Confidentiality and Data Security Guidelines

Much of today's research data will be collected, transmitted, shared, and/or stored electronically. Per federal regulations, the IRB is required to determine the adequacy of provisions to protect the privacy of subjects and to maintain the confidentiality of research data. Investigators need to consider data security issues when developing research protocols that include electronic data collection, transmission, processing, and storage. This guideline provides best practices for managing electronic data within the Franklin University research community. Investigators going abroad should consult the laws and regulations in the countries where they are conducting research to ensure adequate protections are in place.

Principal Investigators (PI) are responsible for ensuring all research data is collected, transmitted, shared, analyzed and/or stored securely. PIs may share raw, identifiable data only with authorized individuals that are listed on the IRB protocol and approved by the IRB. PIs must ensure co-investigators and all personnel are appropriately trained and understand their responsibility to collect and/or interact with research data securely. Data security and data monitoring plans should be discussed regularly in team meetings, and data security details must be included in the study data and safety monitoring plan.

DEFINITIONS

- **Data:** Any information gathered during the research including – but not limited to – documents, databases, spreadsheets, survey responses, text, video recordings, voice recordings, transcripts, field notes, test scores, clicks, IP addresses, searches, social media postings, and photographs.
- **Anonymous Data:** Data that at no time has a code assigned that would permit the data to be traced back to an individual. This includes any information that was recorded or collected without any of the 18 identifiers as defined by HIPAA.[1] Note that IP addresses are considered by the University and some international standards to be identifiable even though the address is linked to the computer and not specifically to the individual.
- **De-Identified Data**: Investigator cannot readily ascertain the identity of the individual.
- **Coded Data**: Identifying information (such as name) that would enable the investigator to readily ascertain the identity of the individual to whom the private information or specimens pertain has been replaced with a code (number, letter, symbol, or any combination) and a key to decipher the code exists, enabling linkage of the identifying information to the private information or specimens.
- **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

[1] We apply the HIPAA identifiers that make health information PHI to our standard definition of personal information. The 18 identifiers are: names; dates, except year; telephone numbers; geographic data; fax numbers; Social Security numbers; email addresses; medical record numbers; account numbers; health plan beneficiary numbers; certificate/license numbers; vehicle identifiers and serial numbers including license plates; web URLs; device identifiers and serial numbers; Internet protocol addresses; full face photos and comparable images; biometric identifiers (i.e. retinal scan, fingerprints); and any unique identifying number or code.

information that may pertain to health. Information/data from patient records are considered PHI.

- **Personal identifying information (PII)**: For the purposes of this policy, this includes information that identify a person including any or all of the following: (1) names; (2) social security numbers; (3) birth dates; (4) addresses; (5) IP addresses; (6) social media identifiers; (7) geographic data; (8) other data that could reasonably lead to discovering a personal identity.
- **Sensitive Research Data**: Data is considered sensitive when disclosure of identifying information could have adverse consequences for subjects or damage their financial standing, employability, insurability, or reputation.

FEDERAL DEFINITIONS

- **Private information** includes information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information that has been provided for specific purposes by an individual and that the individual can reasonably expect will not be made public (e.g., a medical record). Private information must be individually identifiable (i.e., the identity of the subject is or may readily be ascertained by the investigator or associated with the information) in order for obtaining the information to constitute research involving human subjects. 45 CFR 46.102(e)(4)
- **Individually Identifiable**: Identifiable private information is private information for which the identity of the subject is or may readily be ascertained by the investigator or associated with the information. This occurs when private information can be linked to specific individuals by the investigator(s) either directly or indirectly through coding systems. 45 CFR 46.102(e)(5)

DISCUSSION

Researchers have a responsibility to be good data stewards. Much of data is collected, transmitted, processed, and stored on computers and mobile devices. Simply password-protecting a computer may not be sufficient to meet the security standards mandated by the IRB and/or sponsors. Researchers should consult IT staff and data security colleagues who have the expertise to evaluate the security methods most appropriate for the sensitivity of the research data. Doctoral candidates should collaborate with their dissertation committees to devise data security plans for their research.

Data that will be shared with others requires additional oversight to uphold the privacy of the research participant and the confidentiality of their data. If data from the study is to be shared outside the research team, it is important that the researchers obtain the appropriate consent from study participants. Some researchers may request permission to share identifiable data, but the majority will be sharing de-identified data. Many sponsors, including federal agencies, require data sharing as a condition of funding, and this must be reflected in the consent document and, most importantly, in the consent process (discussion). This includes the acknowledgement of the data sharing practices and the possible risk of re-identification when applicable. One should never guarantee that de-identified data cannot be re-linked and the participant's identity disclosed. As technology evolves, so does the potential risk of re-identification.

ASSESSING THE DATA SECURITY METHOD NEEDED

Based on the type of data involved in the study, the IRB is required to 1) assess potential risks to participants, and 2) evaluate the researchers' plan to minimize risks. All research activities result in some type of risk and the researcher has the responsibility to mitigate the risk of improper disclosure.

What is the risk?

- Is the data identifiable, de-identified (coded), or anonymous?
- Is sensitive information being collected that could result in harm to participants?
- What is the risk of harm to the participant or others?

What are the protections against anticipated threats or hazards (during collection, transmission, storage)?

- Encrypting data on devices to protect against loss/theft of device
- Using secure data transmission channels to protect against data interception
- Using strong passwords to protect against unauthorized access
- Storing data behind a secure firewall whenever possible
- Ensuring strong data security controls on all storage sites

DATA SECURITY PLAN

Franklin University has implemented a data security levels model to help researchers identify the type of data security measures required for their study. Please review these levels to determine where your study falls and prepare a data security plan that aligns with the degree of risk associated with your study.

As applicable and at a minimum, all data privacy plans should include a discussion of the following components:

- Data recording method (i.e., survey, de-identified secondary dataset, voice recorder, Zoom, pen and paper, etc.)
- End-to-end encryption
- Data storage (i.e., recording to PC or Cloud)
- Securing recordings and data
- Passwords

DATA COLLECTION TOOLS

Each data collection tool used in a study has its own set of data security considerations. Tools involving the internet are especially vulnerable to security breaches and require layers of protection to mitigate the possibility of exposing human subjects to research-related risks.

ZOOM

Zoom is a virtual meeting tool that is free to anyone who signs up for basic access. The basic plan does not include end-to-end encryption and Cloud recording transcripts. To access this feature, students and other investigators will need to have access to a Zoom business license. There are other virtual meeting options available if you cannot acquire Zoom's business license.

There are several advantages to using an auto transcription feature:

- Manual transcription takes approximately one hour for 15 minutes of clear audio, or roughly four hours for one hour of audio.
- Paying a third-party transcription service will be costly, and auto transcription helps to protect participants' privacy.
- Transcripts will be completed much more quickly than manual transcription, though you should review the transcript for accuracy.

There are data security concerns with using Zoom that can be mitigated by using best practices. If you plan to use Zoom Cloud recording features, you must include the following steps in your IRB application and implement the procedures as part of your research practice. These steps are designed to minimize data breaches and protect participants' privacy.

- Cloud recordings will be processed and stored in Zoom's Cloud after the meeting has ended. These recordings will be password protected and available only to the PI, who will be conducting the interviews.
- The recordings will be stored in both video and audio format.
- Per Zoom policies, Cloud recordings and audio transcripts will be stored encrypted.
- Upon downloading the recordings and transcripts stored in the Cloud, the PI will delete them from the Cloud. The recordings and transcripts will then be saved to the PI's password protected computer in an encrypted folder.
- NOTE – You can record video and audio and choose to download the audio only. You can download the audio file, the transcript, and then delete the recording.

SURVEY SOFTWARE

We recommend that investigators use survey tools such as SurveyMonkey and Qualtrics to support teaching, academic research, and institutional business. Many investigators wish to collect the IP addresses of survey participants to provide a method of determining whether the user has previously completed the survey. The University and some international standards consider IP addresses to be identifiable information. This is important to consider when conducting surveys, especially if the consent process indicates that a participant's responses will be anonymous. When using SurveyMonkey or Qualtrics, check the option to anonymize the data collection process and do not collect the IP address. If IP addresses are necessary to the research, include in the consent process that you will be recording this information.

MOBILE APPS

Many researchers are purchasing mobile apps to interact with study participants. Even if the participant is asked to download a free app or is provided monies for the download, the researcher is still responsible for disclosing potential risks. It is possible that the app the participant downloaded will capture other data stored or linked to the phone on which it is installed (e.g., contact list, GPS information, access to other applications such as Facebook). The researcher has the responsibility to understand known or potential risks and convey them to the study participant. Commercially available apps publish "terms of service" that detail how app data will be used by the vendor and/or shared with third parties. It is the researcher's responsibility to understand these terms, relay that information to participants, and monitor said terms for updates. Additionally, it is important that the researcher collect from the app only the minimum data necessary to answer the research questions.

MINIMIZING DATA SECURITY RISKS

There are several ways to mitigate data security risks that are discussed in this section. An important first step in assessing risk is evaluating the level of risk. Are you collecting sensitive, individually identifiable data? To that end, any PII should be separated from the rest of the data as quickly as possible. For tips on de-identifying data, refer to pp. 4-6 in Data Security Procedures for Researchers.

PASSWORDS

It is imperative that you create and use strong passwords to protect data records. Poor passwords can undue all other efforts to securely store data. Cloud storage is especially vulnerable to hacking and therefore requires strong passwords to supplement encrypted cloud storage. You need to have a strong password for your cloud service and another strong (but different) password for individually encrypted files that are stored in the cloud.

Passwords should never be reused from site-to-site or app-to-app. Use of a password manager such as 1Password, LastPass, or Dashlane is recommend for this purpose. These browser-based plugins and mobile apps will automatically create unique, strong passwords for each login. Users only need to create an exceptionally strong master password to unlock the password manager. Secure notes within these apps can be used to store passwords for encrypted files as well.

Always use complex passwords that are difficult to decode and protect those passwords as you would private, confidential data. You should aim for a password that would take more than one month to

crack. To test the strength of your password, use https://www.my1login.com/resources/password-strength-test/.

ENCRYPTION

Encryption protects data by encoding information so that only authorized parties may read it. Encryption can occur "at-rest" where the data is being stored and "in-transit" as the data is being moved from one location to another.

How to Encrypt Files

1. Create a directory to hold the data you want to encrypt.
2. Right-click (or press and hold) the folder and select **Properties**.
3. Select the **Advanced** button and select the **Encrypt contents to secure data** check box.
4. Select **OK** to close the **Advanced Attributes** window, select **Apply**, and then select **OK**.

You can also use an encryption service to protect your files before sending them to others. See, for example, Encrypto: https://macpaw.com/encrypto

For more detailed guidance on encryption options, see https://www.howtogeek.com/170352/how-to-password-protect-files-and-folders-with-encryption/. Any data that you put into this encrypted directory will itself be encrypted locally.

RECORDING AND TRANSCRIPTION

Audio recording interviews is common practice in research and has many benefits for the researcher and the study. Investigators must consider how recordings will be captured and what happens to those recordings afterward.

If an interview will be recorded, and the subject has specifically consented to audio and/or video recording, you have the option to record locally (i.e., directly to your PC) or to a cloud. While a local recording is a safer choice that cybersecurity experts recommend, you can record to a cloud provided you use a tool that supports end-to-end encryption. If you are recording to a cloud, it is imperative that you use a strong password and encrypt individual files with a different password.

Similarly, if you use your phone to record an interview, a local recording is the preferred, safer option. To record locally, disable auto uploads to your cloud. If you record to a cloud, refer to the data storage guidance below.

Researchers now have several options for transcribing interviews: manual transcription by the investigator; third-party transcription service; and cloud, auto transcription. As mentioned previously (see Zoom section), there are a number of benefits to auto transcription services. Your data security plan should include a discussion of Zoom or another auto transcription service (e.g., iCloud) that you may use and the security measures in place to protect cloud recordings.

If you use a third-party transcription service, it is recommended that you select a transcription professional who will guarantee the privacy of subjects' information and confidentiality of the data in writing. Transcription businesses typically require their employees to sign Non-Disclosure Agreements

(NDA) as a condition of employment, but you should verify this prior to signing an agreement, paying for services, or sending files for transcription. You should review and implement data security measures for transmitting data electronically.

DATA TRANSMISSION

The process of transmitting data is often overlooked as a risk. The plan to protect confidentiality should describe the methods to protect the data during collection and sharing both internally and externally to the University. It is advisable to utilize a secure transmission process even if the data is anonymous, coded, or non-sensitive information. If the research team develops a best practice on using a secure data transmission process, then it is less likely a data breach will occur. Email notifications are generally not secure, except in very limited circumstances, and should not be used to share or transmit research data unless you first compress the files (discussed below). Text messages are stored by the telecommunications provider and therefore are not secure. Data should be encrypted when "in-transit", and there is a wealth of resources available for this. Terms such as Secure Sockets Layer (SSL and HTTPS) or Secure File Transfer Protocol (SFTP) are indications that the data is being encrypted during transmission.

Many compression tools (WinZip, 7-Zip, etc.) can also be used to encrypt files before transmission and are recommended when research files contain PII or other confidential information. This is especially important when transferring files as attachments to email or as files on physical media such as flash memory drives. Compressed files reduce the risk of failed file transfers.

Microsoft

To zip (compress) a file or folder

1. Locate the file or folder that you want to zip.
2. Press and hold (or right-click) the file or folder, select (or point to) **Send to**, and then select **Compressed (zipped) folder**.

   A new zipped folder with the same name is created in the same location. To rename it, press and hold (or right-click) the folder, select **Rename**, and then type the new name.

To unzip (extract) files or folders from a zipped folder

1. Locate the zipped folder that you want to unzip (extract) files or folders from.
2. Do one of the following:
   - To unzip a single file or folder, open the zipped folder, then drag the file or folder from the zipped folder to a new location.
   - To unzip all the contents of the zipped folder, press and hold (or right-click) the folder, select **Extract All**, and then follow the instructions.

Notes

- To add files or folders to a zipped folder you created earlier, drag them to the zipped folder.
- If you add encrypted files to a zipped folder, they'll be unencrypted when they're unzipped, which might result in unintentional disclosure of personal or sensitive information. For that reason, we recommend that you avoid zipping encrypted files.

- Some types of files, like JPEG images, are already highly compressed. If you zip several JPEG pictures into a folder, the total size of the folder will be about the same as the original collection of pictures.

Mac

1. Using Spotlight, type in **terminal**.
2. Change your directory to the location of your folder and hit enter.
3. Type in **ls** to make certain you see your folder.
4. Now type in the command that will simultaneously compress and password protect the files in the form of **zip -er FILENAME.zip FILESorFOLDERStoCOMPRESS**.
5. Hit **return/enter**.
6. Next, you'll be prompted with a **password prompt**. Type in the password you want to use to protect the files. Enter the password twice.
7. You should now see your **compressed and password protected zip file** in Finder.
8. You can **test** the protection by double clicking the file.
9. Enter your **password**.

DATA STORAGE

It has become common practice to store some level of personal information in the Cloud with services such as Google Drive, Dropbox, Office365, and Amazon. Using such services can often result in cost savings; however, special attention must be paid to potential security risks, export control restrictions, and data ownership issues.

Currently, the University's cloud-based storage is OneDrive. Only data that meets HIPAA de-identification standards should be stored unencrypted on OneDrive. While data saved to a personal PC is more secure than cloud storage, you may use cloud storage provided that adequate security measures are in place. This applies to all research data – PII, de-identified data, and audio and/or video recordings. Data stored in clouds must be encrypted and have strong passwords (as determined by the password strength test). See the 'Passwords' section for more information.

Investigators should determine what they will do with their data when a project concludes and include this in their data plan. Federal regulations require that study data and consent forms be stored securely for a minimum of three years after a project closes. If you plan to destroy all research documents after this time, paper records should be shredded and all electronic files must be permanently deleted. If your plan includes longer retention of data, for purposes of future use such as publications or additional analysis and regardless of its form (i.e., hard or electronic copies), you must have a secure storage plan in place (i.e., a safe or locked filing cabinet for hard copies, encrypted and password protected files for electronic copies).

CHILDREN'S ONLINE PRIVACY PROTECTION ACT (COPPA)

The Federal Trade Commission enacted COPPA in 2000 (revised in January 2013), which applies to the online collection of personal information from children under the age of 13. This Act requires websites to display a privacy policy, obtain verifiable parental consent, and disclose how the information will be used. It is important that researchers who plan to collect data from children online carefully review the provisions of the Act. It is the responsibility of the researcher to ensure they are fully compliant with the COPPA regulation.

OTHER CONSIDERATIONS

The data and safety monitoring plan should indicate that research team meetings include discussions about, but not limited to:

- Software on computers to protect against malware
- Data security to ensure all software updates and patches are being applied
- Data collection, transmission, and storage methods employed
- Data collected is only that data necessary to answer the research question
- Codes are not stored with the corresponding de-identified data
- Encryption methods are being used on all portable devices (laptops, mobile devices, and removable storage)
- This section should include a data security plan to describe how the research data will be protected during collection, storage, transmission, and destruction to ensure it meets Franklin University policies.

Think about the Consent process and documentation:

- What are the expectations of the research participant?
    - Investigator will protect their privacy and confidentiality of their data.
- What information is needed to answer the research question?
    - Investigator should collect information that is required to answer research questions and not more just because it is possible.
- Think about personal privacy and do not promise confidentiality.

Include a detailed description of any research activities the research participant will perform that entail the use of the any electronic data (e.g., accessing a website, downloading an app, text messages, completing a survey) so the IRB can determine that risks are minimized.

RESOURCES

U.S. Department of Health & Human Services
Human Subjects Research and the Internet

Federal Trade Commission
Understanding Mobile Apps

HealthIT.gov
Your Mobile Device and Health Information Privacy and Security

---

Modified from the University of Pittsburgh's Human Research Protection Office (HRPO).