIRification PROCESS



1 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.

2 PROJECT ORGANIZATION & DOCUMENTATION

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

- FOLDER STRUCTURE:
  Luckily, the UMCU has recommendations on setting up your folders in a structured way.

- PROJECT DOCUMENTATION:
  The RFS sub-folders provide an initial suggestion of all the administrative materials you should maintain.

- FILE NAMING CONVENTIONS:
  File names should be created and maintained in a logical and consistent way, by all project members. For best practices, see CESSDA's and OSF's guides.

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

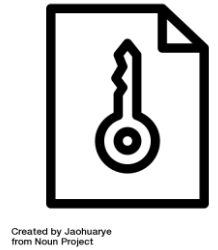
3 GDPR COMPLIANCE

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.

4 COHORT ADMINISTRATION

Is there an alternative to not-very-secure Excel spreadsheets for participant administration/management?

The YOUth cohort utilizes SLIM for this. MS Access is known to be used by another cohort.

What are the pros & cons? Are there more options?
We're still investigating this! 🔎🔎

5 DATA CAPTURE SYSTEMS

The choice of data capture tools can make your data flow more efficient and ultimately, easier to FAIRify. For example, tools which work with non-proprietary file formats are better for interoperability. Moreover, tools which enable easier data exports will eventually make data storage, publication, and sharing easier.

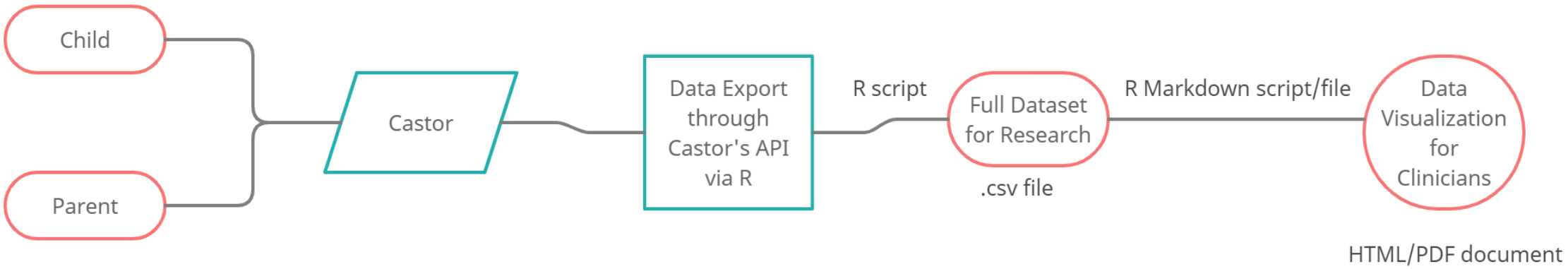
It's good to have some options laid out and do a pros & cons comparison on criteria that are relevant for the project.

	KLIK	CASTOR	REDCAP	LIME SURVEY	QUALTRICS	FORMR	RESEARCH ONLINE
SEPARATE ACCOUNT FOR PARENTS & CHILDREN	✓	✓	✓	✓	✓	✓	✓
ATTRACTIVE OR CUSTOMIZABLE INTERFACE	✓	X	X	✓	✓	✓	?
LINK TO HiX	✓	X	X	X	X	X	?
EASY DATA EXPORT	X	✓	✓	✓	✓	✓	?
UMCU APPROVED/SUPPORTED	✓	✓	✓	X	X	X	?
GDPR COMPLIANCE	✓	✓	✓	✓	✓	X	?
EASE OF USE & SUSTAINABILITY	✓	✓	✓	✓	✓	X	?
COSTS							

6 DATA EXPORT & COMBINATION

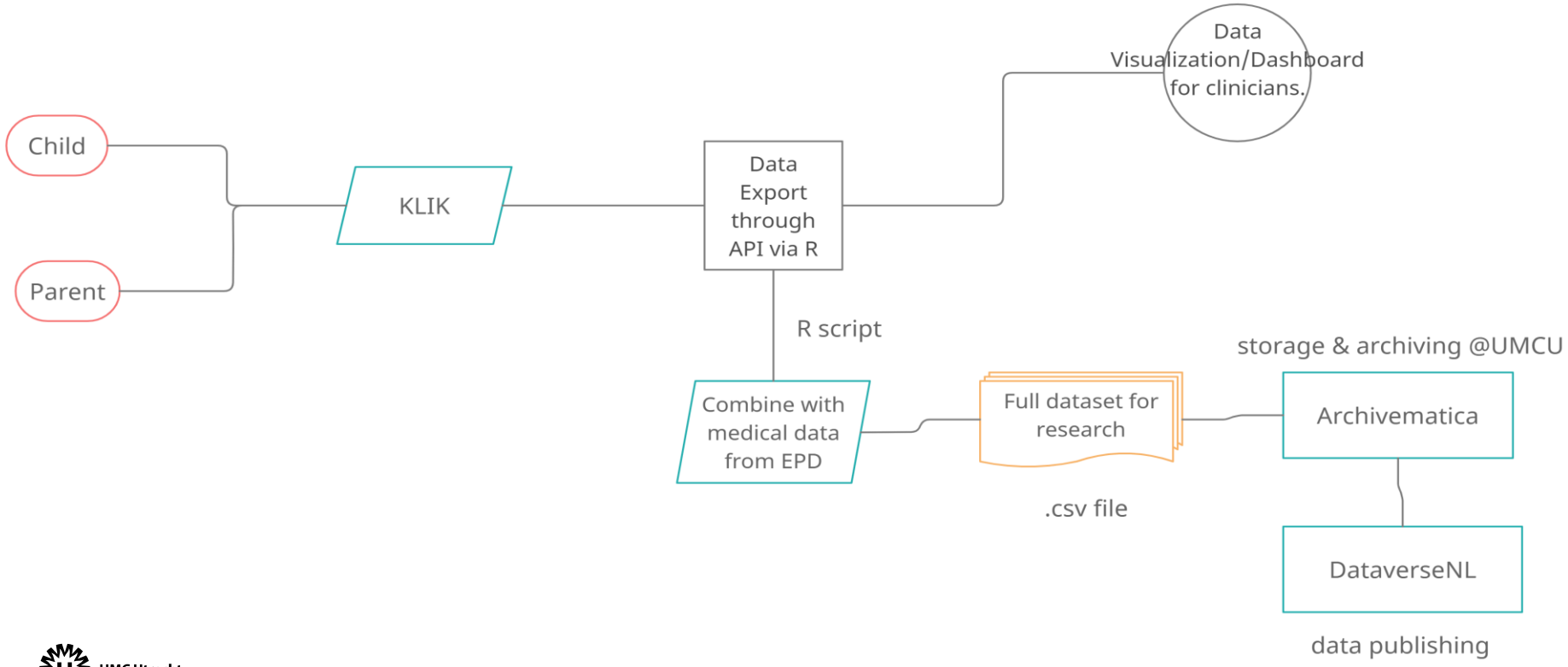
Linked to the previous point, an efficient data flow = easier FAIRification in the future. As far as possible, you want try and automate the process of data export and combination. Moreover, you want to export to data formats that are interoperable.

This (not complete, nor official) example goes from data collection to research and clinical care.



7 DATA STORAGE & ARCHIVING

This is where you can see everything coming together! The data collection tools -> automated data export & combination -> storage (and eventual publication) in UMCU-approved systems.



8 METADATA & DATA HARMONIZATION

Metadata and data harmonization are crucial to making data truly FAIR.

- METADATA:
  In the form of codebooks/data dictionaries + accompanying documents & materials add context to the data. An end user should be able to rely on this metadata to understand the data without having to contact the original researchers.

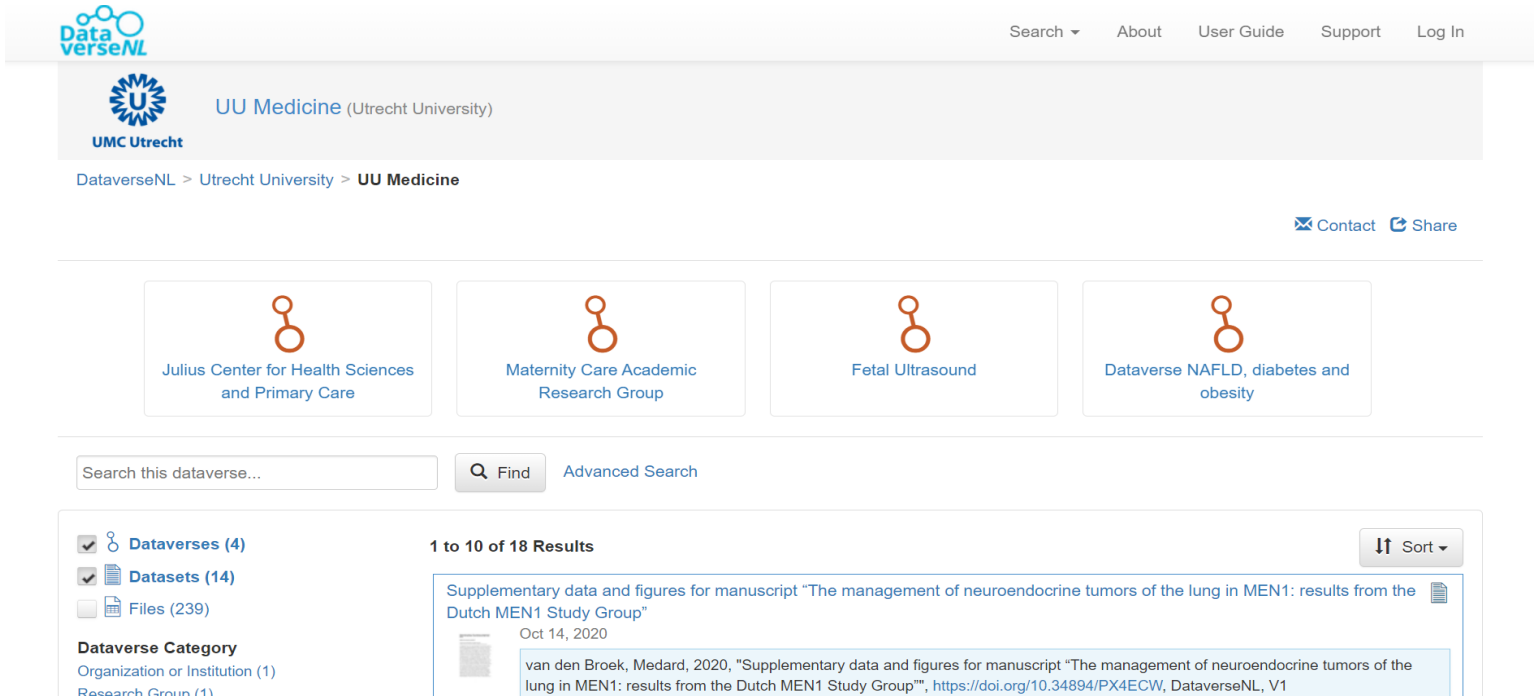
- DATA HARMONIZATION:
  Aligning variables and their meaning + coding across studies. This is no small task, there should be a metadata scheme that one can adhere to / or it should be developed from scratch.

  See this link for an example of how the Consortium on Individual Development is coming together to work on this. Another example is the EU Lifecycle Project's variable catalogue.

9 (META)DATA PUBLICATION

DataverseNL is the repository where UMCU projects can eventually publish their data. Guidelines are currently being developed for this.

DataverseNL offers various levels of access, so you could publish under restricted access (rather than open access), with instructions on how end users can request the data via the Data Access Protocol. You could also choose to only publish the metadata and release the data only when a request is made + approved.



screenshot: https://dataverse.nl/dataverse/UUMedicineUMCU

10 DATA SHARING (1)

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

10 DATA SHARING (2)

The last step! And possibly the most confusing one at times….
How you share your data will depend on a number of factors:

- **Does the end user only need your data alone to answer their research questions?** Then the Digital Research Environment (DRE) may be suitable.

- **Does the end user need to use your dataset in conjunction with another? For example, merging datasets for a larger cohort comparing healthy vs. clinical samples?** Depending on where the data is being transferred to, you can complete a *Data Use Agreement** (within UMCU) or *Data Transfer Agreement* (within and outside EEA, will have to be tailored depending on which). *Consortiums* can set up a DTA specific to their needs & requirements too.

Utrecht University UMC Utrecht

* the DUA does not actually exist, but I'm trying to set one up!

That's the short version!

READING TIP

FAIR, safe and high-quality data: The data infrastructure and accessibility of the YOUth cohort study

QUESTIONS?

You can get in touch anytime!
n.moopen@uu.nl

The information in this presentation has been compiled with the utmost care,
but no rights can be derived from its contents.