**Research Data Management and Security Plan**

**Guidance**

*The objective of the data management questions is to assist researchers with accurate descriptions that are easily understood by reviewers and information security professionals. If you are unsure how to answer a question, feel free to reach out to your normal support staff at UVM, UVM LCOM or UVMMC for assistance.*

**Summary:** The goal for data is for it to be stored in a safe manner, backed up, and available to key personnel in ways that ensure the confidentiality, integrity and accessibility of the data.

The recommendation for data is to code the data by removing any/all of the 18 HIPAA identifiers and replacing with a random (not formulaic) code. You can keep the electronic key in a separate file location should you need to re-identify the data in the future. The electronic key should be password protected and encrypted. Once the data is sufficiently de-identified, the security surrounding the data protections are loosened. Basic physical and computing security should still be in place to protect the intellectual property of your data and to avoid the possibility of data theft by people who could try to re-identify the data manually (remember, we live in a small community.)

If you cannot code the data, you should talk to support personnel at your institution to identify the best way to secure the data. This process can also assist in identifying that your administrative procedures are in place to ensure: annual training of staff in data security; timely adding and removing of permissions; the rule of least privilege as it pertains to what data an individual can see is followed; and that you are storing the data file(s) in a manner that meets your, your sponsor’s (if applicable) and the institution’s security expectations.

**Tips**: It is understood that full identifiers are likely collected during time of treatment and that this information is usually kept in the clinical setting. These questions are specifically asking about the information that you are, in effect, *copying* from the clinical setting or are collecting in a *separate* fashion from the clinical EMR to use in your research data sets.

It will help your efforts if you run the data collection/protection for all studies as similarly as possible. Creating one set of Standard Operating Procedures (SOPs) that can be used as templates for all studies will help you in creating easy to follow, easy to understand, and repeatable processes. Having the SOPs in a location that is accessible by all study personnel will allow for smoother transition when personnel come and go. They should cover topics that are important to your lab, studies and protocols but some suggested topics are: data collection (how it happens, where it goes), data protections (where is it stored, how is it stored, how is it accessed, who can access it), personnel roles/permissions/duties, training content applicable to specific protocol, training logs (who was trained on what and when), and data retention/disposal to specific protocol.

**Reminder** - The 18 HIPAA identifiers are as follows:

* Name
* Address (all geographic subdivisions smaller than state, including street address, city county, and zip code)
* All elements (except years) of dates related to an individual (including birthdate, admission date, discharge date, date of death, and exact age if over 89)
* Telephone numbers
* Fax number
* Email address
* Social Security Number
* Medical record number
* Health plan beneficiary number
* Account number
* Certificate or license number
* Any vehicle or other device serial number
* Web URL
* Internet Protocol (IP) Address
* Finger or voice print
* Photographic image - Photographic images are not limited to images of the face.
* Any other characteristic that could uniquely identify the individual

**Guidance for Select Questions on the Research Data Management and Security Plan form**

**Plan for Protecting the Stored Data**

2.a. Consent forms always include direct identifiers, the participant’s name, we are interested in knowing how the associated paper research data files are identified. Are direct identifiers included on each data point or do you replace that information with a code. Coded means that you are separating the identifiers from the research data, replacing them with a *random* code and keeping the identifiers separate from the research data intend to use a code, explain how the process by which the data are coded. You are requested to explain this process.

2. These questions are to document where electronic data will be stored and how it will be protected. Except for regularly scheduled backups, it is recommended *not* to have multiple copies of your research data and files. It is also recommended that, if you have a key file, you place it in a central location and limit who has access to it. It is strongly discouraged to keep identifiable or sensitive data on portable media including laptops or other smart devices. If identifiable information needs to be transferred, it should be noted why and a procedure must be established for it to be properly protected with encryption.

* If you are a **UVMMC user**, it is likely you are keeping your data on the S: or Shared drive at the hospital. You likely will want to check the UVMMC box and be as detailed as possible where the data is kept (folder names or other details that you can remember.)
* If you are a **COMIS user**, it is likely that you are keeping your data on the L: or Departmental shared drive at the College of Medicine. You will likely want to check the UVM COM box and be as detailed as possible where the data is kept (folder names or other details that you can remember.)
* If you are a **UVM user**, you will need to note as many details as possible where you are keeping your data. It’s possible you are storing in your home space, your shared space, your zoo space or another location that is unfamiliar to the reviewers.

2.b.iii.d. If a third party vendor has already identified, include that information in this section. Request that the vendor/application be reviewed by IT to ensure that the collection and methods meet the privacy and security measures of the institution. Additionally, if you are planning on having something specifically programmed for this study, be sure that you are including the privacy and security personnel early to ensure that those measures are built into the final product.

3.a. Any data use agreements for the Larner College of Medicine (LCOM) will be signed by Gordon Jensen, Senior Associate Dean for Research, and passed through the College of Medicine Technology Services (COMTS) office for review before signature and many of these same questions will be asked.