Guidelines for Personal Research Data Management and Security in the NCSU Psychology Department

Updated March 31, 2014

Purpose

Many students and faculty in the Psychology department conduct personal research. A common example of personal research for graduate students is a first year project. In most cases, personal research projects do not require the creation of a data management plan, as do many funding sources such as the NSF. In the absence of a data management plan (and assuming the project is small enough), following these basic guidelines will ensure your project data is stored and secured in a responsible manner. The guidelines we propose are selected to be the most versatile and quick to implement of sufficient security solutions. These guidelines must be implemented in whole to ensure sufficient security.

Quick Reference

Physical Materials	Electronic Materials
● Store in a key locked file cabinet kept in a locked laboratory. ● Ensure only primary investigators have keys. ● Store documents with personal identifiers in a separate locked file cabinet from data. ● Transport materials in a well-bound, covered manner, so no materials are lost or unknowingly read. ● When transporting, never leave materials unattended. ● On IRB, put "physical materials will be stored in a locked file cabinet."	● Store in NCSU Google Drive ● Share with NCSU Google Drive ● Use strong, unique passwords ● Encrypt sensitive files ● Use passwords both at computer and file level ● Log out of Google and any computer when done with work ● Consider transportable electronic data drives to be physical data and take applicable precautions ● Dispose files with Freeraser

- Have lab-wide discussions about data security and protection whenever a new member or research assistant joins the lab, whether graduate or undergraduate. Meetings should ensure no multiple relationships, verify competent performance, and review the security guidelines in this document.
- Ensure that all new lab members complete training on research with human subjects. The CITI program and NIH programs have excellent online training modules.
- Have a plan for storing data for secondary analyses, and a timeline for destroying data when, if ever, it no longer needs to be stored.
- On the IRB form, be sure to address issues of confidentiality, storage, access, and eventual disposal of data.

Defining Sensitive Materials

When thinking about security, it is best to consider ALL research materials involving participants sensitive. This includes materials collected from participants (e.g. consent forms, demographic questionnaires, notes, experiment data, etc.) and anything else containing participant information (e.g. experimenter notes, experiment schedules, participant lists, etc). Physical materials are those that are tangible, such as paper documents.

Physical materials also include the devices containing electronic data, such as flash drives and hard drives. Electronic materials are those used on computing devices, such as files

Collection

Before you collect data, make sure you have a strong reason for collecting all of the data that you are collecting. Risk can be mitigated by collecting as little sensitive information as possible.

Storage and Disposal

Sensitive materials should only be kept as long as they are needed. When they have no foreseeable use, they should be destroyed/deleted. Reasons to retain information include but are not limited to legal (e.g. meet institutional requirements) and future research (e.g. replications, meta analysis, reuse of data set, etc.). Note the APA considers the facilitation of future replications of research design and analyses to be ethical responsibilities, so data must be retained. For as long as the information needs to be retained, a storage policy should be in place to reasonably secure sensitive materials.

In all cases, personal identifiers (e.g. name) should be separated from data. Coding techniques, such as using a random participant ID number for each participant, should be used instead of personal identifiers. That means everything from putting the information on different pieces of paper to storing in different locked containers.

Physical Data

Store physical data in a low-traffic location behind a locked door in a locked container to which only investigators of that project have access. To dispose of materials, shred or physically destroy them so they are no longer comprehensible.

Electronic Data

Electronic data is easily copied, so there is a temptation to store it in as many locations as possible. Protection from data loss is at odds with data security. More locations means more risk. Researchers should only back up data to the extent necessary to prevent loss and avoid making unnecessary copies of data. To dispose of electronic material, use a tool, such as Freeraser (http://www.freeraser.com/) for individual files or DBAN (http://www.dban.org/) for entire hard drives, to overwrite bits in memory and the disk. Note deleting a file through the Recycling Bin does not remove it from the storage device. A malicious user could still recover that information.

Use NCSU Google Drive - The solution is simple. Store everything on Google Drive tied to your NCSU account. Google Drive offers all the convenience of cloud data storage services. It serves as backup in the event of computer crash. It seamlessly synchronizes information between your computers. It eliminates the need for data transfer (e.g. moving between computers on a thumb drive) which can result in data loss (e.g. corrupt file) or unauthorized access (e.g. stolen or misplaced thumb drive). Uniquely, the Google Drive account tied to your NCSU login is guaranteed to store files only on US servers. Other cloud services, such as Dropbox (or even a non-NCSU Google account), are likely to store data on non-US servers, so those services should not be used.

Security

Physical Data

Secure data with a key lock or combination lock with 30 or more possibilities. Only primary investigators of that

project should have the key or combination to access the data.

Electronic Data

Logout of Google When Done - When logging into Google, tell your browser and Google Drive software to NOT remember your password. Also, be sure to log off of Google when you are done working. This will prevent access to Google Doc files if an unauthorized person uses your computer.

Encrypt Files - Encryption is different from standard password protection. If your storage device (e.g. hard drive) is stolen, files can be extracted if unencrypted. Encryption jumbles up the bits in a file, so computers can't understand the contents without the password. Usually, you won't need to enter the encryption password to work with the file, unless you try to open it on another computer. Your operating system will have a feature to encrypt individual files and may also provide the option to encrypt the entire hard drive, such as with Windows BitLocker (http://windows.microsoft.com/en-us/windows7/products/features/bitlocker). There are also applications available for download that do this. TrueCrypt (http://www.truecrypt.org/) is a safe one. Be careful not to lose the password or key used for encryption. An encrypted file cannot be recovered without the password. You may wish to use a password manager (see guidelines below) if you have many passwords.

Password Protect Individual Files - Your computer logon password is not sufficient for protecting data stored on the hard drive. You should password protect all possible sensitive files, so every time they are opened, the user is prompted for a password. This can be done within the application, such as MS Word and SPSS. Note, files created through the Google Docs interface cannot be password protected. However, files created locally, such as MS Word, that are uploaded and stored on Google Drive should be password protected through the application. Be careful not to lose the password or key used. File may not be recoverable without the password. You may wish to use a password manager (see guidelines below) if you have many passwords.

Password Guidelines - Rules for strong passwords are described in Appendix A: What Makes a Strong Password. You should follow all of those rules. Also, you should never reuse a password, and passwords should be committed to memory. If you have difficulty remembering passwords, there are multiple solutions. First, you can apply techniques to make the password memorable, such as creating an initialism where each letter corresponds to the first letter of the word in a phrase or story. Second, it is ok to use a secret system for personal passwords, such as always having the same first six characters (the base) and adding a unique ending applicable to what the password is used. However, no one else can know your system. Third, you can use a secure password manager to remember your passwords for you. If you choose this option, be very careful the manager is secure. One approved option is LastPass (https://lastpass.com/).

Layered Security - Passwords should be used whenever possible. You should require a password to access your laptop (on boot, awaken, and after timeout), services (e.g. Google Drive), and individual files. Although not yet required, multiple-step verification is highly recommended when possible. For example, Google accounts can be configured to text a code to your phone when you login and that code must be entered to access the account. Note, if you use multiple-step verification, you should enable it for every login, not just the first time a device is used.

Transportation and Sharing

In many research settings, individuals other than the primary researchers may have access to the data. While this brings new security concerns, there are measures that should be taken to minimize risk.

Discuss Data Protocols at Lab Member Intake - When a new person is joining a lab, whether a graduate student or undergraduate student, even for a temporary basis, discuss the importance of and protocols for data security and management. It may be useful to give an example of a problem that could arise, and discuss the repercussions for the lab if data were compromised. An important point is that even if data are viewed as 'non-sensitive' the participant will have to be notified if data is compromised, and the reputation of the lab will be tarnished.

Conduct Human Subjects Training - Before anyone has access to data, or interaction with participants, they should undergo human subjects training. Two such excellent modules are the CITI program, at citiprogram.org, and the NIH human subjects training course, at phrp.nihtraining.com.

Replications and Reusing Data - In many situations another researcher may ask you for your data for the purpose of replication or for extended analyses. You have an ethical responsibility to share non-personally identifiable information, but you must do it carefully. You must require a written agreement from the other researcher outlining a data management plan. Be sure to not share any personally identifiable information.

Physical Data

Don't Leave Data Unattended in Transport - When transporting data, never leave your data unattended, even in your locked car.

Lock Data When Unattended - When there will be nobody in the lab or office, make sure that the room is locked, and that the data are locked in a cabinet or drawer. When data is being entered, by primary researchers or otherwise, make sure there is a protocol in place for storing data before and after it has been entered, and make sure data are not 'left out' overnight, or over weekends. It is tempting to resist 'losing your place' when entering a large amount of data, so make sure individuals other than primary researchers are clear that this practice is unsafe.

Electronic Data

Consider Electronic Data Drives as Physical Data - The practices outlined above for physical data should be followed for data stored on transportable electronic data, such as thumb drives, external hard drives, and mobile computers.

Share Files with Google Drive - It is safe to share files between NCSU Google Drive accounts in the cloud. However, files added to Google Drive (e.g. MS Word documents) should be handled differently from documents created on Google Drive. As described above, files not created on Google Docs (e.g. MS Word docs, SPSS data files, etc.) should be password protected. Files created on Google Docs those files should not be downloaded or moved off the NCSU Google Drive server in any way.

What to Put on IRB Form

The institutional review board (IRB) at NC State examines research proposals for evidence that ethical standards will be met. This section describes language that can be used to convey data management precautions you are planning to take. Keep it mind it is extremely unethical and potentially unlawful to claim you will do something and not do it. You should only use examples in this section to the extent they match your existing plan for data management. That is, develop your plan first and then refer to this section to describe your plan to the IRB.

There are a few main points that must be addressed on an IRB form:

1. If data will be linked to the identity of the participants, explain how data will be linked to identifying information -*or*- describe the method for collecting data anonymously (i.e. with no connection to a participant's identifying information.
2. Describe data security and storage
3. Describe data access
4. Describe how and when data will be disposed of

Below are some examples of how to address these questions. The details of these statements, of course, will differ across different experiments.

Identifying Information

- "Names appear on the Experimetrix system for those who sign up for the study, with participant contact information. Anonymity will be ensured in the informed consent process and in the assignment of ID numbers throughout the course of the study."
- "Participants from PSY 200 will not be asked their names on the survey. They will be asked to email the code that they get at the end of the survey to the lab email address in order to get credits."
- "No identifying information will be collected"
- "All data obtained from the study will be coded and stored with participant ID numbers only."
- "Participants will be assigned participant I.D. numbers to insure that no direct personal information will be stored with their data. No IP addresses will be logged."

When using software to collect data with ID's, such as PsychoPY

- "The software used in the experiment will code each participant's scores to a participant number that has no link to participant's identities."

When using Mechanical Turk

- "No identifying information will be collected. Participants' Mechanical Turk ID's will not be collected or stored. Instead workers will be paid using an 8-digit confirmation code displayed at the termination of the survey."
- "MTurk participants are only identified by worker ID number. This worker ID will not be directly associated with the survey. At the end of the survey, MTurk workers get a randomly generated code. They will enter their code to MTurk to receive compensation.

Security and Storage

- "Data will be recorded in Qualtrics (a secure, password-protected website) and stored in a password locked database."
- "All data will be coded into SPSS files, and stored in password protected files on a password protected computer in a locked, limited access lab at 635 Poe Hall."
- **"**The PI will keep a list of of code numbers and student names in a password protected file, stored on a password-protected computer in a locked office."
- "Data will be exported from a password protected account to a password protected computer."
- "All data obtained from the study will be coded and stored with participant ID numbers only on a password protected computer."
- "After completion, all data will be stored in a locked filing cabinet in POE 738H."

When Recording Video Data

- **"**Video data will be taken by a web camera that is mounted in such a way that only the robot and the robot's path will be visible. The participant will not be in the camera footage and if they happen to move unexpectedly into view, the footage for that trial will be permanently deleted. The footage files for each trial will be stored on a DVD and then placed in a locked filing cabinet."

When using Mechanical Turk

- "The PI will keep a master list of code numbers and worker ID numbers. This list will be stored in a separate password-protected database in a password protected computer until all workers have been paid."

Data Access

- "Only the principal investigator and the faculty sponsor and selected research associates will have access to the data."

Disposing of Data

- **"**After completion of data collection, all data will be stored on a password protected computer for secondary analyses." (No identifiers)

Using Mechanical Turk

- "After all workers have been paid the (identifiers) database will be destroyed."
- "Once the study has been concluded the results in the password protected account will be taken offline and deleted from that server. The results will continue to reside on the computer of the principal investigator until deemed no longer relevant. This will be defined as a decade after the publication of the study or conclusion of data collection, whichever comes first."
- Report whether and how identifying information will be recorded and associated with data. If it will not be associated, report that it will be stored anonymously.

Notes on Physical Data

Physical materials should be stored in a locked, limited access lab or office, and in a locked file cabinet. Be sure to address issues of data transportation and storage that are unique to physical data in the IRB application. These issues include physically moving data, especially if collected off-site, and the storage and creation of physical copies of data, including informed consent forms.

Notes on Electronic Data

All electronic data should be password protected, and stored on a password protected drive in a locked, limited access lab or office. Any shared data will be accessed through NCSU's Google Drive, since these servers are located in the US, minimizing the likelihood that they will be compromised. Make sure to use NCSU's Google Drive, and not a personal account.

Additional Help

Have questions about this material? Contact the Director of METRC (the Media, Education and Technology Resource Center) who is currently Bethany Smith (bethany_smith@ncsu.edu).

If you need to create a data management plan (DMP) for a grant application or another service, see http://www.lib.ncsu.edu/guides/datamanagement.

Appendix A: What Makes a Strong Password

Password	**Time to Crack**	**Characteristics**
password	instant	common password
fido	instant	4 characters, only letters
athena22	11 minutes	8 characters, letters and numbers
bwIdIpDE4?	96 years	10 characters, upper- and lower-case letters and numbers, special character, no dictionary words (initialism for “but when I do I prefer Dos Equis”)
IdadbbwIdIpDE$4	157 billion years	15 characters, upper- and lower-case letters and numbers, special character, no dictionary words (initialism for “I don’t always drink beer, but when I do I prefer Dos Equis”)

As computer technology advances, password cracking will become easier and easier, so it is important to have a stronger password than you think you need to survive the future. From this table, it is clear to see the following guidelines are good to follow:

- Have at least 10 characters.
- Use all four character types: lower-case letters, upper-case letters, numbers, and symbols.
- Avoid dictionary words longer than two characters.