# Practical Tips & Guidelines For Researchers Who Work With Sensitive Data

## DEFINITIONS

*Personal data*
All data that can be used to directly or indirectly distinguish an ordinary person. (WBP, section 1.a).

*Anonymizing data or de-identification of data*
A process that occurs throughout the data collection and analysis phases of research where identifying personal information is removed from the data in order to protect the privacy of research participants, the groups and/or communities that are being examined (Lindsay and Goldring 2010: 25).

*Secondary use (data re-use)*
Refers to using data to examine a question that was not the purpose of the original data collection.

## COLLECTING PERSONAL DATA FOR RESEARCH

According to the Dutch Personal Data Protection Act (*Wet Bescherming Persoonsgegevens - WBP*) a researcher can - for scientific purposes only - collect and analyze personal data. Personal information should be only collected when it is strictly necessary for research analysis and kept archived just for the period needed to complete the research only[1]. Researchers can for scientific and statistical purposes also collect personal data on religion, race, political engagement , union activities, criminal history and medical data.

Personal data can be collected in different ways (VSNU 2005: 12):

- through the respondent;
- using an existing database or dataset containing personal records;
- through family members or others that are familiar with the person;
- through the respondent about others only when this information is relevant to the research project.

Respondents should give explicit consent to the collection and/or analysis of personal data. Consent is not necessary when the respondent published his personal data publicly. Consent for use of personal data for research purposes isn't needed either when it's impossible to request explicit consent or when it would demand disproportionate effort from the researcher (WBP, section 23.2).

It is important to realize that leakage of personal information can damage the person(s) involved. Scientists should protect the personal data of their informants throughout the whole research process. The protection of data and subjects' privacy must be carefully planned and may not be jeopardized by careless storage of data or unprotected electronic data transfers.
The researcher is responsible for the privacy of the respondents during the whole research project and should take all measures necessary to guarantee that all information is stored securely. Also, the researcher(s) should be the only one(s) that can access this information. This information is confidential and cannot be publicly published or disseminated without the explicit permission of the respondent.

[1] Accordingly to the Dutch Personal Data Protection Act, section 10.2, personal information can for scientific, historical or statistical purposes be stored for a longer time than the research period.

## STORING PRIVACY SENSITIVE DATA ON YOUR COMPUTER

It is important to take all measures to protect the privacy of the respondents. During the research process. The Vrije University Amsterdam takes several measures to protect the privacy of the respondents during the research process. The VU provides secure internal network systems. VU computers and notebooks with the green label provide guaranteed protection to the data. The data stored at the **home (H:) or group (G:) drives** are also protected. Researchers may not leave computers unprotected when going for lunch or to meetings. It is important to lock the computer with a password before leaving the room. *The storage of personal data on* ***private*** *notebooks, mobile devices or external hard disks is truly discouraged and should be prevented.*

Researchers can also use the freely available **SURF-Drive** option to store (sensitive) research data during the research project. A folder in SURF-Drive can be shared with anyone you choose. Make sure to delete these users' access rights when they no longer need the data. *Please be aware that the Surf-Drive automatically synchronizes the folders to your workstation.* Therefore, make sure your workstation's virus scanner is up-to-date and use encryption on the hard disk.

**Encryption** is the conversion of data into a form that cannot be easily deciphered by unauthorized people. Encryption can be done with special software. Regular Excel and Word files can also be encrypted using a password: see the instruction video of the Oxford University.
The VU also offers the software TrueCrypt for this purpose. Installation has to be done by the IT Servicedesk, through sending an email to Servicedesk.it@vu.nl with your pc number, VU-Net-ID and the name of the software that you want to install.
VU Windows 7 laptops have the encryption software called **BitLocker**. You can check whether this encryption software is active using the **Control Panel** at **System and Security**.
VU Macbooks use the encryption software **FileVault**. This software option has to be activated by yourself: instructions are available here.

Sending a confidential document by email can be done using the Surffilesender tool. With your VU-Net-ID you can log into this tool remotely. Files of up to 250 MB can be sent securely. Instructions are available here.

The Vrije University Amsterdam also provides advice on the use of technical infrastructure to store and transfer protected data. For More information, please contact the information security office: soc@vu.nl.

During the research project a scientist can also decide to anonymize the research data and keep the personal information in a separate document that can be used for contact purposes with the respondent. By keeping the personal records separate from the research data, the scientist can more easily transfer the non-personal data through devices and network systems. The personal information can stay stored securely in the university's network system. By anonymizing the data the scientist can prevent biases as well – knowing the research subjects can have an effect on the researcher's conclusions.

## ARCHIVING PERSONAL DATA

Personal data may only be collected and saved when strictly necessary and for research purposes only. Organizations that stored personal data may ask permission to the Dutch Data Protection Authority (DPA). Accordingly to DPA scientists don't need to ask for permission when (*Vrijstelling besluit CBP, paragraaf 7, vrijstelling 28*[2]):

***1. Authorized purposes of personal data processing***
*The data processing may only take place for collecting, processing and controlling of the information for a specific investigation or statistical purpose.*

***2. Permitted categories of data to be processed:***
*Only the following personal data may be processed:*

- *Family name, first names , initials , titles, gender , date of birth , address, postal code, phone number and similar information necessary for communication (eg. e mail), like the bank account of the person concerned ;*
- *an administration number ;*
- *every information other than listed above on behalf of a specific investigation or a specific statistical purpose.*

***3. Permitted categories receivers of the data:***
*The personal data may only be provided to or used by:*

*Persons including third parties, that :*
*- Are charged with the activities listed under section 1, or*
*- Are coordinating the persons that are charged with the activities listed under section 1, or*
*- Are necessarily involved with the activities listed under section 1.*
*Others, when:*
*- The person involved has provided unambiguous consent for the data processing, or*
*- The data processing is necessary for compliance with a legal obligation by the organisation for scientific research or statistics , or*
*- The data processing is necessary for the vital interests of the data subject (eg an urgent medical need ), or*
*- The data are further processed for historical, statistical or scientific purposes . Condition is that the organization ensures that the data will only be processed for these specific purposes.*

***4. Retention period***
*All personal data included in the first point under section 2 with the exception of gender , place of residence and date of birth* ***must be removed*** *not later than six months after the information under the third point under section 2 is obtained .*

When scientists for research, historical or statistic purposes need to archive personal data like name, first names, initials, titles, gender, date of birth, address, postal code, phone number and similar information necessary for communication (eg. e mail), and bank account for a period longer than six months, they may report the data collection to the Data Protection Authority. The online form is available in Dutch language only. More information about reporting data, can be asked by DPA - 070-8888500 (workdays between 9:30-12:30 only).
The university has two legal experts on privacy: mr. Petra M. Tolen and mr. Tom H.G. Paffen. Both can help and provide advice on privacy issues to researchers. It is also possible to request a **Privacy Impact Assessment** to determine what the privacy risks are during your research project.

[2] Exemption 28, Archives and Research (Dutch only): http://www.cbpweb.nl/hvb/pages/vwc28.aspx

When research is finished, the scientist should remove the personal information of the data and keep the non-personal information archived in an academic data repository for re-use. Personal data may not be made public or published.

## ANONYMIZING DATA

Anonymizing data may involve several levels and there are many ways it can be achieved.

**Removing personal data for communication purposes**
Personal data are often collected for communication purposes and are not directly part of the research analysis. This information can be removed before data analysis. Reducing this information from the dataset during the analysis has many positive points:

- the researcher can easily share the dataset with other researchers without having to take strictly security measures for privacy data sharing;
- the researcher can store the data in computers outside the university network system without transgressing the security rules of data protection;
- with anonymization the researcher avoids data bias in the evaluation.

Personal data collected for communication purposes are often:

- Name
- Address (also electronic address)
- Phone number

These data may be copied to another document and kept separately in the university computer or university network system. Each person receives a code number that will be used in both documents so that the person can be identified again when needed.

**Bracketing indirect identifiers**
Removal of name and address is not enough to secure the privacy of respondents. The combination of different identifiers can lead to a person and damage his reputation or causes other personal damages. It is important to the security of the respondents that the information doesn't point direct to a person.
An approach that can be used in these cases is called bracketing – the use of categories for generalization of the information. Examples of identifiers that can be changed in categories:

- Birthday or age of the respondent may be replaced by age categories as: 1-10 year old; 11-20; 21-30; etc.
- Income
- When residence is important for the research, the scientist may prefer to use the name of the district instead of the address when the research is taking place in a city only. When the research is taking place in a whole country, the scientist may prefer to use the name of the city, and so on.

For each research the scientist writes a privacy plan for anonymizing the data suiting his research purposes and securing the privacy of the respondents.  This procedure should take place during the research. These personal identifiers should never be publicly accessible without anonymization. Researchers should also ensure that a copy of the original data be kept in a secured storage during the research.

## WORKING WITH SENSITIVE GROUPS

When working with sensitive or marginalized groups, researchers need to make certain that this information is not publicly accessible. Even when this information are stored for research purposes only, the scientist should realize that harmless information today could be used against a group later.  The next categories of information should never be traceable to a person without the consent of the respondent:
- religious belief;
- sexual preferences;
- health matters;
- political activities;
- criminal activities

## FURTHER READINGS

Gedragscode voor het gebruik van persoonsgegevens in wetenschappelijk onderzoek (Dutch only) VSNU, 2005 (volgens CPB is deze gedragscode in januari 2011 verlopen)
http://www.vsnu.nl/files/documenten/Domeinen/Accountability/Codes/Gedragscode%20persoonsgegevens.pdf

Handleiding voor verwerkers van persoonsgegevens (Dutch only), Ministerie van Justitie, 2006
http://www.rijksoverheid.nl/bestanden/documenten-en-publicaties/brochures/2006/07/13/handleiding-wet-bescherming-persoonsgegevens/handleiding-wet-bescherming-persoonsgegevens.pdf

The Netherlands Code of Conduct for Scientific Practice, 2012
http://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/The_Netherlands_Code_of_Conduct_for_Scientific_Practice_2012.pdf

Personal Data Protection Act (unofficial translation from the Data Protection Authority)
http://www.dutchdpa.nl/Pages/en_wetten_wbp.aspx

DANS (Data Archiving and Networked Services) privacy regulation, 2009
http://www.dans.knaw.nl/en/content/dans-privacy-regulations

### Useful Links Protect Data Autority (Dutch only)

Handreiking Vrijstellingsbesluit Wet Bescherming Persoonsgegevens
http://www.cbpweb.nl/hvb/pages/i1.aspx

Melden Verwerking Persoonsgegevens (Dutch only) – College Bescherming Persoonsgegevens
http://www.cbpweb.nl/Pages/ind_melden.aspx