Ethical Digital Data Management

Facilitator: Chandra Kavanagh, PhD

The Problem

The Solution!

- We developed broadly-accessible digital data management education and policy tools.
- These resources are based on extensive research including an original qualitative research project with McMaster University researchers and research ethics board members, as well as an extensive literature review and environmental scan.

Research Data Management Tools

- Digital Data Management Glossary
- Digital Data Management FAQ
- Digital Data Management Matrix
- Digital Data Risk Matrix

Digital Data Management Glossary

- International Research Data Management glossary (IRiDiuM) a project between
- CODATA the Committee on Data of the International Council for Science
- CASRAI the standard dictionary of research administration information.

Glossary: Sample Definition

- 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.

Glossary: Sample Terms

- Cloud Services
- Data Lifecycle
- Data Management Plan
- Data Security
- Data Sharing
- De-identification
- Deletion
- Encryption
- High-Risk Data

Data Management FAQ

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 ?
6. What is the difference between wireless and wired connections?
7. What online survey software should I use?
8. What is the best way to share data with my co-investigators?

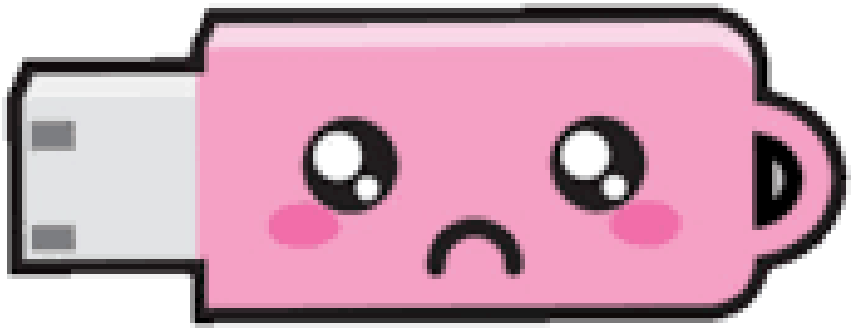
Ethical Data Management FAQ

- **Is a password-protected laptop a secure place to store my data?**

- One of the most common ways of collecting and storing data is through the use of a password-protected laptop. While this method might be secure enough for low-risk data, it is not secure enough for medium and high-risk data. If the data being collected is not low-risk, 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.

Ethical Data Management FAQ

- In some cases storing data on a portable storage device is a good option to add a level of protection. Some portable storage devices, like USB keys, are not connected to the Internet and as such they are less prone to remote access without permission. However these small devices may be more easily lost or stolen.



Ethical Data Management FAQ

- 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 avoid common situations where your laptop may get stolen such as leaving it in a vehicle or public place unattended. Perhaps most importantly, regularly back up and secure your data.

Ethical Data Management FAQ

- **How long can/should I keep my data?**
- The short answer? It depends on your data! But here are some things to look out for:
- 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 is. To learn more about the risk level of your data see the Research Data Management Matrix.

Ethical Data Management FAQ

- **What is encryption? When and how should I encrypt my data?**

- 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 of encrypting your data and they both have pros and cons:

Ethical Data Management FAQ

- Encrypting Individual Files
- Pros: 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.
- Cons: If someone had access to the computer where your data is stored they could break into it and view any non-encrypted files. You also have to remember to individually encrypt each new file you create.

Ethical Data Management FAQ

- Encrypting Your Drive
- Pros: Encrypting your entire drive protects from anyone to accessing any of your data without your authorization. Encrypting your whole device is also more convenient and less prone to error as all files are encrypted automatically.
- Cons: If you experience any corruption on your drive, it may be more difficult or even impossible to retrieve that data.

Ethical Data Management FAQ

- Methods to Try
- To encrypt your whole drive, or individual files, try VeraCrypt (Windows/Linux/OS) or GNU PrivacyGuard (Windows/Linux/OS). To encrypt and compress files you are going to be sending over the internet try 7-Zip.

Ethical Data Management FAQ

- **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. Cloud services can be public or private. 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.

Ethical Data Management FAQ

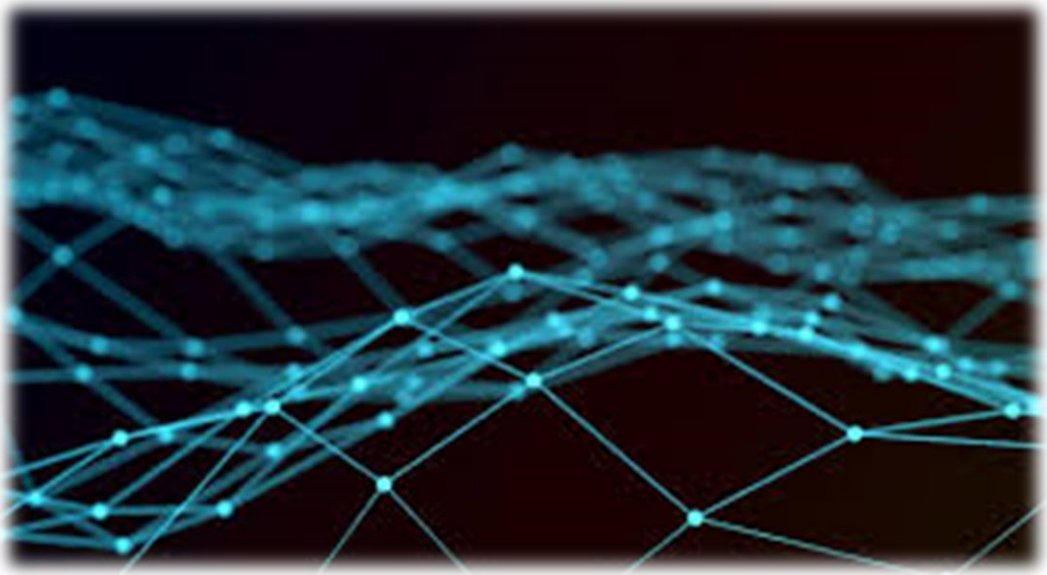
- Server Location: With public cloud storage data is stored in servers that could be anywhere in the world, and thus subject to that country's laws. With private cloud services your data is stored in local servers.

Ethical Data Management FAQ

- Server Control: 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.

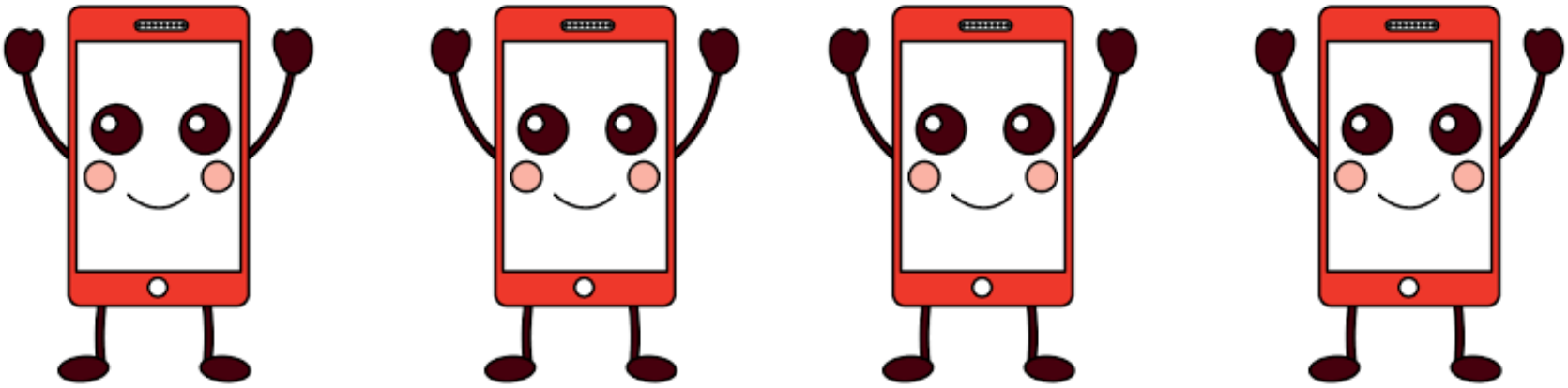
Ethical Data Management FAQ

- Attack Surface: 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.



Ethical Data Management FAQ

- **Is it safe to store my data on mobile devices such as cell phones or USB keys?**
- The answer to this question is different depending on whether we are talking about a portable storage device that has an internet connection, such as a cell phone, or a device that does not have an internet connection, such as a USB key.



Ethical Data Management FAQ

- **What is the difference between wireless and wired internet connections? Is one safer?**

- When it comes to connectivity, computers 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. Wireless connections, especially Bluetooth are the least secure. Wired network access is more secure than wireless. Finally, using a computer that is not connected to the internet is the most secure way to store your data.

Ethical Data Management FAQ

- **What online survey software should I use?**
- The ethics compliant McMaster survey service is 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) Other survey services should be avoided particularly those with servers located in the USA like SurveyMonkey as the information housed there is subject to the US Patriot Act/Domestic Security Enhancement Act.

Ethical Data Management FAQ

- **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. Ideally those collecting the research would remove all identifying personal information before the data was shared with research partners at other institutions.

Ethical Data Management FAQ

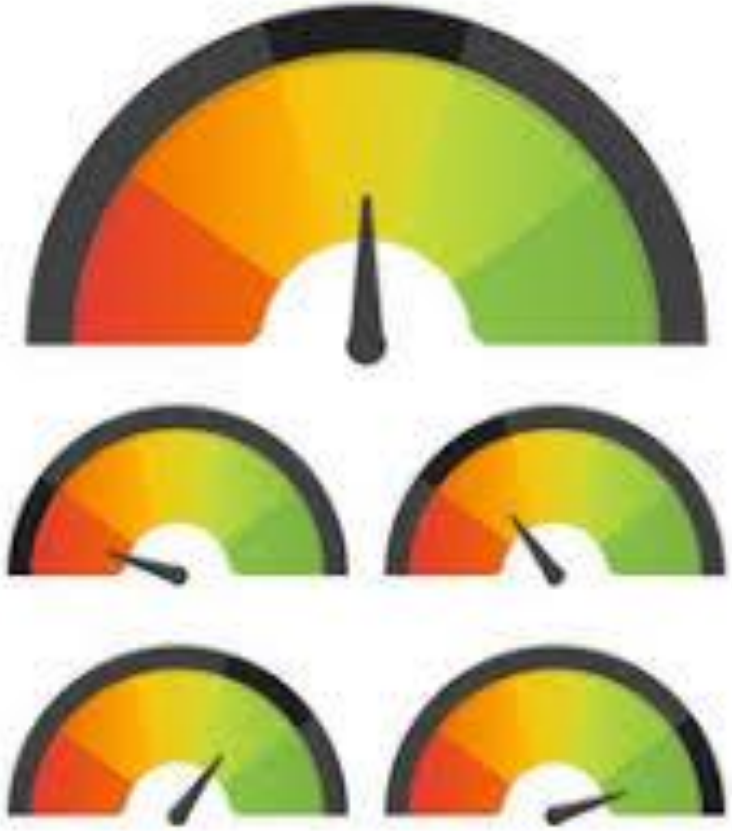
- Low-Risk Data: Can be shared via McMaster email and all cloud services including free personal cloud services (Google Drive, DropBox, iCloud, Onedrive etc.)
- Medium-Risk Data: Encrypted and password-protected files can be shared via McMaster email and McMaster approved cloud services.
- High-Risk Data: Restricted data should be shared hand to hand on a password-protected and encrypted data storage device. Maintaining ethical high-risk data transfer between institutions may require individualized strategies.

Research Data Management Matrix

	Low Risk	Medium Risk	High Risk
Types of Data	- Research data that does not contain any sensitive or identifiable information (e.g. data which has been de-identified). NB If in doubt, assume that data is sensitive. - Non sensitive research documentation - Publicly available information	- Research data that may or does contain sensitive or identifiable information - 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 confidential, restricted or highly sensitive information - Personal health information -Personal financial information -Data and research protocols related to highly sensitive intellectual property
Examples	- Completely de-identified or anonymous data - Blank consent forms and information sheets Information gathered from a public-facing website	- Video or audio recorded interviews depending on the content - Identification keys and signed consent forms - De-identified financial information associated with research payments	- 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
Data Protection	- Research data must always be	- Collect and store data on	- Collect and store data on

Risk Levels

- Low Risk Data
- Medium Risk Data
- High Risk Data



Low Risk Data

- Research data that does not contain any sensitive or identifiable information (e.g. data which has been de-identified). If in doubt, assume that data is sensitive.
- Non sensitive research documentation
- Publicly available information

Low Risk Data: Examples

- Completely de-identified or anonymous data
- Blank consent forms and information sheets
- Information gathered from a public facing website

Medium Risk Data

- Research data that may or does contain sensitive or identifiable information
- Some sensitive research-related documentation
- De-identified records of compensation
- Data and research protocols related to private or sensitive intellectual property

Medium Risk Data: Examples

- Video or audio recorded interviews depending on the content
- Identification keys and signed consent forms
- De-identified financial information associated with research payments

High Risk Data

- Research data that contains confidential, restricted or highly sensitive information
- Personal health information
- Personal financial information
- Data and research protocols related to highly sensitive intellectual property.

High Risk Data: Examples

- 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

Data Lifecycle

- Data Collection
- Data Access
- Data Storage
- Data Transfer
- Data Destruction

Research Data Management Matrix

	Low Risk	Medium Risk	High Risk
Types of Data	- Research data that does not contain any sensitive or identifiable information (e.g. data which has been de-identified). NB If in doubt, assume that data is sensitive. - Non sensitive research documentation - Publicly available information	- Research data that may or does contain sensitive or identifiable information - 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 confidential, restricted or highly sensitive information - Personal health information -Personal financial information -Data and research protocols related to highly sensitive intellectual property
Examples	- Completely de-identified or anonymous data - Blank consent forms and information sheets Information gathered from a public-facing website	- Video or audio recorded interviews depending on the content - Identification keys and signed consent forms - De-identified financial information associated with research payments	- 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
Data Protection	- Research data must always be	- Collect and store data on	- Collect and store data on

Research Data Risk Matrix

	Risks to Research Subjects	Risks to Researchers	Risks to Institutions	Risks to Data	Risk Management
Unauthorized Data Access	Loss of control, disclosure or access to identifiable and/or sensitive information could create significant harm to research participants depending on the severity profile of the data.	It is the researcher's responsibility to assess "threats to the security of information for all stages of the research life cycle, and implement appropriate measures to protect information." Researchers who fail to uphold this obligation face numerous psychological, social, professional and legal risks	"Institutions or organizations where research data are held have a responsibility to establish appropriate institutional security safeguards." Institutions that fail to uphold this obligation face numerous ethical, reputational and legal risks	Fulfilling the duty of controlling access to data is essential to the integrity of the research project. When data has been accessed in an unauthorized way its integrity and usability is thrown into question.	Password protection Encryption Avoid using cloud services Avoid transferring via email To learn more see the Research Data Management Matrix
Confidentiality Breach	Depending on the severity profile of the data participants are subject to a breach of trust at best, and	"Fulfilling the ethical duty of confidentiality is essential to the trust relationship between researcher	"The ethical duty of confidentiality includes obligations to protect information from unauthorized	"Fulfilling the ethical duty of confidentiality is essential to…the integrity of the research project".	All risk management tactics associated with "Unauthorized Data Access"

Research Data Risk Matrix

- Risks to Research Subjects
- Risks to Researchers
- Risks to Institutions
- Risks to Data
- Risk Management

Research Data Risk Matrix

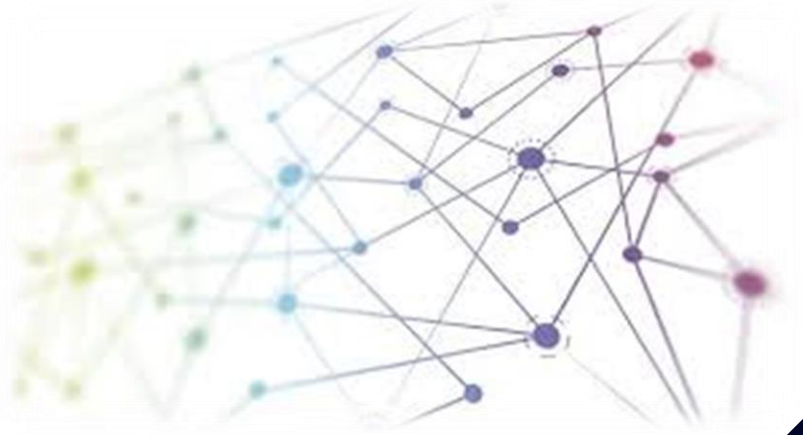
- Unauthorized Data Access
- Confidentiality Breach
- Data Corruption
- Data Loss

Research Data Risk Matrix

	Risks to Research Subjects	Risks to Researchers	Risks to Institutions	Risks to Data	Risk Management
Unauthorized Data Access	Loss of control, disclosure or access to identifiable and/or sensitive information could create significant harm to research participants depending on the severity profile of the data.	It is the researcher's responsibility to assess "threats to the security of information for all stages of the research life cycle, and implement appropriate measures to protect information." Researchers who fail to uphold this obligation face numerous psychological, social, professional and legal risks	"Institutions or organizations where research data are held have a responsibility to establish appropriate institutional security safeguards." Institutions that fail to uphold this obligation face numerous ethical, reputational and legal risks	Fulfilling the duty of controlling access to data is essential to the integrity of the research project. When data has been accessed in an unauthorized way its integrity and usability is thrown into question.	Password protection Encryption Avoid using cloud services Avoid transferring via email To learn more see the Research Data Management Matrix
Confidentiality Breach	Depending on the severity profile of the data participants are subject to a breach of trust at best, and	"Fulfilling the ethical duty of confidentiality is essential to the trust relationship between researcher	"The ethical duty of confidentiality includes obligations to protect information from unauthorized	"Fulfilling the ethical duty of confidentiality is essential to…the integrity of the research project".	All risk management tactics associated with "Unauthorized Data Access"

Connection Between Tools

- One of the novel aspects of these resources is the way they are linked together. You'll notice many of the terms appear in blue, these function as clickable links on the McMaster website. So, when you check out the data management matrix and see the you have high risk data that should be encrypted, you can click on the word encrypted and it will bring you to the definition of encryption in the glossary.



Questions

- For more information please contact:

Chandra Kavanagh, Ethics Officer

Health Research Ethics Authority

Phone: 709 777 8115

Email: ethicsofficer@hrea.ca