CPHPR Data Protection Guide

By Mark Sherman, Ph.D.
January 7, 2026

Goals

To guide researchers in developing their protocol such that it sufficiently protects the data collected about human participants in research. The ultimate goal is to minimize risk of data being "leaked" and causing harm to the participants. Any data containing Personally Identifying Information (PII) must be protected. The entire chain of custody of how it will be handled and processed should be described in the protocol. The risk to participants usually stems from re-identification, so the details of how PII will be removed or encoded are important.

Questions the CPHPR will ask of any protocol include:

- Who will be handling the data?
- How will the data be stored? How is access to this storage controlled?
- If containing PII, how, when, and by whom will the PII be removed?
- If elements of PII need to be preserved for longitudinal or multiple-sampling studies, what will be the process to do this while keeping the analysis data clear of PII?
- **If at any point the data were to be leaked, what harm could it potentially cause the participants?**

The critical elements are *storage,* and *access.* Both should be described in the protocol.

Storage

Secure storage of data is critical. The degree of security needed will vary based on the nature of the data, but all projects should follow the same base guidelines:

- Only accessible by the minimum number of individuals required
- Stored only on encrypted devices or encrypted, trusted cloud platforms.

Minimum Access

When data is stored in a shared medium, such as a networked or cloud-based service, the minimum number of people who can access that data should be the smallest number possible to conduct the research. If the data is stored in a singular, physical place, such as

on a USB storage drive, this is easier to enforce. In that case the protocol should include clear policy for making copies of that data off of the singular location: whether it is allowed, and if so, what conditions it is allowed, and when to destroy the copies.

Dropbox and OneDrive are allowable. BUT OneDrive has a pitfall: the interface to OneDrive encourages sharing via link, and links by default are shared with the entire organization, i.e. all of Emmanuel College. Researchers must pay special attention to ensure that the files and folders containing research data are *only* shared with the specific people who need access, and that they don't create links with unnecessary permissions.

We recommend researchers don't ever use the "send link" feature with OneDrive sharing. Add individuals directly using "Manage Access" instead. They will be notified by email automatically by OneDrive.

Encrypted At Rest

Wherever data is stored, it must be done so securely, with a level of encryption that prevents theft. This rule mostly pertains to devices: laptops, phones, and other computers that can be lost or stolen. All Emmanuel-issued laptops are already required to have full-disk encryption, which makes it so if the laptop is lost, the data on it cannot be forcibly removed without the owner's password. This is sufficient protection for most cases. Android and iPhones also have full-disk encryption turned on by default, and have done so since 2016. If you are unsure if your full-disk encryption is active, the Helpdesk can verify that with you.

A potential pitfall here is backup services. If your computer or phone backs up to a cloud service, your data may be accidentally replicated in the backups, creating more opportunities for exposure. We recommend that you consider your personal backup strategy, if any, when writing your data protection protocol.

Dropbox and OneDrive are allowable for data storage. Both have passed significant security audits by many institutions over many years, and their core services are sufficiently secure that data there is unlikely to be exposed. It is the responsibility of the researchers to ensure that these services are used carefully to ensure only the minimum necessary access to the data is granted.

Other solutions may be acceptable as well.

De-Identification

In almost all cases, data from human subjects will require some degree of de-identification. Avoiding Personally Identifying Information (PII) is important because if the data should be leaked, it will be more difficult for a person's information to trace back to them. This is a necessary precaution, but the depth depends on the risk the particular study.

De-identification might be naturally occurring, such as doing a street survey where no identifiable information is collected, or it might require explicit steps in the protocol to scrub PII before analysis. **The committee needs the de-identification measures clearly described.**

How is the data being de-identified?

Who is doing the de-identification?

If data passed through multiple hands, including third party platforms or different research groups, what is the nature of PII at each step?

Sampling Over Time (Longitudinal)

In a study that samples the same individuals more than once, a mechanism to align the samples of the same individual is required. We recommend implementing some kind of artificial identifier, a random or arbitrarily-generated participant ID, to do this. The goal is to reduce the need for researchers to handle PII.

The participant ID might need to be connected to real PII about the participant in order to identify successive samples. If that is necessary, a "key" containing the associations between real identifier and anonymous participant ID should be created. The key is only required when new samples are collected: the key is the only file that contains both the real identifier and the artificial, anonymous participant ID.

Such a key must be kept separate from the rest of the data, and must be kept secure in a way that accidental release of both data and key is extremely unlikely. One solution is the key is stored only on a USB drive that remains in a locked drawer in the PI's office.

The protocol would describe replacing the real identifier with the anonymous one prior to analysis. This replacement would ideally be done by someone who is not conducting the analysis at all, or would not see the rest of the data.

Third-Party Tools

Third-party tools, like transcription services or other analysis platforms, also need to have attention to data security and conform to the ethical requirements for research.

When a third-party tool will be used to handle protected data, we will look to see the company's policies on data security and retention. They don't need to be explicitly compliant, but value security and have taken paints to ensure that they don't hold onto information unnecessarily, that other people or business partners won't have access to it, and other elements of ethical stewardship of data. The key element is whether the data could end up in the hands of partner businesses, such as advertisers, as those parties may not have the strict adherence to ethical standards.

As an example, Otter is a company that does automated transcription. Their policy is great, and considered compliant: Otter.ai privacy policy.

Dropbox and Onedrive are also trustworthy, as long as you set them up carefully, as described below.

Sharing with OneDrive

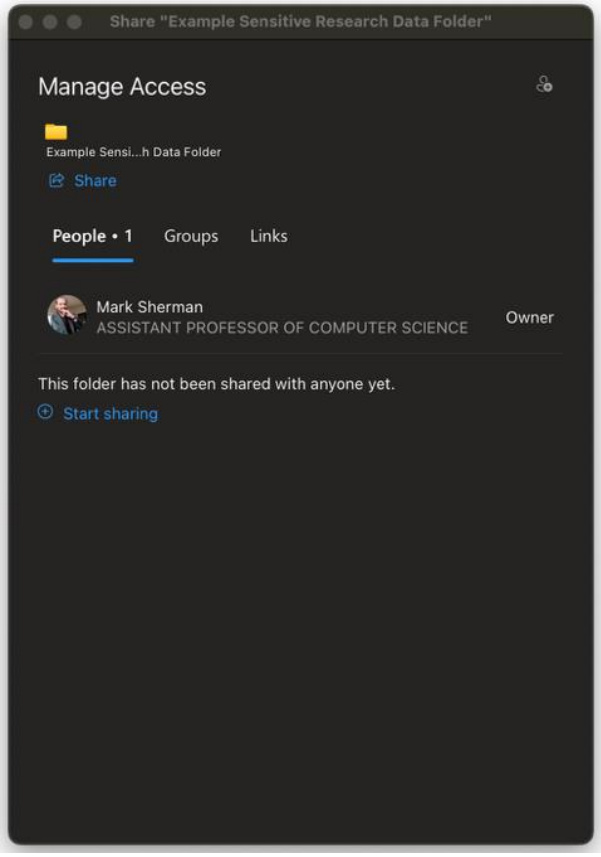
OneDrive is an acceptable storage and collaboration space, but the interface to it makes it easy to mistakenly share that data institution-wide, which is unacceptable. Researchers must be extremely careful in the setup process of their OneDrive folder for human subject research data. Below is, as of January 2026, the process to share through OneDrive safely, highlighting what could go wrong.

The key insight is that sharing by link automatically shares the folder with all of Emmanuel College unless the user is careful.
Avoid sharing by link if possible.

This example is using OneDrive on MacOS. The process is nearly identical on Windows and in the web interface. The location of buttons may be different but the processes are (as of time of writing) the same.

A Fresh Folder

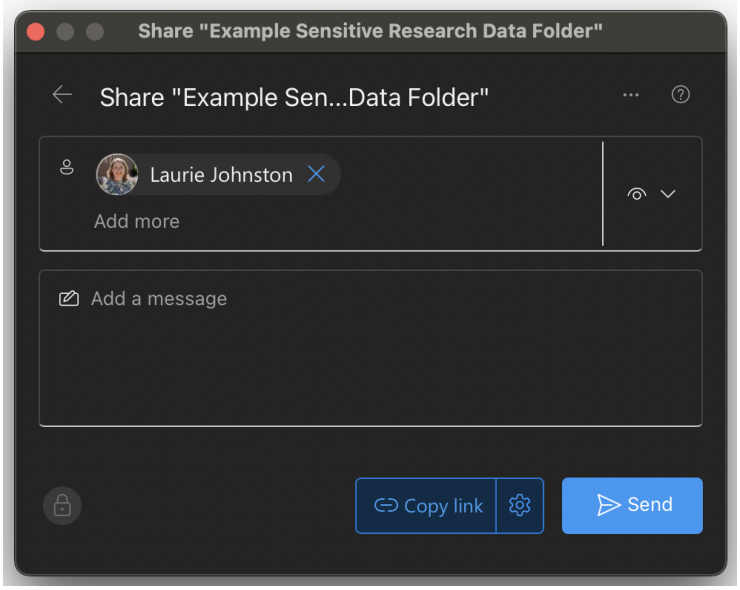
Starting with a fresh, unshared folder, I right-click and select "manage access" from the OneDrive options, giving this window:



It says “This folder has not been shared with anyone yet.”

Adding A Specific Person

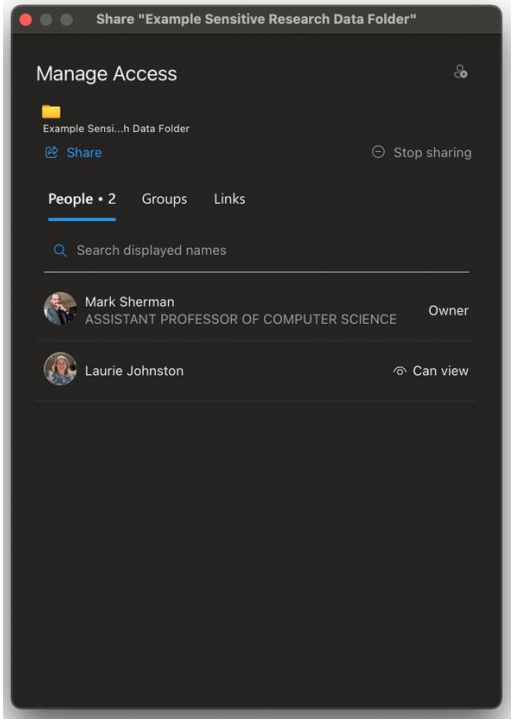
Click “Start sharing” to add a person or people. Don’t click “Copy link” because that can have unexpected share behavior. Ensure the access level for the person is correct, view or edit.



After adding the person, it will send the invite email. This auto-generated email will contain a link, which is safe, as it did not change the sharing scope.

Double Check

Opening the “manage access” window again, seeing correctly configured share.


Share "Example Sensitive Research Data Folder"
Manage Access
Example Sensi...h Data Folder
Share
Stop sharing
People • 2
Groups
Links
Search displayed names
Mark Sherman
ASSISTANT PROFESSOR OF COMPUTER SCIENCE
Owner
Laurie Johnston
Can view