# Data set list – Data management plan

This template is designed to help you creating a datamanagement plan, i.e. an overview of the different sets of data sources you use during your study as well as a description on how you will handle these sources, both during your study as well as afterwards to warrant transparency, cq. research replication.

First you will be asked to provide some general information about the document-version, the people involved, study title etc. Then you are asked to list the sets of data sources used and to provide relevant characteristics with regards to (long term) management, e.g. whether they contain personal data, planned storage location, etc. If there are sets involved containing personal data then you are asked to fill out the [processing grounds](https://gdpr.eu/article-6-how-to-process-personal-data-legally/) and provide details in the second table.

In general, some plans are for all the sets, e.g. the retention period (default ten years) and planned user license. However, in some cases a more differentiated approach is required or special circumstances are involved; for example when some sets contain [special categories of personal data](https://gdpr.eu/article-9-processing-special-categories-of-personal-data-prohibited/), or when data is captured on handhelds and need secure data transport to a server. For those cases you are offered a number of topic-titles which you can use to explain details/issues you cannot leave in the columns, or which hold for all sets. Please fill this out a much as possible. Bottom line is that someone not involved must be able to have good insight in the way you plan to manage each set both during your research as well as afterwards.

Please be aware that data management is always about two separate phases: the time during your actually doing the study 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 data sources in that second phase, as you will be working on a new research project.

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.

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’ nevertheless.

Some of the choices within your data management are fully up to you, others depend on the characteristics of your data sources. If your sources contain no identifiable personal data, you can choose any repository for long term storage which you’d like. On the other hand, if they do contain this type of data, the archival storage must take place in a repository within the European Economic Area. Also be aware that much depends on your choices and ideas. Do you want a fully replicable research or is ‘transparent’ enough. Not every preference is possible and not every possibility is preferred; much qualitative research can’t be ‘replicated’, research with very personal data cannot be fully transparent

If you find yourself struggling with the options involved, please contact the data manager of the [Humanities faculty](mailto:datamanagement.gw@uu.nl).

This DMP-template is accepted by the [ethical committee](https://fetc.hum.uu.nl/landing/?next=/) of the Humanities faculty.

# *Data list <Study title>*

|  |  |  |  |
| --- | --- | --- | --- |
| **Document version** |  | **Date last edited** |  |
| **Principal Investigator** |  | **Writer of this DMP** |  |
| **Funder** |  | **Grant number / URL** |  |
| **Preferred Start date** |  | **Planned end date** |  |
| **Faculty** |  | **Related FETC assessments** |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **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 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

# Remarks

***About the project***

**Project abstract:**

What is, in general terms, the goal and set up (methods) of your research?

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

Who participates, what is their position, how do you get to them, what relation (e.g. hierarchical) ?

***About the GDPR***

If personal data are involved, how will these specific data sets be handled? E.g. if any activities foreseen w.r.t. anonymizing/pseudonymizing, please explain how this takes place? What happens to the RAW data, if any? If based on consent, how and for how long will the consent (forms) be stored? Will there be a separation of research data and administrative data over different archival packages with different accessibilities? Try to list all relevant issues.

***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. Under what use license will datasets be offered (default [CC-BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/deed.nl) – note, this only hold for public accessible data).

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

In case 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, please explain how the datapackage(s) can be identified and localized and what metadata will be available.

***Data transport***

This part is esp. of interest when personal data are involved. Will any data be transported, e.g. from a handheld device to a server. How will 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. If you share data with researchers from other institutes, is there a transport/processing agreement in place. (PO’s have templates for this).

***Software and sustainability***

If software code are part of the data sources which need long term management, are there any specific provisions cq. needs? E.g. with regards to licenses? Also, have you considered the sustainability of your file-formats? A sustainable digital format is one that is compatible, for the foreseeable future, with software needed to open and read it, e.g. by storing Excel-files in a CSV-format.

# Personal data

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

# Explanation of the columns

**Data Set List**

* *Type* – Short description of the data set. Examples: *Raw Interviews, Raw interview transcriptions, Surveys outcomes, Anonymized survey outcomes*
* *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://gdpr.eu/article-4-definitions/).
* *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 be covered by the previous columns.
* *Data subjects –* list the people whose data you will be processing, if any.
* *Legality –* if 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.

**Personal data**

* *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.