**SOP: Safe Practices for Students Working with Research Data**

Most of the departments within the FGB take on students for research internships or as temporary research assistants. When students work with personal data, there may be a risk to the confidentiality of the data if students are not taught how to correctly manage and protect this data. Applying the following principles will help students safely manage and work with personal data.

- All students that are working with confidential information (personal data, patents, data intended for publication) must always sign a confidentiality agreement before starting any research. For students conducting research solely under the supervision of the VU, a confidentiality agreement can be obtained from your section or department head, who will also sign it on behalf of the FGB director of operations. If the student will be working with other parties and collecting data under their supervision (such as in a hospital) a different agreement will need to be set up. IXA can assist with those kinds of agreements.
- All students must read the "SOP: Safe Transfer of Data Outside of the VU"

The supervisor should then determine the privacy risk of the research dataset that the student will use. Appropriate measures to protect the data will be determined based upon this risk level. One method that can help determine risks (in terms of privacy, integrity and security) to data is to complete a data classification (or review a data classification that has already been completed for the research project). The following information provided below also provides examples of data types and the associated privacy risks. Contact the FGB research data officer if uncertain about the risks and for a copy of a data classification template, if desired.

- *Extremely High-risk data:* raw audiovisual recordings; raw interview transcripts; data consisting of directly identifying information (name, address, telephone number, e-mail address etc.)
    - Students must work with such data within the VU on VU workstations. They must save all of their work in a folder on a VU network drive. Access to an appropriate drive is only possible via a functional account and under supervision of the student's supervisor. Contact UCIT for support.
- *High-risk data*: pseudonymized data containing birthdates, 6-digit postal codes, highly specific demographic information (e.g. unique job titles), longitudinal data with several specific event dates and/or open text fields in questionnaire data; blurred video data; voice modulated audio data
    - Whenever possible, if a student is working with high-risk data they should work on VU workstations under supervision as required for extremely high-risk data
    - If this is absolutely not feasible, a laptop can be loaned from TO3 and all data files must either be encrypted with VeraCrypt or all processing of the files (collection, analysis and storage) must be done on an encrypted USB- or external hard drive rather than the laptop hard drive
    - Data should occasionally be transferred to the supervisor for back-up on a VU network. Data may only be transferred to and from TO3 laptops via a secure USB stick or an appropriate data transfer program (see SOP: Safe Transfer of Data Outside of the VU). Data must ***NEVER*** be sent to or from personal e-mail servers, Dropbox, Google Drive or WeTransfer
    - Students must not link TO3 laptops to public Wi-Fi nor use personal e-mail on TO3 laptops
    - Students must not work on high-risk research data in public spaces (working on the VU campus is an exception, but they should find appropriate spaces to do so, e.g. not the cafeteria)
    - Upon completion of the project, all data must be transferred to the appropriate VU network drive and all files must be erased from the laptop and any portable devices used. Software such as KillDisk or Eraser can be used to permanently remove data from a hard drive or USB.

- *Medium-risk data*: pseudonymized data that do not meet the requirements of high-risk data, i.e. there are no highly specific unique cases present in the data. Medium-risk and low-risk data are the most appropriate data for students to use
    - **The same rules apply to medium-risk data as those applied to high-risk data**. However, with medium risk data it is allowed for students to use their own laptops if there are no other options available
    - Prior to sharing data with the student, **if there are any records where indirectly identifying information (e.g. age, gender, region, occupation, other sociodemographic information), when combined create a unique record (N=1) or a record that is the same as only one other record, all such records must be deleted from the student dataset.**
    - Full disk encryption must be activated on the student's laptop and the student must permanently erase all research data from his or her hard drive upon completion of the project, as well as from any portable devices used. Software such as KillDisk or Eraser can be used to permanently remove data from a hard drive or USB-drive.
    - If a personal laptop is being used, the ID codes of individual participants must be deleted from the dataset
    - Data should occasionally be transferred to the supervisor for back-up on a VU network drive in case of disaster (NOTE: any data transfers must be completed as described above)
- *Low-risk data/anonymous data*: low-risk data is essentially impossible to trace back to individuals. Assessing whether the risk is low enough can be extremely time-consuming, therefore for student purposes, data may only be considered low-risk if there are no direct or indirectly identifying variables present in the dataset. Bear in mind that even non-demographic information can be identifying, so even if there is insufficient time to complete a risk assessment of re-identification in the data, take the time to consider whether the variables present in the dataset could be used in some way to re-identify individuals. To determine if the variable is an identifier consider:
    - Is it replicable: your birthdate is consistent over time, whereas your blood sugar varies constantly
    - Is it distinguishable: does the variable single out an individual in the group
    - It is knowable: is the data represented by the variable publicly available, or likely to be known by unauthorized individuals, such as a subject's health care provider? This can be hard to know for sure; when in doubt assume it is known.

  Also note that even if the data is anonymous, it may still contain sensitive business information or intellectual property information. If this is a concern, but the data is otherwise anonymous, treat the data as medium-risk.
    - Low-risk data can be used on any of the above workspace options. No extra measures are necessary to protect the privacy of the data. Students are still strongly advised to back-up their work occasionally on VU network drives in case of disaster.

Finally, regardless of the risk level of the data

- If data must be collected remotely, the student must follow the recommendations in the SOP: Safe Data Use Outside of the VU.
- Students must document all of their work, particularly any code they used to process and analyse the data.
    - Documentation is the backbone of research data management and if students document their work, their supervisors can more effectively review it and, where appropriate, use it for publications in the future.