Confidentiality and Data Security Guidelines for Electronic Research Data

North Dakota State University (NDSU) Institutional Review Board (IRB)

Research practices routinely involve electronic data in a variety of ways. Traditionally provisions for the confidential handling of research data referenced keeping data in locked filing cabinets and in locked offices. While these provisions still have application when research data or materials are in hard-copy, there is a growing complexity in how research data must be protected while it is collected and stored electronically. These guidelines attempt to outline basic protection provisions expected of investigators to protect data during data collection, transmission and storage, realizing that advances in device design, software, and university systems are constantly changing.

To what do these guidelines apply?

This guidelines applies to all studies involving individually identifiable participant data that include information of a personal or health nature. This can include even low to minimal risk studies **if** the information is personal or health related.Thus, even individually identifiable minimal risk research surveys on smoking and/or drug and alcohol use would be included in these guidelines.

Studies which would be excluded from the guidelines include surveys which collect no direct or indirect personal identifiers, or if identifiable, studies which are non-personal in nature, such as participation in hobbies, special or political interests, etc.

Terms:

In this guideline two terms are used for personal information:

1. **Personal health information (PHI)** consistent with HIPAA concepts; and
2. **Personal identifying information (PII)** for other studies not covered by HIPAA.

Data security guidance for investigators who collect, use and store electronic data:

1. **Secure Servers/Desktop computers:** The recommended electronic devices for entering and storing human subjects data are secure servers or desktop computers that have encryption software for all PHI or other identifying data.
   1. Operating systems remain up-to-date with updates and security patches.
   2. Stand-alone PCs can be used for data storage of de-identified data without encryption; however, password protection should be enabled over the use of the desktop computer.
   3. Server based PHI or other identifiable human subjects data should be firewalled and be encrypted.
   4. Internal or external servers housing data include:
      1. Account Control Plan:
         1. Strong passwords/pass-phrases are used and their use enforced.
         2. Accounts on the server are unique and those that are not needed are disabled or removed.
         3. Access to data is on a need to know basis.
      2. Patching Plan:
         1. Operating System Patches are installed in a timely fashion and given a priority.
         2. This plan also includes the application that the server is going to be using and any other 3rd party applications.
      3. Access Control:
         1. All servers have some network access controls enabled, capable of limiting network and Internet access to the server.
         2. The server is in a secured location with limited physical access.
         3. When possible, the applications and services will work in a non-administrative mode.
      4. Malware Control:
         1. Operating Systems that are historically susceptible to malware attacks have protection installed, enabled, and be able to be updated.
      5. Logging:
         1. Operating System level and Application level events are be logged to assist in troubleshooting and forensic investigations.
      6. Backup Plan:
         1. A plan is in place for the backup/recovery of data.
         2. Data backups should be stored in an off-site secure location.
   5. Anonymous data or de-identified data that cannot be traced back to an individual, using cue information in the data set matched to other data sources, can be stored on servers without encryption, but would still require authorized password access.
2. **Laptop data collection devices**. Laptops can be approved for data collection of human subject data when the following are provided:
   1. When available through the University System, two factor authentication is used for storage or access of protected health information (PHI).
   2. If the device stores personal health information nor other identifying information, the device hard drive is encrypted.
   3. The device uses software that encrypts all personal health information or other identifying information.
   4. The data are formatted such that PHI or other identifying data are in separate files or tables from any clinical or research information about the person.
   5. Each user of the device maintains a unique log in and password.
   6. All files containing PII or PHI are password protected.
   7. Laptops can be used for anonymous data collection without encryption.
   8. Laptops can be used for storing and analyzing de-identified data on human subjects.
3. **Jump drives** are only to be used under the following conditions:
   1. The jump drive uses files that have software to automatically encrypt all PHI or other identifying information or the entire jump drive is encrypted.
   2. The data are formatted such that PHI or other identifying data are in separate files or tables from any clinical or research information about the persons.
   3. All files are password protected.
   4. Jump drives can be used for storing and analyzing de-identified data on human subjects.
   5. Jump drives should be stored/transported in a secure location.
4. **Web-based data entry/surveys**.
   1. Research data collected through online portals and/or survey tools should be use secure web server (https) protocols and the server should encrypt any PHI or other identifiers upon submission. The server should be firewalled.
   2. Web-based anonymous or de-identified data need not be encrypted. Firewall protections are advised, but not essential.
5. **Smartphones, tablets, PDAs, and other mobile devices**.
   1. When using mobile devices to collect research data:
      1. Configure the mobile device to be secure by enabling auto-lock and requiring a password or passcode.
      2. Set device to erase data after 10 failed passcode attempts.
   2. When using mobile devices to access research data sets:
      1. The device uses software that encrypts all personal health information or other identifying information.
      2. The data are formatted such that PHI or other identifying data are in separate files or tables from any clinical or research information about the persons.
   3. General best practices:
      1. Connect to secure Wi-Fi networks and disable Wi-Fi when not in use.
      2. Back up your device on a regular basis.
      3. Update mobile devices frequently. Select an automatic update option, if available.
      4. Use appropriate sanitization and disposal procedures for mobile devices. Delete all information stored in a device prior to discarding, exchanging, or donating it.
6. **CDs and DVDs**
   1. PHI, PII or other identifying data should not be stored on CDs or DVDs unless the entire CD or DVD is encrypted.
   2. De-identified human subject data can be stored on CDS or DVDs in open format.
7. **Email**: PHI or PII or other identifying data should not be contained in email communications.

Web-based surveys:

Individual proprietary survey programs incorporate investigators’ own measures but store data on their own servers outside of NDSU. Investigators need to have a clear understanding about the protections that are afforded by the independent proprietary provider. Investigators should obtain information about the tools’ security and privacy protections, including learning whether user IP addresses are captured and saved during completion of the surveys. Most vendors will ‘scrub’ IP addresses from the data at the investigator’s request, or upon submission of the completed survey, but this should be clarified. Despite 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.

[Qualtrics](https://www.ndsu.edu/gdc/?page_id=546), an online survey platform, is available for use by NDSU faculty, staff, and students at no charge. Investigators should be aware that other survey platforms may involve individual software licensing agreements which are subject to review and approval by the [NDSU Chief Information Officer](https://www.ndsu.edu/its/security/).

The window of greatest vulnerability for data is during the time the survey program is open and being used by the research participant. This is the point at which hacking could discover the identity or other personal information. This is not unique to web-based research, but includes any period when a user is online. Investigators need to be assured that when any PHI or PII are being collected in web-based tools, that once the data are transmitted, they are encrypted.

Informed consents should clarify the kinds of protections that are available to the web-using participant. Just as consent forms describe research materials being locked in filing cabinets when data are in hard copy, the consent should describe the specific web-based data security being used. If proprietary vendors are being used to collect the data, and if breach of confidentiality could put respondents at risk due to the nature of the survey questions, consent forms should describe this possibility.

|  |  |  |  |
| --- | --- | --- | --- |
| TOPIC | YES | NO | GUIDANCE/RECOMMENDATION |
| data collection |  |  |  |
| Will your study use personally identifiable information (PII) or personal health information (PHI) about participants? |  |  | If yes, then generally the IRB requires that research data be kept apart from PII or PHI. Separate tables or separate files should be used to maintain confidentiality of individual records. *Note this protection in the consent form. See examples of table structures in Exhibit A below.* |
| Will the PII or PHI be obtained from existing electronic data systems (electronic medical records or institutional data sets)? |  |  | If yes, then ensure that applicable HIPAA, FERPA, or other authorizations have been approved for data access into a new electronic data table.  **On receiving PII or PHI, separate these identifiers from other health or research data. Create crosswalk, if needed, between participant record number and PII/PHI and ensure that the cross walk file is separate from both the research data and the PII/PHI file. (See example of table structures in Exhibit A below).** |
| Will data from participants be entered directly into electronic devices during research surveys or procedures? |  |  | If yes, ensure encryption is installed such that PII/PHI are always encrypted as they are entered, saved or submitted.  The ideal is for the encryption to occur during entry, but it is acceptable to do this at ‘save’ or ‘submit’ functions. *Note this in the consent form.*  An example of language for this might be: “All personal identifying information is encrypted as it is typed into the laptop.” Or “All Personal identifiers are encrypted when data are uploaded.” |
| Will data be loaded into a server system using a virtual private network (VPN)? |  |  | If yes, the VPN-accessing server must be behind a firewall and all identifiers must be encrypted before being deposited in the VPN accessed or uploaded file or the data must be moved immediately into an encrypted file. |
| Will you be using portable devices for data collection? This includes Laptops, Smartphones, iPods, iPads, PDA’s, etc. |  |  | If yes, then ensure that encryption is installed such that PII/PHI are always encrypted as they are entered or as they are saved or submitted. The ideal is for the encryption to occur during entry, but is acceptable at save or submit functions (for VPN or FTP or other uploading). There are two ways of encrypting data for portable devices: 1) Encrypting the entire device so that a password is needed to even open any operation of the device; and 2) encryption only o the specific file being used for research. |
| Are you planning to put data on small portable storage devices such as jump drives? |  |  | If yes, **ensure that PII or PHI are not stored on jump drives.** Other research data can be stored on jump drives as long as there is no way that the data could be traceable to a participant identity. |
| Are you using a university desktop PC or MAC for entering study data? |  |  | If yes, then ensure that your device is behind the university firewall. Ensure that backups are to secure system servers or on an external hard drive is used for backups, ensure that it contains only encrypted PII or PHI. |
| Are personally owned desktop PCs or MACs being used in the research? |  |  | If yes, do not have PII or PHI stored on personal desktop devices. For all other research data on human subjects, ensure that a firewall is installed and turned on at all times. |
| Will web-based survey tools be used to collect data? |  |  | If yes, be sure to find out to what extent access to the server is limited, what protections are in place to protect the data against unauthorized access, and whether the data can be encrypted upon transmission. Consider using [Qualtrics](https://www.ndsu.edu/gdc/?page_id=546) which is available free of charge to NDSU faculty, staff, or students, or ensure that your data collection platform has been approved by the [NDSU Chief Information Officer](https://www.ndsu.edu/its/security/). |
| Are survey questions of a sensitive nature such that a breach of confidentiality could put subjects at risk? |  |  | If yes, then consent forms should address the possibility of breach of confidentiality and that anonymity cannot be guaranteed. |

EXHIBIT A – CROSSWALK EXAMPLE BETWEEN PII AND DATA TABLES

This table would show personally identifying information (PII) associated with the research record number and would need to be encrypted.

Participant Crosswalk Table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Participant ID | Name | Address | Telephone | EMPLID | DOB |
| 10001 | John Doe | 123 Cherry Ave., Fargo, ND 58103 | 701.555.1234 | 123-45-6789 | 7/6/1982 |
| 10002 | Jessie Smith | 468 Alder Dr., Harwood, ND 58042 | 701.555.9876 | 451-88-8888 | 1/5/1959 |
| 10003 | Jayne Doe | 4963 2nd Ave., West Fargo, ND 58078 | 701.555.1397 | 999-55-1111 | 6/13/1975 |
| 10004 | Jacob Larson | 1656 9th St. W, Moorhead, MN 56560 | 218.555.4682 | 999-19-7777 | 4/30/1985 |

Baseline Data Table:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Participant ID | Gender | age | Question 1 | Question 2 | Question 3 | Question 4 |
| 10001 | M | 33 | 5 | 2 | 4 | 3 |
| 10002 | F | 56 | 4 | 6 | 3 | 1 |
| 10003 | F | 40 | 2 | 5 | 6 | 1 |
| 10004 | M | 30 | 3 | 6 | 1 | 2 |

Definitions for electronic data collection:

Protected health information (PHI): Individually identifiable health information transmitted or maintained by a covered entity or its business associates in any form or medium (45 CFR 160.103). The definition exempts a small number of categories of individually identifiable health information, such as individually identifiable health information found in employment records held by a covered entity in its role as an employer.

Personal identifying information (PII): For the purposes of these guidelines, this includes information that identify a person including any or all of the following: (1) names; (2) social security numbers; (3) birthdates; (4) addresses; (5) IP addresses; (6) other data that could reasonably lead to discovering a personal identity.

Server: A server is a computer device with software that networks/links PCs and databases or web applications.

PC/Personal computer: A stand alone or networked desktop computer.

Laptop: A portable computer that includes traditional laptops, netbooks, and other portable computing devices that generally have full range PC capacities.

External drives: This includes everything from jump drives to external hard drives.

Compact disks and DVDs: Plastic disks for storing electronic data.

Security tokens: Jump drive-like devices that contain security codes or de-encryptions to allow access to secure web-based data sets.

Smartphones, tablets, PDAs, and other mobile devices: Portable computing devices that send and receive emails, text or phone messages, or other communications and that include data entry and data storage capacity.

Virtual Private Networks (VPNs): The university allows access to selected drives and folders on university servers from remote locations using software provided by the NDSU phone system. With a VPN, a researcher can connect to files from off-site computers using either Ethernet or wireless connectivity. VPNs are password protected.

File Transfer Protocols (FTPs): FTPs are used to transfer data from off-site computers to main campus servers via web connections to specified server files/