Data Management and Integrity in Human Research

Revised December 2017

Researchers must develop and follow protocols to manage and protect the security and integrity of participant data, particularly Protected Data. This document provides practical recommendations to help TWU's researchers protect participant data.

Types of Protected Data

Personally Identifiable Information (PII)

Personally Identifiable Information (PII) can be used on its own or with other information to identify, contact, or locate a single person, or to identify an individual in context. Examples include name, social security number, date and place of birth, mother's maiden name, or biometric records; and any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information.

> **So if I leave off the social security number, I'm ok, right?**
> Wrong. While the SSN should not be collected at all, each of these examples are PII and when multiple PII data are collected on a single participant, that person's personal information and security are at risk.

FERPA

FERPA protects the privacy of student educational records. If you must collect student records through the Registrar, Colleague, a SQL report, or through some other method, you are expected to:

- Use the information only for purposes of the approved research project. Any new use of the information requires new approval.
- Provide adequate protection for the information to ensure that it is not compromised or subject to unauthorized access.
- Ensure that no one outside the research team has access to the information.
- Destroy the information within a reasonable time after completion of the research.

HIPAA

HIPAA provides comprehensive federal protection for the privacy of health information. Learn more at the NIH website. The Privacy Rule permits certain incidental uses and disclosures that occur as a by-product of another permissible or required use or disclosure, as long as the covered entity has applied reasonable safeguards and implemented the minimum necessary standard, where applicable, with respect to the primary use or disclosure.

Data Collection Guidelines

Researchers must comply with TWU's Office of Research & Sponsored Programs IRB Policies. Further, if the researcher is **videotaping a classroom**, consent must be obtained from everyone in the classroom or the camera must be situated so that there is no chance people who haven't signed a consent form are captured on the video or audio.

If collecting data via an **online survey** (such as Survey Monkey, Qualtrics, PsychData), the researcher will likely be storing PII. As soon as data collection is complete, data should be exported and the survey and results removed from the survey website. As with all research data, we recommend both encryption and regular backup. Each of these data collection services offer differing levels of security. Survey Monkey is the least secure, though their professional accounts offer SSL encryption.

Ernst and Young suggest a holistic approach to fight the battle of data loss, over internet or e-mail exchanges and to maintain and secure sensitive/ confidential data. They suggest using the data loss prevention (DLP) approach, supported by integrating the information technology infrastructure. The essential element is data governance, which, when coupled with the DLP approach, can achieve a multilayered defense mechanism (E&Y).

Data Management and Protection

Data management includes ownership, collection, storage, protection, retention, analysis, sharing, and reporting. **Data storage and protection** are critical aspects of human research data management. Proper data management includes protection.

ORI suggests protection is best ensured by **limiting** access to data, **protecting** the computer systems with an updated antivirus, and **encrypting** the data (ORI, Steneck 2007).

Guidelines for Human Research Data Storage and Protection

1. Data Management Plan. As you begin the research project, work with your fellow researchers to **discuss how data** will be collected, maintained, archived, and protected.

Once data are collected, these files are to be protected as they will likely contain PII, FERPA and/or HIPAA info. We recommend assigning each participant a unique identifier which is not at all related to their personal data. This unique identifier could then be used in shared datasets to link the participants between the protected dataset and the shared datasets.

Unique identifiers can easily be created in Microsoft Excel. Import the data you need to share into Excel. Insert a column. Give it a title, like "Unique ID." Type in an alphanumeric string. For ease of use, the string should end in a number and the number should contain enough integers to account for the maximum participants expected in the study. To the right, you see the first UniqueID created is 1001. It could

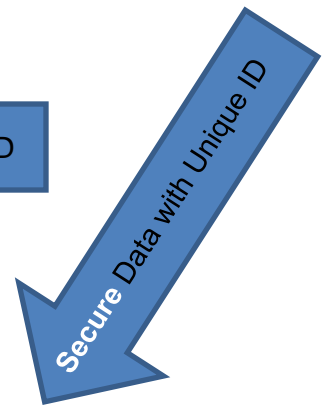|	A	B	C	D	E	F
1	UniqueID	InterviewDate	Q1	Q2	Q3	Q4
2	1001	9/5/2013	0	0	2	7
3	1002	8/31/2013	1	0	0	3
4		9/2/2013	1	0	1	0
5		9/4/2013	1	2	4	8
6		8/31/2013	0	1	3	5
7		9/2/2013	1	1	0	5
8		9/6/2013	0	2	4	7
9						

also be EX101 or any other string of characters ending in a multidigit integer. Repeat the naming convention in the cell below, increasing the value of the string by one. *(If EX101, then EX102 below)*

You can drag this single value down the column to create a non-repeating set of unique identifiers. Highlight both cells containing unique ID values. Place your cursor over the lower-right corner of the bottom cell. A plus sign appears. Click and drag the plus sign down through your spreadsheet until all data have a unique ID.

	A	B	C
1	UniqueID	InterviewDate	Q1
2	1001	9/5/2013	0
3	1002	8/31/2013	1
4		9/2/2013	1
5		9/4/2013	1
6		8/31/2013	0
7		9/2/2013	1
8		9/6/2013	0

	A	B	C
1	UniqueID	InterviewDate	Q1
2	1001	9/5/2013	0
3	1002	8/31/2013	1
4	1003	9/2/2013	1
5	1004	9/4/2013	1
6	1005	8/31/2013	0
7	1006	9/2/2013	1
8	1007	9/6/2013	0

Shared Data with Unique ID

Secure Data with Unique ID

	A	B	C	D	E	F	G	H
1	UniqueID	InterviewDate	Q1	Q2	Q3	Q4		
2	1001	9/5/2013	0	0	2	7		
3	1002	8/31/2013	1	0	0	3		
4	1003	9/2/2013	1	0	1	0		
5	1004	9/4/2013	1	2	4	8		
6	1005	8/31/2013	0	1	3	5		
7	1006							
8	1007							
9								

	A	B	C	D	E	F	G
1	UniqueID	InterviewDate	Gender (Q1)	DOB	Age Range (Q2)	Daily Avg B.S. Tests (Q3)	Yrs Since Diagnosis (Q4)
2	1001	9/5/2013	Male	10/8/1955	55+	2	7
3	1002	8/31/2013	Female	9/1/1945	55+	0	3
4	1003	9/2/2013	Female	12/26/1950	55+	1	0
5	1004	9/4/2013	Female	3/5/1991	18-30	4	8
6	1005	8/31/2013	Male	6/7/1967	31-54	3	5
7	1006	9/2/2013	Female	4/18/1974	31-54	0	5
8	1007	9/6/2013	Male	11/17/1980	18-30	4	7
9							

Properly Ejecting a USB Drive

USB drives are notoriously unstable and can damage data on hard drives. Properly removing USB drives is essential to data protection.

1. Click the Show Hidden Icons button on your desktop tray (lower-right corner).
2. Select the Safely Remove USB Drive icon. Select your drive and remove.

Customize...
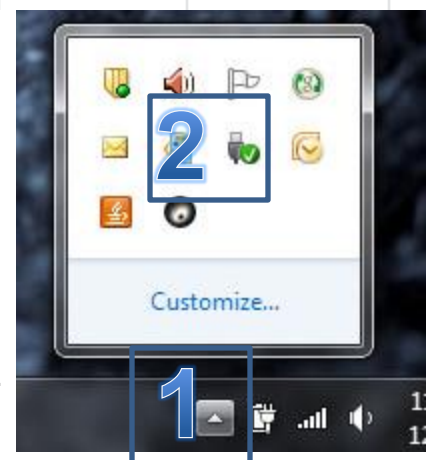
2. **Data Security**. There are many schemes for securing data in a personal computer. The most secure method is to use a removable hard drive that resides in a safe when not in use. To process data one simply installs the hard drive, disconnects all other storage devices (e.g. USB drives), disconnects the network, and then boots the computer to do processing. This procedure is very secure, but often impractical. **Minimal effort at data security is password protection.** You can easily password protect SPSS files, Excel files, SAS files.

Encryption is the key to secure data and it is TWU's policy. Symmetric cryptosystem key lengths should be at least 128 bits for confidential data and other agency-sensitive information identified by TWU. Confidential or agency-sensitive information transmitted over the internet or as an email message must be encrypted.

There are many software solutions for encrypting data. For Windows, we recommend using BitLocker Drive Encryption. Need help encrypting files? Contact the TWU Service Desk. [An earlier version of this document recommended using TrueCrypt – this product is no longer supported and files are subject to risk. Migrate from TrueCrypt to BitLocker.]

Store physical data (either electronic devices - such as external hard drive, flash drive, DVD/CD - or paper hard copies) in an **area with key access**. Restrict key access to only top research personnel. It is the responsibility of the principal investigator to make the decision regarding the access to confidential research data and training the research personnel as required (HRSA Video Training).

Video and audio recordings are physical data that should receive high-level security. We recommend assuming these data contain PII and storing them in an area with key access because a participant may share something personal on the recording such as medical history or a home address.

3. **Data Retention**. Typically, research data must be retained **for at least 5 years** from the date of the most recent IRB approval. Retention requirements may vary by discipline, and any requirements stipulated in an IRB approval must be followed. The informed consent forms or the short form and the research summary must be retained by the investigator on behalf of the institution for at least 3 years unless it has been waived by the IRB (HHS).[1] We recommend storing two copies on two separate USB Drives. **Identifiable data** (such as video recordings) are recommended to be destroyed as soon as a transcribed copy has been created and backed up.

Data Destruction. When appropriate, all data containing personal, medical, or educational record data should be properly and permanently erased. Paper should be shredded in a cross-cut shredder; CDs and DVDs should also be processed through an industrial shredder. Portable hard drives, dedicated hard drives, and USB flash drives should be wiped using software.

After the three years, can I just hit my USB drive with a hammer to destroy it?
The TWU Service Desk has a Degausser, which safely and securely removes data from removable drives – this includes drives used in video cameras. Videotape can also be erased through the Service Desk tool. The Degausser permanently destroys the drive and tape, so they cannot be used again, but they can be disposed of without risking the data.

---

[1] If the investigator designated for the purpose leaves the institution, it is the responsibility of the institution to replace the investigator by a responsible representative for the research data management as suggested by the HHS. In addition, HHS also requires that the research data are secured in hard copy or soft copy format and be accessible for either inspection or copying purposes by authorized representatives of HHS.

4. **Data Documentation.** How are the data stored and backed up?

Storage Device	Advantages	Disadvantages
Paper Copies	Inexpensive Low Corruptibility	Susceptible to damage Difficult to share and manage data
USB Thumb / Flash Drive **Removable Hard Drive**	Ease of use Inexpensive	Questionable reliability Prone to corruption Not integrated well for backups Low read/write speed
Cloud Storage (such as Dropbox, Sharepoint, or Google Drive)	Ease of use Accessibility Multiple users can access	Many security issues, not recommended for research data unless data are de-identified as explained earlier in this document Intellectual property ownership concerns

USB drives can only be secured through encryption software. Any electronic files should be both encrypted and password protected. *The easier it is for you and your team to access the files, the easier it is for that data to be compromised.*

5. **Planning for Data Disaster**.

- Regularly **back up your data** on an encrypted drive at a key length of at least 128.

CAUTION! Do not forget your password. Data are unrecoverable if the secure password is lost.

- Monitor the progress of research study or clinical investigations according to the approved protocols and the safety of participants. The **PI must assign the duty of regular monitoring of processes and reporting** and have communicated procedures to be followed by the personnel monitoring the process in cases of unanticipated situations or circumstances.
- Adhere to the policies and assure compliance regarding **unanticipated situations** and reporting the same to IRB. Quality control plans must be in place to respond appropriately.
- If the study requires a **temporary suspension or modification of protocol** in order to resolve unforeseen problems, a procedure must be established to first secure the data that were obtained. Then a decision can be made if the data should be discarded, considering possible implications of both data usage and destruction. A plan to this end must be included in the original protocol.
- Find additional **detailed information regarding data and safety management plans**.[2]

---

[2] For Phase I and II trials NIH does not require a data and safety management board, but for multi-site phase I and II studies, it does require investigators to keep reports of adverse events reported in a timely manner. For additional information refer to the NIH website http://grants.nih.gov/grants/guide/notice-files/not99-107.html)

FAQs

How do I know if a drive is secure?
Student researchers should contact the TWU Service Desk. They can discuss data security options with you. Faculty should contact TWU Office of Research & Sponsored Programs.

What if I store data on a flash drive?
While inexpensive and easy to use, flash drives are prone to corruption and difficult to back-up. Use precaution and encrypt the drive.

What if I collect data on a laptop?
Data is not secure on a laptop, even if password protected and frequently backed up, unless the laptop has full encryption of stored data. The laptop should be stored in a secure location. Updated antivirus software should be running on the laptop.

What if I store data on DropBox or Google Drive?
Files should be encrypted and password protected. However, these are not recommended storage solutions for research data.

How can I password protect a data file?
Excel spreadsheets are easily protected. SPSS files (v21+) can be password protected.

What if I collect data on an iPad or iPhone?
As with any other data collection method, routine backup measures are recommended and security protocols should be established, particularly if the device is shared among multiple users.

What about if I email data files? Is that safe/okay?
Email is not secure. Do not use email for transmitting PII data.

What about video/audio recordings stored on a camera or audio recorder?
Recordings should be backed up for security purposes. Devices should be stored in a secure location.

Can I carry data storage devices out of my workplace?
This is not a recommended practice. Data are vulnerable, particularly when they are not in a secure location. You may carry devices out of the workplace if they are encrypted. Password protection alone is not enough.

Can I save personal stuff on the research storage devices?
Your personal data should be stored separately from your research data.

I have a database connected to the internet. Is it secure?
Even an encrypted database can be at risk online. Files must be encrypted and password protected to assure security. An outside review of security is recommended.

Can graduate students take their research data with them?
In most cases, research data will belong to the University.

I am traveling abroad. How should I access my data?
The TWU Office of Technology will provide employees who plan to travel to foreign countries with laptops that meet all the criteria for the "tools of trade" exemption. Employees are required to complete the form on this page to arrange the loan of a clean laptop for trips abroad. [policy]