HUMBER
Research Ethics Board

Data Privacy, Storage and Disposal

Guidelines

The following is a set of guidelines that should help researchers design the handling, storing and disposing of data as they carry out their research projects. While each project is different and no method is 100% secure, following these principles and practices should make unfortunate events related to data far less likely, protecting the integrity of the project.

Tri-Council policies on the ethical conduct of research do not prescribe a time period for storing research-related data. Depending on the nature of the project, researchers will determine a reasonable or necessary length of time for which they will need access to the information collected. Regulatory standards may exist for documents held for audit purposes.

- On the REB application, a clear distinction should be made between **electronic** and **hard copy** data being collected. In each section, ensure you detail what will happen to each electronic and hard copies of information.
- Whenever possible, **de-identify** the information you collect: while the consent forms will need to identify the participant by name, you can create a coding sheet to identify surveys or other research instruments and then **separate** the identified from the de-identified information. For example, once you assign codes on signed consent forms and to surveys answered by the participants, immediately put the consent forms away and handle only the coded survey forms.
- If possible, refrain from transferring raw data over the internet. If you must transfer data, ensure the transfer is **encrypted.** For how to do this, please consult the *Electronic Security Guidelines* below.
- **Delete any emails** that identify your participants.
- Create a computer **password** that is at least 13 characters. Short passwords can be broken into in minutes. For more information, please consult the *Electronic Security Guidelines* below.
- When **in transit**, store all paper documents in closed folders or envelopes. Lock your computer.
- **Do not read** research data in public (on the train, bus, in cafes or busy public spaces). It is easy for passers-by to read potentially confidential information.
- **Do not leave** your computer, research documents and other data unattended. Lock your office. When you leave, put all data in a locked cabinet and log off the computer.
- Files stored on **USB keys** or **external drives should be encrypted** since these devices can be easily lost or stolen.
- A password on your laptop is minimal protection**. Files should be encrypted** to ensure that they cannot be read if the laptop is stolen.

- **Shred hard copies** using secure shredding services or deposit them into secured (locked) bins that are subsequently taken for secure shredding. Colleges and universities should have secure shredding facilities or services available to their teaching and research staff.

HUMBER
Research Ethics Board

Electronic Security Guidelines

Overview

Electronic security involves protecting the confidentiality and integrity of documents. Confidentiality ensures that others cannot read the documents while integrity ensures that other cannot tamper with the documents.

The degree of effort required to protect your documents is directly proportional to who is trying to read them. If you are in the espionage business, you will go to far greater lengths to protect your data than a researcher will. Researchers need to protect their data from unsophisticated attacks by the general public or more sophisticated attacks by thieves who are after other data and just happen to access research data.

Most data loss and theft is a result of physical theft, not electronic theft. You are working on something, go to the washroom for a minute and leave your door open. Someone comes along and reads what is on your desk. You have a working version and you throw it in the garbage rather than shredding it. Someone wants to work on a job at home, takes some documents home and does not dispose of them properly. Locking doors and shredding documents are the first steps to achieving security.

Laptop or Desktop Security

Most Humber laptops are password protected. This provides a minimal level of security. If the laptop was stolen, anyone could remove the hard drive, insert it in another laptop and read it. This is why it is important to encrypt the documents rather than relying on a password protected computer. If your data is stored on a USB key or external hard drive, a password protected laptop provides no security since these devices could simply be plugged into another computer and read.

A few years ago, the Data Encryption Standard (DES) was broken by someone using about 10 PC's in 4 months. You should avoid using DES or its variants like Triple DES. Encryption algorithms such as RSA, Blowfish, Twofish or AES are currently considered trustworthy. Some of these algorithms are public domain and available in free software.

An encrypted computer is a computer with its entire hard drive encrypted. This can be done by add-on software or by purchasing an encrypted drive from a manufacturer. This will provide a high-level of data security, but might involve additional costs.

Many people simply delete a file from their computer and assume it is gone. Sadly, it is still there. Hard drives have a table of contents and the file is removed from the contents table. The file contents are not removed, since that would take additional time. It is marked for reuse and will, eventually, be reused. Until it is reused, any good programmer with access to the hard drive can retrieve the file. If it is unencrypted, it is immediately readable.

To safely remove an unencrypted file, you should remove the file and then do a low-level format of the hard drive or USB key. This will write zeroes over the entire file system, erasing everything on the disk.

This is only practical on external hard drives and USB drives since if you did it on the drive inside a computer the drive would need everything re-installed.

The take-away from this is that **encrypting files is the best way to ensure their security** and that a reliable encryption algorithm should be chosen.

Server Security

Many researchers store data on servers, either at Humber College or elsewhere on the internet. They feel these servers have taken care of security for them and there is nothing to worry about. Let's consider some of the scenarios and examine how secure they are.

Files are sometimes stored on shared drives at Humber College. There might be an entire department or group of people who have access to these files. If you have had a special directory created with access limited to only the research team, then you have done the first step. However, an administrator in ITS can read these files. They likely have no reason to, but they do have the capability. If you want to protect this data, you need to encrypt it before it is stored on the shared file system.

Storing data outside of Humber College presents its own set of challenges. You need to be concerned with whether the data is stored securely, where it is stored, and whether it is transported across the internet securely.

Many people use Fluid Survey or Survey Monkey to gather data. Both these websites have security policies listed that describe the physical security on their data centres and that encrypted transport across the internet is available. For Survey Monkey, encrypted transport is automatic, for Fluid Surveys it is an option which must be enabled. Whatever survey software you use you should read their security policy, ensure they secure their storage and transport. For transport security, look for Transport Level Security (TLS) or Secure Socket Layer (SSL) and make sure these are enabled for your survey.

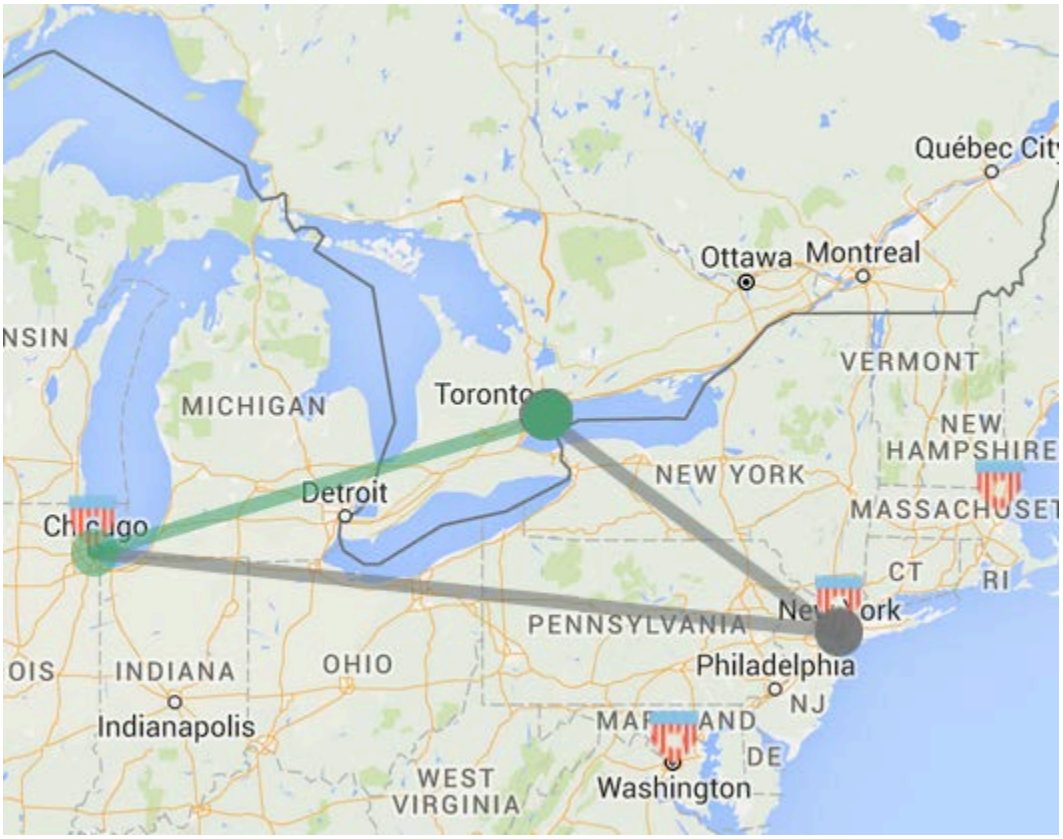
The government of the United States has stated that they have the right to access any information on any computer in the US or in transit through the US. They are looking for terrorist activity and will likely have no interest in your data. However, if you are doing surveys on people's attitudes towards terrorism, surveys related to crime or possibly even political parties in the US, you might attract their attention. For this reason, many people prefer to store data outside the USA and this is required for most research in Canada.

So, you have selected a server based in Canada, located deep inside a mountain, protected by 3 foot thick steel doors and guarded 24/7 by a team of heavily armed mercenaries. Sounds good! Or is it?

You set up encrypted transport using SSL and the people doing your survey are filling it out on Canadian Computers. What could go wrong? The answer lies in how the internet routes traffic. Computers fail for many reasons all the time. Thus, the internet routers dynamically change the routes all the time to go around crashed or powered off computers and get the data to its destination in the most efficient manner. At least that was the original intention. There is a command called traceroute which will show the exact path a message takes between two computers. The following map from https://www.ixmaps.ca/tour.php shows a traceroute from a Toronto location to a nearby location.

HUMBER
Research Ethics Board

Apparently, the shortest distance from Toronto to Toronto is via New York and Chicago. Now, New York and Chicago are both suspected to have large data centres operated the National Security Agency (NSA) of the US. Could they be looking at your data? The answer is: “No one knows.” So, even though all your data is stored in Canada and being transported to a Canadian destination, it is going through the US, which has the right to read it.



Your data is encrypted using SSL, which is reasonably secure. Could the NSA read it? Maybe. They were caught altering code in some network cards so that they could predict the SSL keys generated. There has also been situations where the operating systems were compromised so data sent by them could be read.

OK, but I am just a researcher! How am I going to protect my data? You stored it on a Canadian server that promises security and used encrypted transport. This is all you can do, unless you have a budget in the billions. Your data is probably not safe from national governments but the only solution is to not use the internet and have the data transported on encrypted USB via a security company like Brinks. Sadly, that solution is beyond the budgets of most research projects.

HUMBER
Research Ethics Board

Passwords

Passwords are the most common way of securing data and one of the weakest links in the security chain. The most common problems are passwords which are poorly chosen and passwords which are managed incorrectly.

Many people use common words as passwords. I see many passwords which are the name of the project with the year tacked on the end. These passwords are very easy to guess. Another common password includes one or more words from dictionaries. To make better passwords, we need to understand how people can crack passwords.

The simplest way of cracking a password is to take a guess based on what you know about the project or person. In far too many cases, this is successful. The second approach is to use a dictionary attack. This simply tries every word in the dictionary to see if you used any real words. The third approach is the brute force attack which tries every combination of symbols until it gets the one that opens the lock. The only way to defeat this is to use longer passwords, which take longer to find.

The fourth approach only applies to logins on computers. Often hackers get into a machine and steal the encrypted password file. This should not make any difference since they are not the passwords but encrypted versions of them. Unfortunately, someone has pre-calculated the encrypted version of all possible passwords for strings up to 12 characters in length. This allows them to simply look up your encrypted password in a table and get a password which will generate that encrypted value and open your account. This can be done in about 5 minutes. **The only defense is passwords longer than 12 characters.**

Many people use the same passwords in multiple places. The same password is used at work and at many websites on the internet. Websites are now hacked regularly with the result that many people have their passwords stolen. The way to protect against this is to use different passwords everywhere.

So, you end up with 47 different 13 character passwords that do not contain words. This is now impossible to remember, so the solution is a password safe. This is a program that stores your login names and passwords in encrypted form and protects them all with a master password. There are many password safes available. 1Password and Wallet Guard are a couple of the common ones for the iPhone. Android will have some as well. Using these, the password you need is secure and no further away than your phone.