# Data set list – Data management plan

This template is designed to help you with creating an overview of the different sets of data sources you use during your research for that research. This list is meant to use as a lightweight data management plan, i.e. a plan in which you describe how you will handle the sources you use within your research during that research and how you will deposit these sources afterwards to warrant transparency, cq. research replication.

This form consists of two tables. In the first table you can list the sets and in the second one the ‘personal elements’ per set. The exact meaning of the columns for the first table is explained on the final page.

Between the two tables you find a number of headers and some example texts. These are there to inspire you to give extra information. Some of that information is simply mandatory, e.g. the retention period of the package. Bottom line is that someone not involved must be able to have good insight in the central plans regarding your data handling.

Please be aware that data management is aways about two separate phases: the time during your research project and the, longer, phase after you have finalized your research. Please be aware that in practice, you will have little time or attention for your research data in that second phase, as you will be working on a new research project.

Also be aware that much depends on your choices and ideas. Do you want a fully replicable research or is ‘transparent’ enough; much has to do with what is possible and/or preferred (much qualitative research can’t be ‘replicated’ ; much research with very personal data cannot be fully transparent).

Be as complete as possible but list only sets that you will manage yourself. The secondary literature and archival sources are managed by other parties (libraries, archives), so you do not need to account for those (the exception being the consent-forms, participant-lists and key-files; strictly speaking are those not research data but they need to be ‘managed’).

Differentiate as much as makes sense. If you have two participant groups each going through the same interview protocol, you might consider these as one dataset. On the other hand if you have one set of participants, but two different interventions this will end up as two sets.

This DMP-template is accepted by the FETC of the Humanities faculty

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ID** | **Type** | **Format** | **Storage during research** | **Storage after research** | **GDPR?** | **Art. 9** | **Other rights** | **Data subjects** | **Legality** | **Anon/pseud deriv** | **Accessibility****during research** | **Accessibility****after research** | **Access restricted to** | **Remark** |
| 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 8 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 9 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 10 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

# Remarks

***About the project***

Who is PI, who is all involved.

***About the sample (size & sampling method)***

Who participates, what is their position?

***About the GDPR***

Will personal data be involved and on the basis of what [legal ground](https://gdpr.eu/article-6-how-to-process-personal-data-legally/) (Consent or Public Interest? Are any activities foreseen w.r.t. anonymizing/pseudonimyzing? What happens to the RAW data, if any? If based on consent, how and for how long will the consent (forms) be stored?

***Retention Period and use license (conditions for use)***

Default retention period is a minimum of 10 years. If sets needs to be treated GDPR-compliant (i.e. personal data involved), than it is a maximum of 10 years.

***Identification and description of archival data packages***

If a non-standard UU solution (i.e. not [Yoda](https://www.uu.nl/en/research/yoda) or [DataVerse](https://www.uu.nl/en/research/research-data-management/tools-services/tools-for-storing-and-managing-data/data-repository-finder/the-research-data-repository-dataversenl)) is chosen for data archival

***Handling personal data***

Any specificalities, e.g. with regards to storage of RAW data, separation of research data and administrative data over different archival packages with different accessibilies.

***Data transport***

Will any data be transported, e.g. from a handheld device to a server. How ewill that take place (esp. interesting in certain countries. You might want to consider use of VPN and a procedure for immediate deletion on the (encrypted) handheld.

***About the sets***

**Ad(1) –** anything of special interest

# Personal data

|  |  |  |  |
| --- | --- | --- | --- |
| **Personal element** | **Reason for processing** | **Dataset** | **Method of anonymizing** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Explanation of the columns**

* *Personal element* – The element of personal data you will register. Try to be as specific as possible.
* *Reason for processing –* Explain why you need this particular element of personal data.
* *Dataset –* Identify the datasetsin which this particular element will occur.
* *Method of anonymization –* if you create a derivative without this particular element, indicate how you will anonymize/pseudonymize.

**Explanation of the Data Set List**

* *Type* – Short description of the data set.
* *Format –* indication of what format the data is in, e.g. text, tabular, audio, video – helps to identify what different things you can do. For example: a video is hard to transcribe and anonymize.
* *Storage during research –* list all locations where the data of this specific set will be stored during research. This could be more than one, e.g. a recording device and a central server.
* *Storage after research –* list all the location of the data archive where the data of this specific set will be stored after research, if at all.
* *GDPR* – indicate whether the data set contains [personal data](https://cdh.uu.nl/research/data-management-privacy/guide/working-with-personal-data/).
* *Art 9 –* to indicate whether a dataset contains [special categories of personal data](https://cdh.uu.nl/research/data-management-privacy/guide/working-with-personal-data/) (as defined in Article 9 of the GDPR).
* *Other rights –* Are there any rights, e.g. Intellectual property - or Copy Rights resting on the data set. These could also be laid down in consortium-, transport- or processing agreements. Any constrictions from Informed Consent agreements are considered to b covered by the previous columns.
* *Data subjects –* list the people whose data you will be processing, if any.
* *Legality –* in case of working with personal data, what is (are) the legal grounds for your data processing*.*
* *Anon/pseud deriv –* If a data set is an anonymized/pseudonomynized, derivative, what other dataset is it derived from (use the i*d).*
* *Access during –* who has access to the data during research cycle.
* *Access after –* What is the accessibility of the data once archived (Closed/Restricted/Open).
* *Access restricted to –* In case of restricted access, to who will it be restricted (e.g. *Research team)*
* *Remark –* Any remark to explain details.
	+ **NB 1 If you store data on different locations** e.g. first on your laptop and later on Yoda, please use this field to describe the transport protocol to get data from location A to B. This holds specifically when personal data is involved
	+ **NB 2** If you **don’t** follow up this template with a data management plan, but want to use this for an internal UU procedure, e.g. a FETC assessment, then please add the archival retention period (default 10 years) in the *Remark* field.