MREB Data Storage and Security Tools: Cover Page
Last Revised: 2018-06-19
Version: 0.1

This document is a collection of four complimentary data storage and security tools, which are intended to guide researchers in determining the data storage and security procedures for handling their research data.

Data Storage & Security Guide: This guide provides an overview of the main issues in the storage and security of research data. It also lists data storage services available at McMaster.

Frequently Asked Questions: The FAQ answers some of the common questions about data storage and security – including questions on encryption, cloud storage, and sharing data with co-investigators at other institutions.

Research Data Management Matrix: The matrix provides data storage and security guidance for different risk levels of data – low, medium, and high.

Data and Information Storage Glossary: Defines some of the data storage and security terms used in the other documents (e.g. cloud service, portable storage device, anonymous data, etc.).

MREB Data Storage & Security Guide

Last Revised: 2018-06-19
Version: 0.1

The following Data Storage & Security Guide provides information on collecting and storing data for research involving human participants. This guide refers to research data (or 'data'), which is "data that are used as primary sources to support technical or scientific enquiry, research, scholarship, or artistic activity, and that are used as evidence in the research process and/or are commonly accepted in the research community as necessary to validate research findings and results" (CASRAI, 2018).

Why is data storage relevant?

Creating a data storage plan is essential to research projects at any scale. A thorough and purposeful plan ensures that data are easily retrievable in the short and long term, facilitates appropriate access, and prevents potential data loss.

Why is data security relevant for researchers?

Researchers often collect personal, confidential or sensitive data. It is important to ensure that appropriate measures are established to prevent unauthorized access to the information to protect participants, researchers, and the University.

Regulation of Data Storage & Security

In June 2016, The Tri-Agency released a statement of principles on digital data management. It re-emphasized the importance for data to be stored using secure software and formats, and for data to be stored in a manner that enables preservation of and access to the data after the research project is completed.

In Ontario, the Personal Health Information Protection Act (PHIPA) 2004 regulates the collection, use, and disclosure of personal health information, stressing the importance for appropriate measures of data storage and security relative to the level of data sensitivity. All researchers must comply with this legislation.

Considerations for Data Storage

- Inform study participants of the data storage procedures during the consent process
- Always attempt to collect the least amount of specific information possible (e.g. initials instead of name, age instead of date of birth). Data should be anonymized as soon as possible given the context of the study (e.g. for a longitudinal study anonymization is not feasible).
- Consider the longevity of the file formats when storing data. Store data using more basic file formats (.csv, .txt) rather than Microsoft Office or Adobe Suites

- Provide instructions on how to interpret the data in case data are used in future projects (e.g., a code book, readme file, structured metadata etc.)
- When backing up your data, use the *3-2-1 rule*
    - Save 3 copies of the data, stored in 2 different storage mediums, and have 1 copy off-site (using a trusted service provider)

Considerations for Data Security

Below are data security considerations for three types of security: physical, administrative, and technical.

Physical Security:

- Control access to buildings, rooms, cabinets where data, computers, media or hardcopy materials are held
- Log the removal of, and access to, media or hard copy material in storage rooms
- Transport sensitive data only under exceptional circumstances, even for repair purposes. *For example, giving a failed hard drive containing sensitive data to a computer manufacturer may cause a breach of security*

Administrative Security:

- Develop organizational rules about who has access to research data
- Impose non-disclosure agreements for managers or users of confidential data
- Do not store sensitive data such as those containing personal information on servers or computers connected to an external network, particularly servers that host internet services
- Apply firewall protection and security-related upgrades and patches to operating systems to avoid viruses and malicious code

Technical Security:

- Ensure that all computer systems storing confidential data are password protected and that data are stored on encrypted drives wherever possible
- Ensure computer software is up-to-date, including anti-virus software
- Implement password protection and controlled access to data files. *For example, 'no access', 'read-only', 'read and write' or 'administrator-only' permission*
- Control access to files, folders or entire hard drives
- Anonymize personal information that is collected
- De-identify personal information upon data collection
- Do not send personal or confidential data via email

Prior to selecting the data storage and security measure for a specific project, consider the following questions:

- What kind of storage will the research data require? *For example, will you need a physical space to store consent forms? How much digital storage will you need to hold all project files?*
- What types of data are you collecting for the project? Will different levels of security will be necessary for this type of data? *For example, if you are collecting personal health information, how will you ensure that the data are kept secure? Will the data be anonymized? Will the data be encrypted, with identifying information removed? Where will the data be stored?*
- How will the project be managed, if multiple collaborators are involved? *Will the data storage require sharing capabilities? How will you ensure data are only authorized to those who have permission?*
- How will the data be backed up? How long will the data need to be stored?

Appendix I (McMaster University File Storage and Backup Services Matrix) contains a list of data storage options offered at McMaster University.

NOTE: While Cloud-based storage (Dropbox) is easy to use, there are limitations to the data security. Some of these Cloud-based storages are based overseas (e.g. Dropbox server is housed in the US), meaning governments may be able to acquire access to the data upon request of the company. While there is no specific legislation in Ontario or McMaster University banning the use of Cloud-based storage, best practice suggests that commercial storage systems, like Dropbox, should be avoided for confidential data. Alternatively, McMaster does host Cloud-based storage options which include, MacDrive, MacDrop, MCloud, and Dataverse.

Additional Resources

Regulatory Documents on Data Management

- Tri-Agency Statement of Principles on Digital Data Management
- Tri-Council Policy Statement (TCPS) II
- Personal Health Information Protection Act, 2004

Data Management Support at McMaster University

- Research Data Management (RDM)
- MacDATA Institute
- Research & High Performance Computing Support

External Resources

- Portage Network
- UK Data Service
- DMP Tool

For any questions regarding the contents of this document, please contact Mark Lee.

Appendix I - McMaster University Campus Wide File Storage and Backup Services Matrix

Last Revised: 2018-05-11

	RHPCS - Backup Services (Research & High-Performance Computing Support)	RHPCS - Hosted Server Packages	MacDrive	Microsoft OneDrive / Teams
Storage Quota	1 TB; more available for fee	1 TB; more available for fee	300 GB per account	1 TB per account; up to 5 TB by request
Rates / cost	$500 / yr + one time set up fee ($125 / machine) Additional space: $150 / TB / yr Restore services: $125 / hour	$500 - $1000 / yr Setup fee: $500 - $1000 Additional space: $150 / TB / yr	No cost to users	No cost to users
Backups / versioning	Nightly, 14-day rotating cycle; Restore services through RHPCS	Nightly, 14-day rotating cycle; Restore services through RHPCS Nextcloud sync service available.	Ongoing real-time sync 4-month version history Full Library restore through UTS	Ongoing real-time sync Unlimited version history (?)
Who can use this service?	Any subscribing users or research group	Any subscribing users or research group	McMaster Faculty and Staff Graduate students can obtain zero-quota accounts	All McMaster faculty, staff and students
Server location	A.N. Bourns building	A.N. Bourns building	Replicated clusters in Gilmour Hall and JHE	**OneDrive**: Canadian servers **Teams**: Soon in Canadian servers only

Other notes			Supports encrypted libraries, file and directory sharing, Desktop client, web interface	Supports file and directory sharing, Desktop client, web interface
More info	RHPCS Rates	RHPCS Rates	MacDrive MacDrive Documentation	Portal Portal Documentation

Note that individual Departments and Faculties may also offer network storage solutions that are suitable for research data. Please consult with your local IT provider for more information

Frequently Asked Questions

Last Revised: 2018-06-19
Version: 0.1

The following is a list of common questions that MREB staff frequently receive about collecting and storing data for research involving human participants:

1. Is a password-protected laptop a secure place to store my data?
2. How long can/should I keep my data?
3. What is encryption? When and how should I encrypt my data?
4. What is Cloud storage? Is it safe to store my data in the Cloud?
5. Is it safe to store my data on mobile devices such as cell phones or USB keys?
6. What is the best way to share data with my co-investigators at other institutions?
7. What online survey software should I use?
8. What is the difference between wireless and wired internet connections? Is one safer?

NOTE: Many of the answers to these questions will depend on the level of risk associated with the data you have collected. To learn more about the risk level of your data see the *Research Data Management Matrix*.

1. Is a password protected laptop a secure place to store my data?

Using a password-protected laptop is one of the most common ways to collect and store data. While this method might be secure enough for low-risk data, it is not secure enough for medium and high-risk data. If a laptop must be used for medium to high-risk data, then the files or the hard drive must be encrypted.

Even if you are collecting low-risk data, there are ways to make storage on a password-protected laptop safer. For example, encrypt your hard drive, use anti-virus software and anti-malware regularly, update your computer as soon as updates are available, and use secure browser plugins. Perhaps most importantly, regularly back up and secure your data (see the Data Storage & Security Guide).

The use of physical security is also recommended for the storage of your data. Avoid common situations where your laptop may get stolen such as leaving it in a vehicle or public place unattended. If the data being collected are not low-risk data, additional steps should be taken to protect it including storing the data on a password-protected and encrypted desktop in a locked office, storing the data on a password-protected server and so on.

Encryption is recommended for low-risk data and required for medium to high-risk data. The use of portable storage is not recommended as these small devices may be more easily lost or stolen. If a portable storage device cannot be avoided, it must be encrypted regardless of the risk level of the data.

2. How long can/should I keep my data?

The short answer: It depends on your data! But here are some things to consider:

- Researchers need to find a balance between the risks and benefits of retaining or deleting their data, paying special attention to how identifiable or risky the data are. To learn more about the risk level of your data see the *Research Data Management Matrix*.

- Researchers need a plan to manage their data securely on an ongoing basis. Ask yourself: what is my data management plan if I move on from this institution, this field, or this career?

- While it is preferred that you delete your data, it is not inevitable. If you have a good reason to keep your data, and a robust data management plan that describes your plans for how you will steward the data in the future, you may not have to delete it. Depending on the project, it is important to de-identify data at the earliest possible step in the acquisition/analysis process. Note that some scholarly journals, organizations, and scholarly societies have guidelines about data retention for publication.

- Pay attention to where you are in the world! Canada's TCPS-2 and the UK's *Data Protection Act* do "not set out any specific minimum or maximum periods for retaining personal data". (7) However, the U.S. *Office for Human Research Protections* requires research "records frequently held by investigators" such as consent forms or transcribed anonymized interviews to be held "for at least three years after completion of the research." (8)

- Finally, in the case that "research is supported by a contract with or a grant to the University that includes specific provisions regarding… data retention the provisions of that agreement will take precedence". (9).

3. What is encryption? When and how should I encrypt my data?

Encryption is a method of protecting your data so that only you, or someone you authorize, can access it. The *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* states that "in general, identifiable data obtained through research that is kept on a computer and connected to the Internet should be encrypted." (10) There are a couple of different methods of encrypting your data such as encrypting individual files and encrypting your drive (or volume). Both of these methods have their pros and cons.

Encrypting Individual Files

Encrypting only select files such as those that are research-related, or those that contain identifying information, keeps your data safe without any extra complications. Programs such as MS Office and Adobe, offer file-level encryption natively. These programs are recommended when there are few files to encrypt.

Encrypting Your Drive (or volume)

Encrypting your entire drive or the volume on which your data are stored protects against anyone accessing any of your data without your authorization. Encrypting your drive or volume is often more convenient and less prone to human error as all files are encrypted automatically.

For more information on **Encryption** and password protection, **Mobile Devices** (cell phones, laptops, USB keys), **Wi-Fi Security** (on public networks, and when traveling to other institutions), **Passwords** (best practices), **IT Security** (contacts, anti-virus software, spam/phishing emails, network drive security), and **Resources** (confidential waste bins, privacy impact assessments, other resources), please visit the McMaster Privacy Tools and Resource page.

4. What are Cloud services? Is it safe to store, transfer or share my data using the Cloud?

Cloud services store and share data by keeping it on remote servers accessed from the internet. For our purposes we can think of Cloud services as being internal (i.e. provided by or endorsed by McMaster University) and external (i.e. provided by someone else). You may also see the terms public and private but these can be somewhat misleading. While any use of Cloud services comes with some inherent risk, the risks for internal and external Cloud services servers are somewhat different. The most important differences relate to where data are stored, and how users are authenticated. With external Cloud storage, data are stored on servers that could be located anywhere in the world, and thus subject to that country's laws. With internal Cloud services, your data are stored on locally controlled servers, in our case located at McMaster University, on servers that are run by trusted partners. Which Cloud services you should use will depend on the risk level of your data. To learn more about the risk level of your data see the *Research Data Management Matrix*.

External Cloud Services

- *Example: Dropbox & iCloud*- McMaster has no agreement in place with these services. Users must use personal accounts, unaffiliated with the University, to store their data. MREB recommends against using these types of services for identifiable data.
- *Example: OneDrive*- McMaster currently has agreements in place. These services are only recommended for low risk data. Data considered to have a medium-risk must be encrypted before uploading.

Internal, McMaster Endorsed Cloud Services

McMaster University has 4 Cloud services that it endorses (MCloud, MacDrive, Dataverse, and MacDrop). MCloud, MacDrive, and MacDrop are all very similar. They are all EFSS (Enterprise File Storage and Synchronization) solutions, similar in function to Dropbox, and house the data on-site. It is recommended that any data uploaded to these services are also encrypted. Dataverse is an open source research data repository system for archiving, describing, and publishing datasets to enable their long-term preservation and reuse. McMaster faculty, staff, and students have access to Scholars Portal Dataverse, a service provided by the Ontario Council of University Libraries (OCUL).

- Example, MCloud: MacDrive and MCloud offer the same kind of functionality but MCloud is tied in to research workgroup servers and can handle large file quotas routinely.
- Example, MacDrive: MacDrive and MCloud offer the same kind of functionality but MacDrive is a standalone system and typically deals with smaller quotas. However, it has the advantage of being tied into the university's central authentication system and is a centrally supported service.
- Example, MacDrop: Similar to MCloud and MacDrive, MacDrop differs in that it is managed by the Computer Services Unit, the IT group within the Faculty of Health Science (FHS). The only users are within FHS.

5. Is it safe to store my data on portable storage devices such as cell phones or USB keys?

Data should generally not be stored on any type of portable storage device (other than a laptop), no matter what the risk level is associated with the data. The biggest risk associated with all portable storage devices is that they can be easily lost or stolen. Portable storage devices that have an internet connection (such as a cell phone) have additional risks that a device that does not have an internet connection (such as a USB key) would not have. If data must be stored on a portable storage device, it must be stored in an encrypted format.

For internet-connected portable storage devices:

Pros: Collecting data on an internet connected portable storage device such as a cell phone can be a good choice because the technology is ubiquitous, familiar and convenient, it is fast, accurate and portable, and requires low power at a relatively low cost to the researcher. (17)

Cons: When data are stored on portable storage devices "it can potentially be stolen or improperly accessed – the same holds true during data transmission." (18) However, "the use of encryption at both the device level and during transmission can greatly mitigate such risks." (19).

For non-connected portable storage devices:

Pros: Non-connected portable storage devices do not have the same vulnerabilities as internet-connected portable storage devices, while still providing storage and data transfer options.

Cons: Given that they are not connected to the Internet data transfers can be less convenient. Additionally, some portable storage devices are easily corruptible and not built for long-term storage, for example inexpensive flash drives.

6. What is the best way to share data with my co-investigators at other institutions?

Before you share any data collected from human participants in any way, the key is to render that data as low-risk as possible--for instance, by de-identifying it. If working with identifiable data, ideally those collecting the research would remove all identifying personal information before the data was shared with research partners at other institutions; it is important to be mindful of legislation that may be applicable to your co-investigators (e.g. the US Patriot Act / Domestic Security Enhancement Act). If you intend to share data, you require a data sharing agreement and this must be described in your ethics application.

Select a method of communicating your data that is consistent with its risk level. To learn more about the risk level of your data see the Research Data Management Matrix.

Low-Risk Data: All McMaster hosted Cloud services and McMaster email.
Medium-Risk Data: Encrypted and password-protected files can be shared via McMaster approved Cloud services, encrypted files via OneDrive, and by McMaster email.
High-Risk Data: Restricted data should be shared hand to hand on a password-protected and encrypted data storage device. Encrypted and password-protected files may be shared via McMaster approved Cloud services if approved by the MREB. Maintaining ethical high-risk data transfer between institutions may require individualized strategies. Contact the MREB more information.

7. What online survey software should I use?

McMaster University, through the Office of the VP Research, provides a survey service called LimeSurvey. (20) LimeSurvey allows users to create "online question-and-answer surveys that can work for tens to thousands of participants without much effort. The online survey software itself is self-guiding for the respondents who are participating." (21) The MREB LimeSurvey services stores data locally and has templates in place that were designed with the principles of the TCPS statement on ethics in mind. For this reason, we

recommend all researchers avail themselves of this service rather than using alternatives which may have higher inherent risk associated with them.

8. What is the difference between wireless and wired internet connections? Is one safer?

When it comes to connectivity, computers at McMaster University fall into 3 categories: computers that connect to the Internet wirelessly, computers that connect via wired networks, and computers with no internet connection at all.

When it comes to wired and wireless connections, the risks involved are relatively the same. Risks on the endpoint are much greater than those on the network (i.e., a user more likely to contract malware). The difference in security between wired and wireless connections at McMaster is marginal. In fact, it is more important to maintain a clean computer.

Notes

1. Tri-Council (CIHR, NSERC and SSHRC), "Tri-Agency Statement of Principles on Digital Data Management." Government of Canada, Innovation, Science and Economic Development Canada, Office of the Deputy Minister, Communications and Marketing Branch. December 21, 2016. Accessed August 28, 2017. http://www.science.gc.ca/default.asp?lang=En&n=83F7624E-1.
2. Tri-Council, *Digital Data Management*.
3. Ibid.
4. Ibid.
5. Oxford University. "University of Oxford Policy on the management of research data and records." Research Data Oxford. February 04, 2016. Accessed August 28, 2017. http://researchdata.ox.ac.uk/university-of-oxford-policy-on-the-management-of-research-data-and-records/.
6. Oxford University, *Research Data Oxford*.
7. Government of Canada Interagency Advisory Panel on Research Ethics. *"Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans" (TCPS 2)* Pre.ethics.gc.ca. November 18, 2016. Accessed August 28, 2017. http://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/chapter5-chapitre5/. Article 5.3.
8. Information Commissioner's Office. "Retaining personal data (Principle 5)." Data Protection Act: Data Protection Principles. Accessed August 28, 2017. https://ico.org.uk/for-organisations/guide-to-data-protection/principle-5-retention/.
9. Office for Human Research Protections. "Investigator Responsibilities FAQs." Department of Health and Human Services. Accessed August 28, 2017. https://www.hhs.gov/ohrp/regulations-and-policy/guidance/faq/investigator-responsibilities/index.html. 45 CFR 46.115(b).
10. Interagency Advisory Panel on Research Ethics, *TCPS 2*, Article 5.3.
11. Techopedia, "What is Cloud Storage? - Definition from Techopedia." Techopedia.com. Accessed August 28, 2017. https://www.techopedia.com/definition/26535/cloud-storage.
12. McMaster Library Maps, Data & GIS, "Frequently Asked Questions." McMaster University Library, Hamilton, Ontario, Canada. July 12, 2016. Accessed August 28, 2017. https://library.mcmaster.ca/rdm/faq#storagesecurity.
13. University of Michigan Safe Computing, "Safely Use the Cloud." safecomputing.umich.edu. Accessed August 28, 2017. https://www.safecomputing.umich.edu/protect-the-u/protect-your-unit/safely-use-the-cloud.

14. University of British Columbia Okanagan Behavioral Research Ethics Board, “Research Data Storage and Security.” Office of Research Services. Accessed August 28, 2017.
15. Nextcloud, "Nextcloud 12 User Manual." Nextcloud 12 User Manual Introduction — Nextcloud 12 User Manual Documentation. Accessed August 28, 2017. https://docs.nextcloud.com/server/12/user_manual/.
16. Scholar’s Portal Dataverse, "Guides: Scholars Portal Dataverse Guide: About Dataverse." About Dataverse - Scholars Portal Dataverse Guide - Guides at Scholars Portal. Accessed August 28, 2017. http://guides.scholarsportal.info/dataverse.
17. Courtney Sheppard, "MacDrive documentation v1.3." Google Docs. Accessed August 28, 2017. https://docs.google.com/document/d/16DCGEmi-5AVi1HA4gFk6Dq3JIYsGCvIUzWlFHdDKvvo/edit#.
18. Trucano, Michael. "Using mobile phones in data collection: Opportunities, issues and challenges." Edutech. April 18, 2014. Accessed August 28, 2017. http://blogs.worldbank.org/edutech/using-mobile-phones-data-collection-opportunities-issues-and-challenges.
19. Trucano, *Mobile Phones*.
20. McMaster Research Ethics Board. "Ethics Compliant McMaster Survey Service.". Accessed August 29, 2017. https://reo.mcmaster.ca/limesurvey.
21. LimeSurvey. "LimeSurvey Manual." Accessed August 29, 2017. https://manual.limesurvey.org/.
22. Saint Mary's University Research Ethics Board. “Frequently Asked Questions - Pre-Approval” Accessed August 29, 2017. http://www.smu.ca/research/reb-frequently-asked-questions-pre-approval.html.
23. McMaster University Engineering Computing and Software Wiki."Wi-Fi." Computing and Software Wiki RSS. Accessed August 29, 2017. http://wiki.cas.mcmaster.ca/index.php/Wi-Fi.
24. McMaster University Technology Services. "Access and Services. Accessed August 29, 2017. http://www.mcmaster.ca/uts/network/access.htm.

Works Cited

Government of Canada Interagency Advisory Panel on Research Ethics. *"Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans" (TCPS 2)* Pre.ethics.gc.ca. November 18, 2016. Accessed August 28, 2017. http://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/chapter5-chapitre5/

Information Commissioner's Office. "Retaining personal data (Principle 5)." Data Protection Act: Data Protection Principles. Accessed August 28, 2017. https://ico.org.uk/for-organisations/guide-to-data-protection/principle-5-retention/

LimeSurvey. "LimeSurvey Manual." Accessed August 29, 2017. https://manual.limesurvey.org/.

McMaster University Engineering Computing and Software Wiki."Wi-Fi." Computing and Software Wiki RSS. Accessed August 29, 2017. http://wiki.cas.mcmaster.ca/index.php/Wi-Fi.

McMaster Library Maps, Data & GIS, "Frequently Asked Questions." McMaster University Library, Hamilton, Ontario, Canada. July 12, 2016. Accessed August 28, 2017. https://library.mcmaster.ca/rdm/faq#storagesecurity

McMaster Research Ethics Board. "Ethics Compliant McMaster Survey Service.". Accessed August 29, 2017. https://reo.mcmaster.ca/limesurvey.

McMaster University Technology Services. "Access and Services. Accessed August 29, 2017. http://www.mcmaster.ca/uts/network/access.htm.

Nextcloud, "Nextcloud 12 User Manual." Nextcloud 12 User Manual Introduction — Nextcloud 12 User Manual Documentation. Accessed August 28, 2017. https://docs.nextcloud.com/server/12/user_manual/

Office for Human Research Protections. "Investigator Responsibilities FAQs." Department of Health and Human Services. Accessed August 28, 2017. https://www.hhs.gov/ohrp/regulations-and-policy/guidance/faq/investigator-responsibilities/index.html

Oxford University. "University of Oxford Policy on the management of research data and records." Research Data Oxford. February 04, 2016. Accessed August 28, 2017. http://researchdata.ox.ac.uk/university-of-oxford-policy-on-the-management-of-research-data-and-records/

Saint Mary's University Research Ethics Board. "Frequently Asked Questions - Pre-Approval" Accessed August 29, 2017. http://www.smu.ca/research/reb-frequently-asked-questions-pre-approval.html

Scholar's Portal Dataverse, "Guides: Scholars Portal Dataverse Guide: About Dataverse." About Dataverse - Scholars Portal Dataverse Guide - Guides at Scholars Portal. Accessed August 28, 2017

Sheppard, Courtney Sheppard. "MacDrive documentation v1.3." Google Docs. Accessed August 28, 2017. https://docs.google.com/document/d/16DCGEmi-5AVi1HA4gFk6Dq3JIYsGCvIUzWlFHdDKvvo/edit#

Techopedia, "What is Cloud Storage? - Definition from Techopedia." Techopedia.com. Accessed August 28, 2017. https://www.techopedia.com/definition/26535/cloud-storage

Tri-Council (CIHR, NSERC and SSHRC), "Tri-Agency Statement of Principles on Digital Data Management." Government of Canada, Innovation, Science and Economic Development Canada, Office of the Deputy Minister, Communications and Marketing Branch. December 21, 2016. Accessed August 28, 2017. http://www.science.gc.ca/default.asp?lang=En&n=83F7624E-1

Trucano, Michael. "Using mobile phones in data collection: Opportunities, issues and challenges." Edutech. April 18, 2014. Accessed August 28, 2017. http://blogs.worldbank.org/edutech/using-mobile-phones-data-collection-opportunities-issues-and-challenges

University of British Columbia Okanagan Behavioral Research Ethics Board, “Research Data Storage and Security.” Office of Research Services. Accessed August 28, 2017

University of Michigan Safe Computing, "Safely Use the Cloud." safecomputing.umich.edu. Accessed August 28, 2017. https://www.safecomputing.umich.edu/protect-the-u/protect-your-unit/safely-use-the-cloud

Research Data Management Matrix

Last Revised: 2018-06-19
Version: 0.1

The following Research Data Management Matrix is a guideline from the McMaster Research Ethics Board for collecting and storing data for research involving human participants.

	LOW RISK	MEDIUM RISK	HIGH RISK
TYPES OF DATA	Research data that does not contain any sensitive or identifiable information about individuals, organizations or communities (e.g. data which have been de-identified). NOTE: If in doubt, assume that data are sensitive. Non-sensitive research documentation (e.g. non-confidential protocols and information sheets) Publicly facing information. While public facing information is often considered low-risk, there are cases where informed consent/risk of harm should be closely considered. For example, information regarding racial or ethnic origin could be found on	Research data that may or does contain confidential, sensitive or identifiable information about individuals, organizations, or communities Some sensitive research-related documentation Personally identifiable information De-identified records of compensation Data and research protocols related to private or sensitive intellectual property	Research data that contains highly sensitive information about individuals, organizations, or communities (e.g. information about criminal activity) Personal health information Personal financial information such as banking information, income tax returns Data and research protocols related to highly sensitive intellectual property Identifiable data where disclosure, loss, or unauthorized modification of information may result in significant risk for the research participant including reputational damage, significant professional or

	public facing websites, but in certain study contexts could be considered medium or high risk data.		personal disruption, financial consequences, physical or psychological harm, and legal liability.
EXAMPLES	Completely de-identified or anonymous data Blank consent forms and information sheets Information gathered from a public-facing website	De-identified financial information associated with research payments Identifiable demographic data and/or information about participants' beliefs, opinions, health, etc., that in the context of the study would be considered medium risk.	Depending on the study context, examples of high risk data could include information with regard to racial or ethnic origin; political opinions; religious beliefs or other beliefs of a similar nature; trade union membership; physical or mental health or condition; sexual life; the commission or alleged commission by the data subject of any offence. Video or audio recorded interviews depending on the content Identification keys and signed consent forms
DATA PROTECTION	Research data must always be stored according to protocols approved by the appropriate Research Ethics Board	Collect and store data on password-protected devices, preferably static devices in a secure location such as on a desktop computer in a locked office or an appropriately protected server. Consider encryption where possible. All research data are subject to the TCPS2 which states "identifiable data	Collect and store data on password-protected and encrypted devices. Physical security of the data is required (i.e., stored in a locked office or on a protected server). All Research data are subject to the TCPS2 which states "identifiable

		obtained through research that are kept on a computer and connected to the Internet should be encrypted." See below for more information about secure data storage, access and transfer.	data obtained through research that is kept on a computer and connected to the Internet should be encrypted."
DATA STORAGE	Local hard drive (e.g., C: drive, "My Documents") Removable storage media (e.g., USB drives, portable hard drives, etc.) University hosted file sharing and storage (e.g., UTS hosted shared network drives) Department hosted file sharing and storage (e.g., department shared network drives) University hosted Cloud based storage (e.g., MacDrive, MacDrop, MCloud, Dataverse) University sanctioned Cloud based or third party storage (e.g., One Drive)	A computer that meets the data protection requirements. Public Cloud services (DropBox, iCloud, etc.) for data storage or transfer might be suitable if identifiable data is encrypted. Institutional Cloud services (e.g. MacDrive, MCloud, Dataverse, and MacDrop) might be suitable if specified in the REB protocol. Privacy and security risk are the reasons for preferring internal services over external, particularly those for which there is not an enterprise agreement. Central, departmental and lab file shares that meet data protection requirements and have been identified in the REB protocol.	A computer or external electronic storage device that meets the data protection requirements. Central, departmental and lab file shares that meet data protection requirements and have been identified in the REB protocol. University hosted file sharing and storage (e.g., UTS hosted shared network drives) All considerations above for Confidential are applicable, plus: Must never be stored in any unsanctioned storage location. Must not be shared via email.

	Department sanctioned Cloud based or third party storage (e.g., DropBox for Business) Personal Cloud storage (e.g., OneDrive, Dropbox)		Research data are subject to the TCPS2 which states "identifiable data obtained through research that is kept on a computer and connected to the Internet should be encrypted." All public Cloud services (DropBox, iCloud, OneDrive, etc.) for data storage or transfer are strictly prohibited. Use of private Cloud services for data storage and transfer are subject to the restrictions detailed below.
DATA ACCESS	No special handling required.	Access to confidential information must be restricted to authorized individuals who have been identified in the REB protocol only.	Access to confidential information must be restricted to authorized individuals only who have been identified in the REB protocol. Note for reviewers: Access should be restricted to the fewest number of individuals possible.

DATA TRANSFER	Can be shared via all Cloud services including public Cloud services (OneDrive etc.)	Encrypted and password-protected files can be shared via McMaster email and McMaster approved Cloud services.	Restricted data should be shared using direct system to system encrypted (TLS) transfer instead of portable devices. Files may be shared using properly encrypted, password-protected, expiring links.

MREB Data and Information Storage Glossary

Last Revised: 2018-06-19
Version: 0.1

NOTE: For a generalized and comprehensive research data management glossary, visit *CASRAI*, the standard dictionary of research administration information. Many of the definitions below are drawn from the CASRAI dictionary.

Anonymized Data

Data which has had all identifying information irrevocably stripped out, with the risk of identification of individuals being very low. Note that this is not necessarily the same as de-identified data, which could include the use of a code that would allow re-identification of an individual with the aid of the coding key.

Anonymous Data

Data that never had identifiers associated with it (e.g. anonymous surveys), and risk of identification of individuals is very low.

Cloud Services

A method of storing and sharing data by keeping it on remote servers accessed from the Internet. Cloud services are maintained, operated and managed by a cloud service provider on storage servers. Cloud services can be public or private. Public cloud services include DropBox, iCloud and OneDrive. McMaster University endorsed private cloud services include Dataverse, MacDrive, MacDrop, and MCloud. While any use of cloud services comes with some inherent risk, the risks for public and private servers are different. Some main differences include server location, server control, and attack surface. With public cloud storage, data are stored in servers that could be anywhere in the world, and thus subject to that country's laws. With private cloud services your data are stored in local servers. Private companies control public cloud services and the data that is stored there. Access to data stored in private cloud services is controlled by McMaster University. Finally, public cloud services have sprawling infrastructure with many different points where an unauthorized user could attempt to extract data; in some cases, private services are less open to such attacks.

Cloud Computing

A large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted, virtualized, dynamically- scalable, managed computing power, storage, platforms and services are delivered on demand to external customers over the Internet.

Data Lifecycle

The data lifecycle refers to all of the stages in the existence of data from collection to destruction. A lifecycle view is used to enable active management of the data over time, thus maintaining security, accessibility, and utility.

Data Management Plan (DMP)

A DMP is a formal statement describing how research data will be managed and documented throughout a research project. Almost all DMPs contain the following core elements: metadata, policies for access and sharing, policies for re-use and redistribution, and plans for archiving preservation and destruction. McMaster encourages the use of DMP assistant by Portage, a bilingual tool for preparing DMPs that follows best practices in data stewardship and walks researchers step-by-step through key questions about data management.

Data Security

A description of security measures to protect the data - e.g. will data storage require additional security levels (not on network, encryption etc.).

Data Sharing

The practice of making data available for reuse. This may be done, for example, by depositing the data in a repository or through data publication.

De-Identification

The act of minimally perturbing individual-level data to decrease the probability of discovering an individual's identity. It involves masking direct identifiers (e.g., name, phone number, address) as well as transforming indirect identifiers that could be used alone or in combination to-identify an individual (e.g., birth dates, geographic details, dates of key events). If done correctly, de-identification is a defensible, repeatable, and auditable process that consistently provides assurance, based on generally accepted and repeatable statistical methodologies, so that there is a very small risk of re-identification of any data that are released.

Deletion

The process of destroying data stored on hard disks, mobile devices and other forms of electronic media so that it is completely unreadable and cannot be accessed or used.

Encryption

Encryption is a method of encoding your data so that only you, or someone you authorize, can access it. The Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans states that "in general, identifiable data obtained through research that is kept on a computer and connected to the Internet should be encrypted." There are a couple of different methods for encrypting your data including encrypting individual files and encrypting entire devices. Both options have pros and cons.

High-Risk Data

High-risk data requires very strong controls against unauthorized disclosure, loss, and modification. Disclosure, loss, or unauthorized modification of restricted information may result in significant risk for the research subject and the researcher including reputational damage, significant professional or personal disruption, financial consequences and legal liability. Depending on the study context, examples of high-risk data could include information with regard to racial or ethnic origin; political opinions; religious beliefs or other beliefs of a similar nature; trade union membership; physical or mental health or condition; sexual life; the commission or alleged commission by the data subject of any offence or criminal activity. Other examples of high-risk data may include Personally Identifiable Information (PII) (where a breach of confidentiality would carry a high risk for research participants), Personal Health Information (PHI) and credit card information (PCI). *See the Research Data Management Matrix for information handling guidance.*

Low-Risk Data

Low-risk data requires controls against unauthorized modification for the sake of data integrity rather than to prevent risk to researchers or research participants. Examples of unrestricted information may include completely de-identified or anonymous data, blank consent forms and information sheets, and information gathered from a public-facing website. *See the Research Data Management Matrix for information handling guidance.*

Medium-Risk Data

Medium-risk data requires strong controls against unauthorized disclosure, loss, and modification. Disclosure, loss, or unauthorized modification of confidential information may result in putting research subjects at risk. *See the Research Data Management Matrix for information handling guidance.*

Metadata

Best described as "data about data. Metadata define and describe the characteristics of other data, used to improve both understanding of data and data-related processes. Business metadata includes the names and business definitions of subject areas, entities and attributes, attribute data types and other attribute properties, range descriptions, valid domain values and their definitions. Technical metadata include physical database table and column names, column properties, and the properties of other database objects, including how data are stored. Process metadata are data that define and describe the characteristics of other system elements (processes, business rules, programs, jobs, tools, etc.). Data stewardship metadata are data about data stewards, stewardship processes and responsibility assignments.

Non-Identifiable Data

Data that cannot lead to the identification of a specific individual, to distinguishing one person from another, or to personally identifiable information. These may be data that have been de-identified, or that could not lead to personally identifiable information in the first place (e.g., shopping preferences).

Online Survey Software

Online survey software provides questionnaires that research subjects can complete over the Internet. They are usually Web forms along with a database to store the answers, and statistical software to provide analytics.

Personal Health Information (PHI)

Personal health information (PHI), also referred to as protected health information, generally refers to demographic information, medical history, test and laboratory results, insurance information and other data that a healthcare professional collects to identify an individual and determine appropriate care.

Personally Identifiable Information (PII)

Personally Identifiable Information (PII) is information that, when used alone or with other relevant data, can identify an individual. PII may contain direct identifiers (e.g. passport information) that can identify a person uniquely, or quasi-identifiers (e.g. race) that can be combined with other quasi-identifiers (e.g. date of birth) to successfully recognize an individual.

Portable Storage Device

A portable device is any device that can easily be carried. It is a small form factor of a computing device that is designed to be held and used in the hands. Portable devices are becoming an increasingly important part of personal computing as the capabilities of devices like laptops, tablets and smartphones continue to improve. A portable device may also be called a handheld device or mobile device.

Public Facing

A public facing resource accepts anonymous connection requests from any public internet protocol address. In other words public facing resource are externally accessible resources that the public can access.

Raw Data

Data that have not been processed for meaningful use. Although raw data have the potential to become "information," they require selective extraction, organization, and sometimes analysis and formatting for presentation. Raw data have yet to be de-identified and thus, if there is any stage of the data lifecycle wherein your data will contain PII, it is this stage.

Research Data

Data that are used as primary sources to support technical or scientific enquiry, research, scholarship, or artistic activity, and that are used as evidence in the research process and/or are commonly accepted in the research community as necessary to validate research findings and results. All other digital and non-digital content have the potential of becoming research data. Research data may be experimental data, observational data, operational data, third party data, public sector data, monitoring data, processed data, or repurposed data.

Research Data Management (RDM)

Data Management refers to the storage, access and preservation of data produced from a given investigation. Data management practices cover the entire lifecycle of the data, from planning the investigation to conducting it, and from backing up data as it is created and used to long term preservation of data deliverables after the research investigation has concluded. Specific activities and issues that fall within the category of data management include: File naming (the proper way to name computer files); data quality control and quality assurance; data access; data documentation (including levels of uncertainty); metadata creation and controlled vocabularies; data storage; data archiving and preservation; data sharing and reuse; data integrity; data security; data privacy; data rights; notebook protocols (lab or field).