OFFICE OF THE INSTITUTIONAL REVIEW BOARD

Human Research Data Security Standards
UNM Main and Branch Campuses

Introduction

It is the policy of The University of New Mexico (UNM) Institutional Review Board (IRB) for Main and Branch Campuses to consider whether adequate provisions exist for the security of research data. Researchers are entrusted with confidential and privileged human subject information, whether in paper or electronic form and must take measures to protect the security of this information.

Researchers should pay special attention to data security because of the use of electronic devices including portable devices and drives, as well as web-based survey tools. Given the wide range of diversity in studies, methods, and electronic data devices, researchers need to give extra consideration to confidentiality and data security when electronic data are collected and/or stored.

Each member of the campus community is responsible for the security and protection of information resources over which they control. All researchers and research team members must be familiar with information security policies and procedures of their department or unit, UNM, the State of New Mexico and Federal privacy laws (such as HIPAA, FERPA, FOIA, New Mexico IPRA) as well as the data confidentiality requirements associated with research funding (e.g. National Institutes of Health, Department of Justice (DOJ), etc.

Definitions:

- **Data:** For the purpose of this policy, data refers to information gathered from or about individuals during the course of scientific research.
- **Subject identifiers:** For the purpose of this policy, this includes information that identifies a person participating in research including any or all of the following: (1) names; (2) social security numbers; (3) birth dates; (4) addresses; (5) IP addresses; (6) other data that could reasonably lead to discovering a personal identity (e.g. direct and indirect identifiers).
- **Personal health information (PHI):** This is defined by HIPAA law and includes personal identifiers that are associated with medical information other than patient/subject self-reported information that may pertain to health. Information/data from medical records are considered PHI.
- **Link:** A document that contains a crosswalk connecting subject identifiers to unique study identification number or code.
- **Encryption:** The process of encoding information in such a way that only the person (or computer) with the **key** can decode it.
- **PC/Personal computer:** A stand alone or networked computer such as a desktop device.
- **Portable devices:** A portable computer that includes traditional laptops, netbooks, tablets, smart phones and other portable computing devices that generally have full range PC capacities. Small plug-and-play devices such as USB drives, external hard drives and webcams can also be considered portable devices.
- **Sensitive data:** Data for which any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.
- **Storage:** Secure storage can include both on-site and off-site facilities.

Guidelines to Consider when Submitting a Protocol

The IRB Protocol requires researchers to address issues related to subject privacy and confidentiality, information security, data management and HIPAA, if applicable. The following standards should be considered and incorporated into the IRB protocol:

1. Collect only the minimum necessary subject identifiers.
2. Remove/destroy subject identifiers as soon as they are no longer needed. See below for record retention requirements.
3. Limit physical access to any area or computer or device that contains subject identifiers.
4. Limit electronic access to any computer that contains subject identifiers.
5. Avoid storing subject identifiers or PHI on portable devices (such as laptop computers, digital cameras, portable hard drives including flash drives, USB memory sticks, iPods or similar storage devices) as these devices are particularly susceptible to loss or theft. If there is a necessity to use portable devices for initial collection of sensitive data with subject identifiers, the data files or devices must be **encrypted** (more information on encryption provided below), and data must be transferred to a secure system (e.g. server behind firewall) as soon as possible. Then, the data must be deleted from the portable device.
6. Remove subject identifiers from data files whenever possible. Identifiers should be stored in a separate and secure location from the coded data files, and associated with the data files through a link that is also stored in a separate and secure location. Coded (de-identified) data files stored electronically must, at a minimum, be password protected.
7. If subject identifiers will be retained in the data files because of the specific needs of the study, additional justification must be provided by the researcher in the IRB protocol to justify retention. **If identifiable sensitive data are stored electronically the files or device must be encrypted.**
8. Use only secure modes of transmission of data; identifiable data or PHI submitted over a public network or email must be encrypted (note: UNM main/branch campus email is NOT encrypted).
9. Never store identifiable human research data on public online services or cloud storage (including Google Drive, Dropbox, Basecamp, etc.). UNM IT and the University Library provide alternative storage solutions.
10. If storing data off of the UNM campus, the electronic storage devices must, at a minimum, be password protected. If data is sensitive and identifiable or is PHI, it must also be encrypted, as described previously. Hard copy data must also be kept secure (locked file in a locked home office). Hard drive or mobile devices should also be stored in this manner until such time that the data can be removed from the portable device.
11. Review the UNM IT website for additional recommendations on how to best secure confidential data.
12. In the Confidentiality section of the IRB protocol, researchers must address the method of collecting, recording, coding, and maintaining data, as well as specify who will have access to the data and at what point subject identifiable data will be de-identified or destroyed.
13. In the Informed Consent document, researchers must describe the extent, if any, to which they will maintain confidentiality of records identifying the subject.
14. The PI must report any inadvertent breach of confidentiality of the research data which causes harm or places subjects or others at a greater risk of harm (including physical, psychological, economic, or social harm) to the IRB within 7 calendar days of the researcher becoming aware of the event (see OIRB SOP 401).

Web Surveys

There is increasing interest in using web-based survey tools for research involving human subjects. There are two major forms of web-based surveys:

1. researcher-devised and programmed tools that are housed on university servers under the control of the researcher; and
2. independent proprietary survey programs that incorporate researchers' own measures but data reside on servers owned by the survey company.

Researchers need to have clear assurances about the protections that are afforded by the independent proprietary providers, whereas those that are developed by the researcher can have total control in-house. While proprietary vendors of web-based survey tools are generally ethical, researchers should obtain information about the tool's security and privacy protections, including learning whether user IP addresses are captured and saved during completion of the surveys. Most vendors will remove IP addresses from data at the researcher's request. Some vendors also remove them upon submission of the completed form, but this should be clarified. Despite their stated privacy policies, many vendors, especially those who promote freeware, do in fact share IP addresses with their consortium of investors, and thus absolute anonymity cannot be guaranteed to survey respondents.

With either approach, the window of greatest vulnerability for data is during the time the program is open and being used by the participant. This is the point at which hacking could discover identity or other personal information. Researchers need to be assured that any data collected in web-based tools that contain subject identifiers or PHI are being encrypted before and after transmission.

Several things need to be considered with web-based data collection. The informed consent forms or scripts used in lieu of consent forms will need to clarify (in simple terms) the kinds of protections that are available to the web-using participant, if applicable. Just as consent forms describe research materials being in locked file cabinets when the data are in hard copy, the consent form should describe the specific web-based data security being used. If proprietary vendors are being used to collect data, and if breach of confidentiality could put respondents at risk due to the nature of the survey questions, consent forms should explain this possibility to potential respondents.

Online statistical software would be treated similarly, unless working with completely de-identified data.

Encryption

Encryption can be applied to storage devices or files (data "at rest") and to network data (data "in flight"). The type of computing device, the network communicating from/to the device, and whether identifiable sensitive data or PHI is involved will dictate whether or not encryption is required. See the UNM IT website for recommended encryption options. Open source file level encryption applications can also be obtained at: https://www.gnupg.org/.

Encryption is not needed if you do not store or work with research data that includes subject identifiers or PHI. Therefore, it is best not to collect any of this information unless it is actually needed.

Scenarios in which storage encryption is **REQUIRED**:

- Possession of sensitive data that includes subject identifiers and/or PHI, AND
- Computing device is a mobile device, OR

- Computing device is a personal system, OR
- Storage device is removable (portable), OR
- Access to the storage device is not in a physically secure environment, OR
- Data storage is on shared computers (i.e. servers) to which multiple users have access over a network.

Scenarios in which file level encryption is **REQUIRED**:

- Use of sensitive data that includes subject identifiers and/or PHI over a network if:
    - The information is not already encrypted by means of storage encryption, AND
    - Any part of the data transmission is outside of a trusted (UNM-managed) network, OR
- Access to a system containing research data that is identifiable and/or includes PHI that is not entirely over a trusted network.

Examples of common tasks where encryption is **REQUIRED**:

- Use of electronic research data that includes subject identifiers and/or PHI, AND
- The information is being sent by:
    - Email, OR
    - Webmail, OR
    - Web browser, OR
    - Traditional mail (US Post Office), OR
    - Courier, OR
    - Instant Messenger, OR
    - Peer-To-Peer network, OR
    - Wireless (Wi-Fi, cell phone, Blackberry, SMS, etc.), OR
- A backup of the information is created.

DATA LOSS OR SECURITY BREACH

Store data access information securely where the Principal Investigator and research team can find it. If data is lost or there is a breach of confidentiality of identifiable research data, the event must be reported to the IRB within 7 days of discovery. There may be additional institutional and sponsor reporting requirements, depending on the type of data (e.g. PHI). To initiate sponsor-related reporting, please contact the UNM Office of Sponsored Projects.

RECORDS RETENTION REQUIREMENTS

Researchers are advised to retain all study records for 3 years after closure of the project. This includes approved IRB documents, as well as recordings, tapes or transcripts (unless destroyed earlier according to approved protocol), and all other data-collection instruments and source documents. Confidential data must be stored in such a way to prevent breach or loss. Data may also be retained for copyright and intellectual property applications as well.

Longer retention periods are recommended for certain research records:

- Records involving the generation, disclosure, and/or use of Protected Health Information (PHI) should be retained for six years.
- Records for funded research must be retained in compliance with Sponsor requirements.

- The National Institutes of Health and National Science Foundation require that grant recipients keep all data three years beyond grant final expenditure report.
- American Psychological Association expects its members to retain data for a minimum of five years.

In sum, the recommended approach is to determine which regulation applies to your research. Additional examples of regulations that might apply to your research include Office for Human Research Protections (OHRP), HIPAA, Food and Drug Administration (FDA), Department of Veterans Affairs (VA), DOJ, etc. Different regulations have different timelines. For example, current OHRP regulations mandate that data be kept for at least 3 years, current HIPAA rules call for at least 6 years, while the Department of Veterans Affairs' norms impose the requirement that data be retained indefinitely. If multiple regulations apply, the researcher should keep the data for the longest required amount of time.

In addition to the regulations, if your study is under a sponsored project, you must comply with any terms for confidentiality and record retention detailed in the award from the sponsor or associated contracts such as MTAs, CTAs, NDAs, etc.

You should also determine what information you should keep. Generally, you should keep the following: signed participant informed consent/assent documents, signed parental/guardian informed consent documents, IRB correspondence and written research summary. In most cases, these records can be stored electronically or in hard copy. Again, look to the specific language of the applicable regulation, policy or contract for guidance.