Utrecht University School of Economics

Data Management Plan

Draft Version

Date first version:	YYYY-MM-DD
Current version:	YYYY-MM-DD

Table of Contents

Organisational Context

Name of researchers:	[…]
Name of project:	[…]
Funding bodies:	[…]
Partner organisations:	[…]
Project duration:	Start: YYYY-MM-DD End: YYYY-MM-DD
Responsible for data management:	[…]

Description of the Research

Utrecht University School of Economics plans to conduct a controlled field experiment (randomised controlled trial) to assess the (dis)advantages of alternative social assistance configurations. For a period of two years groups of social assistance claimants in Utrecht will receive benefit payments under varied conditions. These conditions concern rewards and sanctions, as well as reintegration obligations. The experiment aims at providing sound scientific evidence about which forms of social assistance work best to achieve successful and sustainable reintegration of claimants into the labour market and society. To assess the effectiveness and efficiency of different tested forms of social assistance we collect data on claimants behaviour, well-being, satisfaction and financial situation, as well as on the cost of a scheme. To collect data we will make use of subsequent surveys among participants and caseworkers as well as the municipalities' administrative data pools.

Utrecht University School of Economics is responsible for the scientific aspects of the experiment. This includes primarily experimental design, data collection and data analysis.

1. Preparation

1.1 Data Collection

The objective of the research is to evaluate the effectiveness and efficiency of different forms of social assistance. To do so we collect data on *individuals* and the involved *administrative bodies*. Individuals thereby refers to (i) participating social assistance claimants, (ii) non-participating claimants, and (iii) caseworkers. Administrative bodies includes the involved municipalities.

The remainder of this section is organised as follows. We describe separately for individuals and administrative bodies the information we are planning to collect as well as data collection procedures. Besides, we provide a classification of data into different categories depending on the stage of data processing.

1.1.1 Individuals

In what follows we specify for each of the three parties what kind of information we aim to collect from which data sources. Next to collecting our own data (survey data) we make use of existing data (administrative data) which we collect from the involved municipalities' data pools. Most of the data we plan to collect is personal data, which is subject to strict privacy regulations. As different types of personal data require different treatment we provide a classification based on the classes of personal data outlined in the Dutch *Data Protection Act* (WPB – Wet Bescherming Persoonsgegevens). Those are: (i) directly identifying data, (ii) indirectly identifying data, and (iii) sensitive data. Please see the glossary at the end of the document for further information on data classes and subclasses. Table 1 provides an overview of data types and sources.

To collect personal data we will make use of two data sources: individuals and the involved municipalities' data pools. To collect data from individuals we will make use of surveys. Our survey method will be face-to-face computer-assisted personal interviews (CAPI) with questionnaire elements for personal topics such as health and well-being. Collecting information on health and other sensitive topics is necessary as the evaluation of the different schemes will also be based on indicators such as mental and physical health, or work stress.

Table 1. Data from individuals

Data source	Data class	Data subclass	Data specification
Participating claimants			
Municipality (external data)	Personal data	Directly identifying	• First name • Last name • Initials • Address
Municipality (external data)	Personal data	Indirectly identifying	• Gender • Arrangement[1] • Welfare history • (Non)compliance[2]
Individual	Personal data	Directly identifying	• Date of birth • Phone number • E-mail address • IBAN[3]
Individual	Personal data	Indirectly identifying	• Country of birth • Place of residence • Civil status • Educational attainment • Household information • Attitude towards work
Individual	Personal data	Sensitive data	• Ethnicity • Physical and mental health • Mental capacity
Non-participating claimants			
Municipality (external data)	Anonymised data	–	• Age • Arrangement • Civil status • Ethnicity • Compliance • Cooperation
Caseworkers			
Individual	Personal data	Directly identifying	• First name • Last name • Initials • Date of birth
Individual	Personal data	Indirectly identifying	• Gender • Employing municipality • Duration of employment • Work satisfaction
Individual	Personal data	Sensitive data	• Work stress

The surveys will be administered digitally making use of tablets and CAPI software programmes. Such programmes usually have powerful exporting tools, allowing us to export survey results into spread sheet format. We choose Excel to be our spread sheet file format during the research. Our spread sheet file format for long-term storage will be comma-

[1] Arrangement: A claimant's distance to the labour market.
[2] Compliance: Sanctioning of claimants in case of non-compliance.
[3] Collecting participants' bank details is necessary to transfer fees for filling in surveys.

separated values (.csv) as it is non-proprietary and future-proof. Files containing survey data will be protected as they contain (in)directly identifying and sensitive personal data (see 3.2 Data Access and Security).

To collect personal data from municipalities the involved administrative bodies will provide us with files that contain selected data from their administrative data pools. The data will be extracted from the data pools by administrative officials at the different points of measurement. Data files will be provided in Excel format. We will collect two types of data from the municipalities: Personal data on participating claimants and anonymised data on non-participating claimants. Data files from the municipality will not contain directly identifying information (see below). See 3.2 Data Access and Security for more information on data ownership and responsibility for external data.

In total there will be four points of data collection: Before the experiment starts, after the first year, after the second year, and six months after the experiment has ended. Data collection thus comprises a period of 2.5 years.

To guarantee confidentiality of personal data all information on individuals (data subjects) will be cleaned from directly identifying information during the research. Directly identifying information will be replaced by a unique personal identifier (pseudonymisation), which is a meaningless administration number unrelated to personal characteristics of the data subject. In order to communicate with participants and match data from subsequent collections a separate protected file that contains contact information and personal identifiers will be created. Access to contact data is restricted (see 3.2 Data Access and Security). Pseudonymisation takes place at the universities (for survey data) and the municipalities (for administrative personal data) and is executed by trusted third persons that are not in an hierarchical relation to other members of the research team.

1.1.2 Administrative bodies

Next to data on individuals we aim to collect financial data from the involved administrative bodies.

Table 2. Data from administrative bodies

Data source	Data class	Data subclass	Data specification
Municipality	Administrative financial data	–	• Direct cost • Personnel cost • Administrative cost

1.1.3 Overview of data

In total we distinguish between six categories of data depending on the stage of data processing. Those are:

- Contact data
- Survey data
- Administrative individual data
- Administrative financial data
- Processed data
- Statistical data

Table 3 provides a descriptive overview of the data collected and processed during the research using the above mentioned categories. It also provides further information on formats, software, sizes and reproducibility. With regard to the latter we depend to a large extent on survey data, which is non-reproducible. As a consequence, the master files for raw survey data will be stored separately and write and access protected (see also 3.2 Data Access and Security).

Table 3. Data overview

Data category	Description	Collection	Format	Software	Est. size	Est. tot. size	Source for pub.
Contact data	Protected file containing (i) directly identifying personal data such as name or address for communication purposes, and (ii) unique personal identifiers to match and clean data.	Provided by municipalities and completed by survey data; non-reproducible	.xls	MS Excel	KBs	KBs	No
Survey data	Protected and locked files containing raw survey data; cleaned from directly identifying information.	Provided by researchers (surveys); non-reproducible	.xls	MS Excel			Indirect
Administrative individual data	Protected locked files containing raw administrative data including (i) personal data from participating claimants and (ii) anonymous data from non-participating claimants; cleaned from directly identifying information.	Provided by municipalities; reproducible.	.xls	MS Excel			Indirect
Administrative financial data	Files containing financial data from the participating administrative bodies.	Provided by municipalities; reproducible	tba	tba			Indirect
Processed data	Protected files containing quality checked and adjusted data for further statistical analysis.	Provided by researchers; reproducible	.xls	MS Excel			Indirect
Statistical data	Files containing the results of statistical analyses.	Provided by researchers; reproducible	tba	tba			Yes

1.2 Data Documentation

1.2.1 Metadata

During the research we make use of descriptive meta data schemes for the files we work with in order to find and interpret specific data more quickly and effectively. We will develop templates of metadata schemes in the form of Excel worksheets that can be filled with controlled vocabulary (e.g. by drop down). The information provided by our metadata schemes depends on the file:

- Raw survey data: The metadata includes e.g. information on locations, survey wave, collection period, interviewer, etc.
- Processed data: The metadata includes information such as author, changes to last version, data sources, codes and abbreviations, etc.

We thereby aim to make use of the DDI standard (Data Documentation Initiative), which is a widely used, international standard for describing data from the social, behavioural, and economic sciences. DDI is particularly suited to manage longitudinal datasets.

After the research has finished we compile data packages that will be transferred to a public repository. We will then make use of the repository's metadata standard to provide information about our data.

1.2.2 Documentation

Documenting our research we compile three documents following DDI standards.

- A codebook that provides variable descriptions and coding to make coded data understandable.
- A manual explaining our experimental design and methodological approach including e.g. sampling and randomization.
- A survey guide that documents the process of data collection among individuals, including our questionnaires.

1.2.2 Directory and file naming convention

During the research we will work on a shared networked drive. We plan to establish the following scalable folder structure. Folders that contain privacy sensitive information will be encrypted and access restricted. Our contact file with directly identifiable personal data and unique personal identifiers will be stored in a separate location.

Project Folder

- **1 Project Management**
  - 1 Proposals
  - 2 Planning
  - 3 Financials
    - 1 Budget
    - 2 Funding
  - 5 HR
  - 6 Internal Communication
  - 7 Meetings, Notes and Minutes
- **2 Ethics / Governance**
  - 1 Guidelines and Policies
  - 2 Information Material
  - 3 Consent Forms
- **3 Theory**
  - 1 Theoretical Literature
  - 2 Models
- **4 Empirical Data**
  - 1 Surveys
  - 2 Raw Data [Protected]
    - 1 Measurement 1
      - 1 Survey Data
      - 2 Administrative Data
    - 2 Measurement 2
    - 3 Measurement 3
    - 4 Measurement 4
    - 5 Caseworkers
  - 3 Processed Data [Protected]
    - 1 Master Files Claimants
    - 2 Master Files Caseworkers
    - 3 Financial Data
  - 4 Data Analysis
  - 5 Outputs
- **5 Dissemination**
  - 1 Publications
  - 2 Reports
  - 3 Publicity
  - 4 Conferences
  - 5 Presentations
- **6 Team Folder**
  - 1 Personal folder team member 1
  - 2 Personal folder team member 2
  - 3 …
- **7 Miscellaneous**
- **8 Archive**

As we share a file space and exchange files with partner organisations we decide to apply a standardised file-naming convention. Our file names consist of (i) an abbreviation for the organisation, (ii) a content description (generic to specific), (iii) the date of modification (international standard: YYYYMMDD), and (iv) a version number (v0.0), separated by an underscore. We do not use special characters, full stops or spaces.

Example: UU_SurveyQuestions_20160424_v5.0.docx

2 Data Handling

2.1 Data Storage and Backup

2.1.1 Daily storage

During our research we collect, process and analyse digital data. All our data (except contact data) will be stored in an access restricted project folder. The project folder will be stored on generic networked infrastructure of Utrecht University. Currently, the strategic theme 'Institutions for Open Societies' of Utrecht University is building the so-called I-Lab which will provide facilities for safely storing research data. As soon as this facility is available (presumably already late 2016), it will be used for daily data storage during the research. Contact data with directly identifiable information will not be stored on a shared drive, but at a protected and encrypted distinct location.

The project folder will be our master copy location, thus the location of the most current and correct file and basis for all back-ups. Using the university's networked infrastructure comes with several advantages. First, it allows all team members to access the data. Second, access to the data is not device specific. Third, the university's infrastructure is a secure storage environment. Fourth, there exists a back-up regime that backs up data automatically, regularly and encrypted.

As most of the research data is privacy sensitive it is not planned to use personal computers or portable devices such as USB flash drives to (temporarily) store research data. An exception is a hardware encrypted and robust back-up hard drive. It will be used to back up data in a second and physically distinct location. Back up to this location will be incremental and take place once a month. Responsible for these manual back-ups is the team member responsible for data management.

2.1.2 Data exchange

As exchange of indirectly identifiable data takes place with involved municipalities a secure data exchange solution is required. Portable devices such as USB flash drives pose a security risk and need to be exchanged manually. Cloud services allow for fast and easy data exchange, but also come with considerable data safety concerns. We thus choose to exchange data via SURFfilesender, a service offered by the Dutch national data centre SURF with which researchers can send research data and other confidential files safely and quickly.

SURFfilesender comes with several advantages. The uploaded files are stored in the Netherlands for no more than 21 days. Extra security in the form of encryption is possible.

Users of SURFfilesender don't have to install anything in order to send and receive files, the sender and recipient only need a modern browser.

Full SURFfilesender functionality is available to Dutch education and research institutes. Guest access is possible, to enable the safe exchange of files with individuals without a SURFfilesender licence. Data exchange with the involved municipalities using SURFfilesender is thus possible as well.

2.1.3 Data streams

The streams of data are visualised in Figure 1 and 2. Before the experiment municipality and university have to exchange data as the university is responsible for randomisation of participants into control and treatment conditions. The chosen procedure makes sure that social assistance claimants' privacy is protected:

- The municipality compiles a list of possible participants (target population) according to several criteria, e.g. receiving social assistance for more than six months, or no personal insolvency. Thereafter the list will be pseudonymised. Next to personal identifiers the cleaned list contains personal information on age, arrangement, nationality (native/foreign), and civil status.
- The cleaned list is then sent to the university, where the randomisation of participants takes place and is added to the list. The randomised list is sent back to the municipality.
- The municipality invites claimants on the list to participate in the experiment. In case claimants agree to participate they are asked to sign a consent form that allows the collection and processing of their personal data.
- Signing claimants become participants. A contact file which lists all participants and their unique personal identifiers is compiled and shared with the university. This contact file will be used by the municipality and the university for following data subjects over time and pseudonymisation during the experiment.

During the experiment, the municipality collects data from its administrative data pools and pseudonymises the data before is sends the data to the university. The university collects and pseudonymises survey data. Both data files can be matched based on unique personal identifiers.

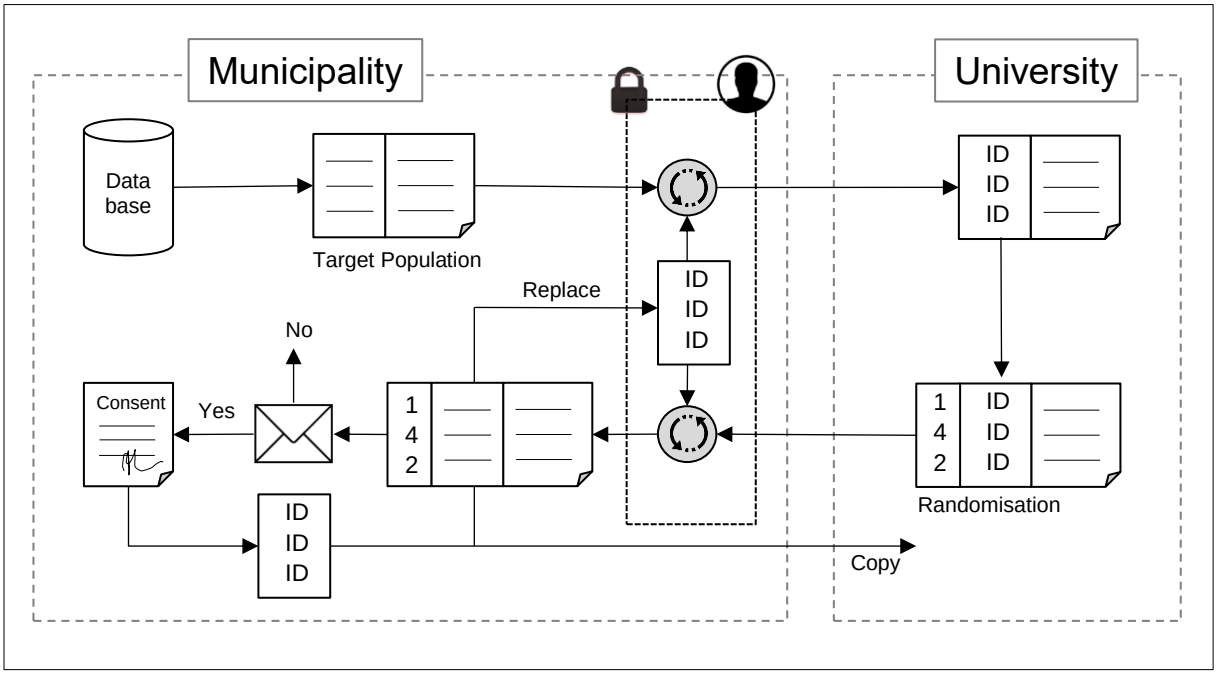

Figure 1. Data stream before the experiment

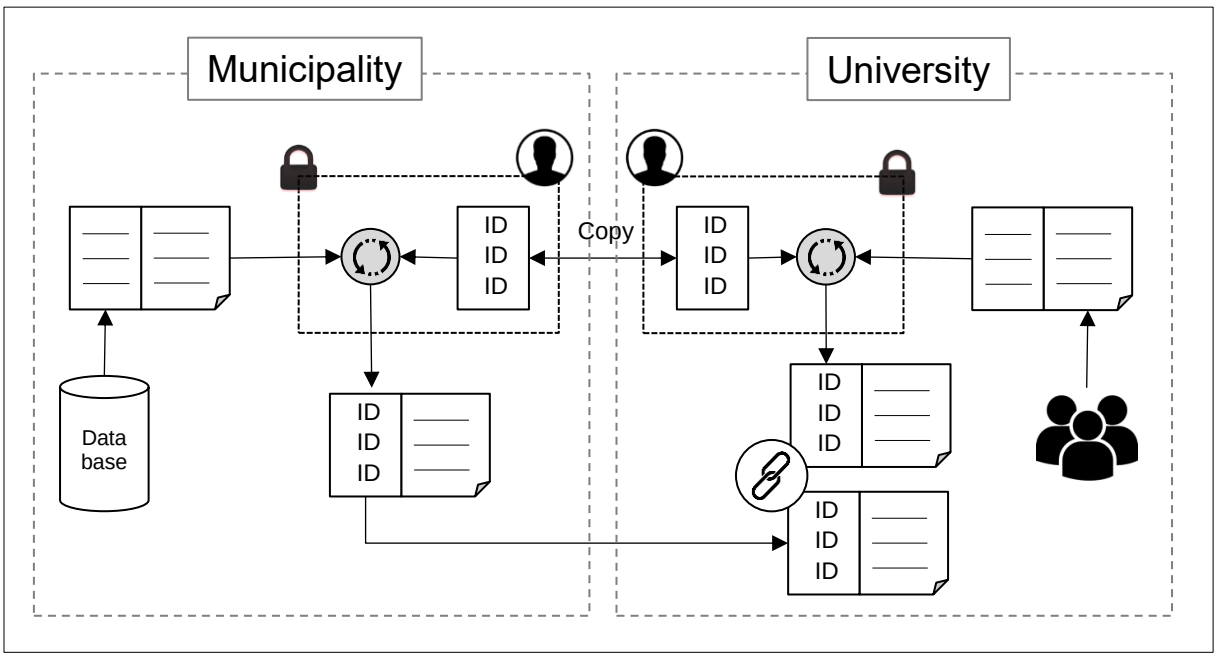

Figure 2. Data stream during the experiment

2.1.4 Version control

As data is being used by several members of the research team and exchanged with other involved parties we will implement a version control regime. In doing so we will follow best practices, which includes:

- Working with a Master copy and a Master copy location, which are the respective subfolders of our project;
- Not to overwrite old versions, but creating a new version of the Master copy in case of changes;
- Using the extension "v0.0" in the file-naming convention to track major and minor changes;
- Document the changes associated with the versions in a 'version history file';
- Maintaining the original as well as the most current version and moving intermediate versions to an 'Old Versions' folder that is cleaned up regularly.

The versioning process we plan to follow is visualised in Figure 3.

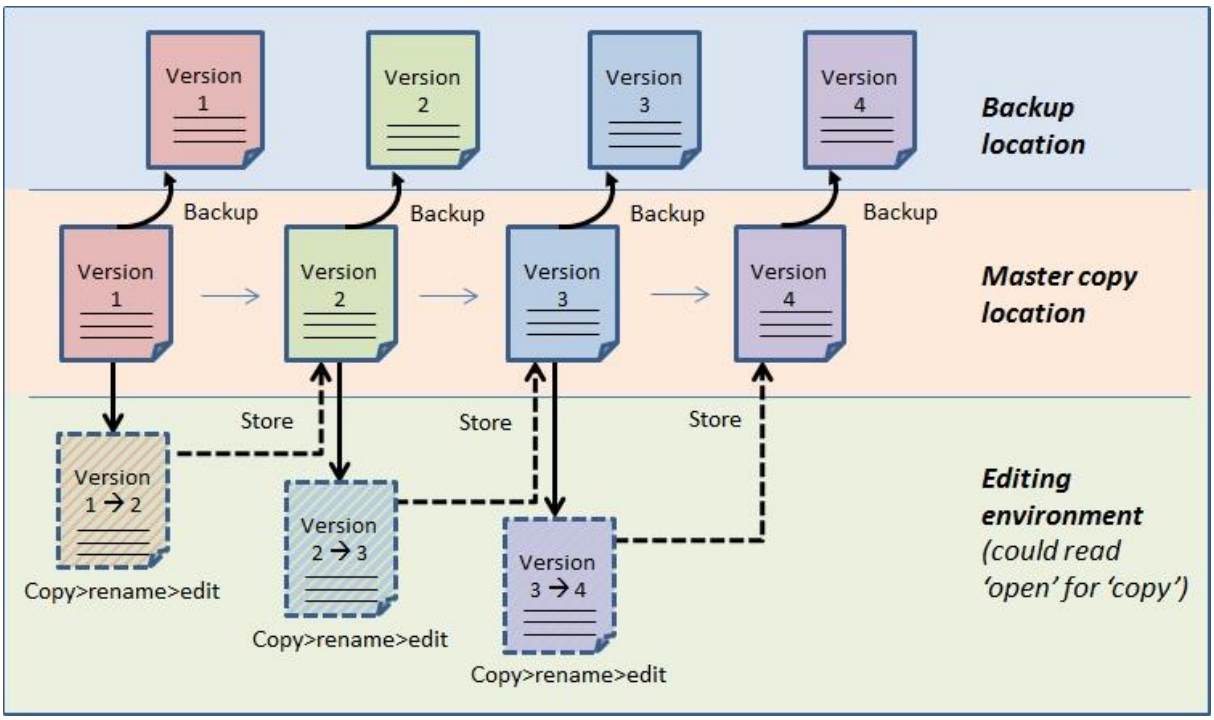

Figure 3. Versioning process (*Source*: Utrecht University Library)

2.2 Data Access and Security

2.2.1 Data ownership and responsibility

With regard to data ownership and responsibility we have to distinguish between data that is collected by the university (survey data) and external data from the municipality, which is shared with the university (administrative individual and financial data). Ownership and responsibility for the survey data lies with the research team. Ownership and responsibility for raw and uncleaned administrative data lies with the municipality. Likewise, the municipality is owner of cleaned administrative data. The municipality allows the university to process, store and analyse cleaned administrative data within the scope of the research and agrees to the publication of anonymised data after the termination of the research. After the exchange of data, the university is responsible for cleaned administrative data. However, with the exchange of data between the two parties ownership of data does not change. Data ownership and responsibility is agreed on in the consortium agreement signed by all participating organisations.

We collect data on individuals and administrative bodies, which is why we have to consider the rights of our data subjects and the rights of the municipalities. With a view to data subject's rights the protection of personal data is of large importance and we will adopt data management procedures that protect data subject's privacy (see following section).

2.2.2 Protection of privacy sensitive data

During our research we collect privacy sensitive information from participating individuals and caseworkers. Those are directly identifiable, indirectly identifiable and sensitive personal data. Collecting, processing and storing this kind of data requires to develop data management procedures that comply with the legal frameworks and guarantee confidentiality and data protection.

Dutch laws that are related to our research are first and foremost the *Data Protection Act* (WPB – Wet Bescherming Persoonsgegevens). The WPB has been translated into a code of conduct for researchers that is approved by the *Dutch Board of Data Protection* (CBP – College Bescherming Persoonsgegevens). Our data management procedures strictly follow the principles outlined in this *Code of Conduct for Scientific Research* (Gedragscode voor gebruik van persoonsgegevens in wetenschappelijk onderzoek). In what follows we will apply those principles to our research:

Art. 3.1 – Legal use

- We will collect, process and store personal data according to the legal obligations and restrictions.

Art. 3.2 – Restrictive data collection

- We will only collect personal data that is necessary to answer our research questions. This holds particularly true for sensitive personal information .
- Directly identifiable information (contact data) will be collected as we need to contact our data subject for subsequent measurements.

Art. 3.4 – Collecting data from individuals

- Data subjects will be asked for their consent before we start collecting, processing and storing their personal data (see below). They can stop participating in the research at any point of time.
- We will only collect, process and store personal data from consenting data subjects.
- Before data collection starts data subjects will be informed about the goal of the research, the contracting authority, the organisation executing the research and a contact person they can approach for further questions and information.
- There will be no video or audio recording.

Art. 3.5 – Collecting data from existing data bases

- Data subjects will be asked for their consent to collect, process and store their personal data collected from the municipality's data bases.

Art. 3.6 – Directly identifiable information

- We will separate directly identifiable information (contact data) from other research data. Contact data will only be used to contact and follow data subjects throughout the research.
- Contact and other research data will be stored at different locations and are subject to different access and security regulations.
- The data will be connected by a meaningless administration number (unique personal identifier).
- Contact data will be destroyed after the termination of the research.

Art. 3.7 – Data collection from third party individuals

- There will be no collection of data about data subjects from third party individuals (e.g. neighbours, colleagues, family members). We will not ask data subjects to provide information about other data subjects.

Art. 3.8 – Informing the Dutch Board of Data Protection

- The Dutch Board of Data Protection does not have to be informed about the research even though directly identifiable information will be collected, as directly identifiable

data (contact data) (i) will be stored separately from other research data, (ii) will not be stored for more than six months after the research has finished (except sex, place of residence and year of birth), and (iii) will only be used for the respective research.

- The Corporate Information Security Officer of Utrecht University will be informed about the research.

Art. 3.9 – Use of data

- The personal data collected will only be used for the outlined research.

Art. 3.10 – Storing personal data

- Contact data will not be stored any longer than six months after the research has finished.
- Anonymised data will be transferred to long-term storage facilities to allow for scientific replication. The files transferred to long-term storage will not contain any directly or indirectly identifiable information.

Art. 3.11 – Transparency and data quality

- Directly identifiable personal data will only be shared with the Municipality of Utrecht.

Art. 4 – Security

- We will adopt the necessary security measures to protect the personal data collected. This includes access control and encryption of data.
- All researchers with access to personal data will sign a non-disclosure agreement.

Art. 5 – Sharing data with third parties

- Personal data will not be shared with parties other than the partner organisations.

Art. 7 – Publication

- Scientific results will be published in a way that does not allow to trace back information to data subjects.

2.2.3 Informed consent and opt-out

Social welfare claimants that have been chosen for participation in the experiment will be invited to participate. Participation is voluntary. In case claimants agree to participate they will be informed about the consequences of their participation, the goal of the research, the contracting authority, the organisation executing the research and a contact person they can approach for further questions and information. They will be asked to sign a consent form that allows the municipality and Utrecht University to collect and process their personal data. Data collection only starts after participants have given their informed consent. Subjects can

stop participating in the experiment at any point of time. In case of an opt out no further data will be collected. Existing data will be anonymised and used for future analysis.

2.2.4 Data access and security during the project

During the research access to the research data is only permitted to the researchers involved. The project directory is therefore password-protected. Some data will be exchanged with partner organisations during the research. In general only pseudonymised data will be exchanged and data exchange will take place in a secure way (see 3.1 Data Storage and Backup). Data exchange is regulated in the consortium agreement. Due to the different types of data involved different levels of access and security need to be distinguished. For confidentiality reasons it will not be allowed to store privacy sensitive data on any other locations (e.g. portable devices, personal devices) than those mentioned below.

Contact data [privacy sensitive] – This file contains directly identifiable personal data such as name or address and is therefore subject to strict confidentiality requirements. The file is encrypted and stored on a password protected distinct location. Access to the storage location and the file is only granted to two trusted third party individuals responsible for data cleaning. These persons are [Name] and [Name]. A copy of the file will be stored under the same conditions at the municipality. The contact information file will be needed for cleaning and pseudonymisation of survey data, implying that the persons mentioned above are responsible for this task. In accordance with the Dutch Code of Conduct for Scientific Research contact data will not be stored longer than six months after the research project has finished (excluding sex, place of residence and year of birth). Contact data will be destroyed using a secure erasing software.

Raw survey data, uncleaned [privacy sensitive] – Raw survey data contains directly identifiable information, which means that until data cleaning has taken place it needs to be stored under the same strict access and security conditions as contact data. The files will be encrypted and stored on a password-protected distinct location separated from the contact information and the rest of the research data. Access will be granted to the trusted third party individuals responsible for pseudonymisation. During that process the directly identifiable information will be replaced by a meaningless unique personal identifier. The files created in this process contain raw *cleaned* survey data. The files containing raw *uncleaned* survey data will be destroyed after the cleaning process using a secure erasing software.

Raw survey data, cleaned [privacy sensitive] – Raw cleaned survey data does not contain any directly identifiable information anymore. However, indirectly identifiable and sensitive personal data are still included. To secure confidentiality accordingly, the data

files have to be protected. The according subfolders will therefore be password-protected and encrypted. As survey data is not reproducible the files will be write protected.

Raw administrative individual data, cleaned [privacy sensitive] – Cleaning of raw administrative individual data takes place at the municipality under the same conditions outlined above. After pseudonymisation the files only contain indirectly identifiable personal information. The municipality will exchange cleaned data files using the secure service SURFfilesender. Storage location will be password-protected subfolders of the project folder. All team members will have access to these files.

Processed survey and administrative individual data [privacy sensitive] – Processed data will be stored in password-protected subfolders of the project folder. All team members will have access to these files.

Remaining data – The remaining data is not privacy sensitive. It will be stored in the password-protected project folder. All team members will have access to these files.

2.2.5 Data access and security after the project

After the termination of the project research data will be anonymised and transferred to long term storage. All privacy sensitive files will be destroyed using a secure erasing software.

3 Preserve and Share

3.1 Data Preservation and Archiving

Preserving and archiving our data we follow the requirements of the Dutch *Code of Conduct for Scientific Research* (Nederlandse Gedragscode Wetenschapsbeoefening), the *Utrecht University Policy Framework* (Beleidskader Onderzoeksdata Universiteit Utrecht), and the *Protocol on Research Data of Utrecht University's Faculty of Social and Behavioural Sciences*.

Accordingly, our data will be stored for at least ten years after the research has been finished. Due to privacy regulations not all collected data can be stored for the long-term. To allow for data preservation we will destroy directly identifiable data and anonymise other research data once our data is static and not subject to any further changes. Our data will be made public at the latest nine months after a publication.

We will submit our research data to an archive in form of data packages. For every publication we will compile a separate data package with all the data and information needed for replication. In addition, we will compile a final data package will all data gathered during the research after the research has finished. Data packages will include:

- Metadata
- A text file describing the files included in the package and their relation
- Primary, raw data (anonymised)
- Secondary, processed data (anonymised)
- A code book explaining the variables and coding used
- A manual explaining our experimental design
- A survey guide explaining the process of data collection, including questionnaires
- A text file describing our statistical approach including programming code to replicate our analyses

To archive our data we choose for a trusted and certified data archive. Utrecht University is currently developing an own institutional archive, called I-Lab, which is planned to be available from December 2016 on. The data package shall be archived at there for 10 years. The data package will be documented using the metadata standard of the I-Lab.

To ensure data integrity and prevent the loss of data we choose to include data files into our package that are in future-proof long-lived formats. Those are comma-separated values (.csv) for spread sheet data and text files (.txt) and PDF files for textual data. In addition, we will use a verifier that allows us to check if our data is still the same. A copy of the data

package will be stored on the networked drive of the University, which means that it will be backed up automatically and securely on University servers.

Finally, our publications will be registered with Pure, the university's research data registration system.

3.2 Data Sharing and Reuse

We are planning to share our data in order to allow for further research and replication. Although we mostly collect personal data during the research we will be able to share our data as it will have undergone anonymisation before archiving. As stated in the previous section our data will be transferred to the institutional I-Lab archive after the research. I-Lab does not only function as an archive but also as a public repository which allows everyone to access the data. Using I-Lab our data will be provided with a Persistent Identifier (ePIC DOI) and disseminated via the I-Lab catalogue. As we make use of non-proprietary data formats no specific software will be needed to make use of our data.

To provide clarity and certainty for potential users of the data about how they are allowed to use it we will make use of a creative commons (CC) license. We choose the Creative Commons Attribution Only license (CC-BY 4.0), which allows potential users to distribute, remix, tweak, and build upon our work, even commercially, as long as we are credited for the original creation

Glossary

Please see the website of the Utrecht University Research Data Management Support for further explanations.

Code book: A file that comes with data files, explaining variable names and used codes.

Differential back-up: Back up all files that are changed or created since the first, full back-up of all files.

Documentation: Information about data, human readable.

External data: Data for which ownership lies completely outside of the university. Also 'third party data'.

Identifiable personal data: Data that without much effort leads to the identity of a person. This can be *directly* (i.e. by name, address) or *indirectly* (e.g. a rare occupation with age).

Incremental back-up: Backup files that are changed or created since the most recent (partial) back-up.

Informed consent: A voluntary, specific and unambiguous expression of will by a research subject, based on adequate information, to accept the processing of his/her personal data.

Master file: A file in which all changes are documented. It is the basis for backups (duplicates).

Metadata: Information about data, with a potential for machine-to-machine interoperability. Mostly a small set of vocabulary words used to describe a source.

Metadata schema: A list or description of metadata fields that needs to be filled out.

Metadata standard: Metadata schemas that are developed and actively maintained by responsible parties for reuse.

Non-disclosure agreement: A legally binding contract between two parties in a professional relationship to ensure confidentiality of sensitive information.

Personal data: All information collected with respect to a person.

Proprietary format: a file format for which encoding is either secret or published with its use restricted through licences.

Pseudonymisation: Identifying fields within a data record are replaced by one or more artificial identifiers, or pseudonyms. Either with or without the possibility of re-identifying the subject of the data (reversible or irreversible). It allows for data on the same subject to be linked across data records without revealing the identities.

Repository: A central storage to preserve, manage, and provide access to many types of digital materials, so these can be searched, discovered, and reused.

Sensitive personal data: All information about a person that is delicate such as race, ethnic origin, political opinion, physical or mental health, criminal record, sexual orientation, religious or other beliefs, economical status.

Versioning: The creation and management of multiple releases of a product, all with the same general function but improved or customized.

Relevant Laws, Guidelines and Policy Documents

1. Internal Documents

[to be added]

2. Laws and Regulations

- Wet Bescherming Persoonsgegevens (WBP)
- Gedragscode voor gebruik van persoonsgegevens in wetenschappelijk onderzoek
- Nederlandse Gedragscode Wetenschapsbeoefening
- Wet Meldplicht Datalekken

3. UU Policy Documents

- Informatiebeveiliging Beleid en Basisregels UU [UU Intranet]
- Beleidskader Onderzoeksdata UU
- Code Zorgvuldige en Integere Wetenschap UU
- Protocol Onderzoeksdata Faculteit Sociale Wetenschappen UU