Data Privacy Best Practices

Trudi Wright, CRM

Do More with Digital Scholarship Workshop Series

March 10, 2022

Lewis & Ruth Sherman Centre for Digital Scholarship
scds.ca

McMaster University | Library

McMaster University sits on the traditional Territories of the Mississauga and Haudenosaunee Nations, and within the lands protected by the "Dish With One Spoon" wampum agreement.

- McMaster University sits on the traditional Territories of the Mississauga and Haudenosaunee Nations, and

Session Recording and Privacy

This session is being recorded with the intention of being shared publicly via the web for future audiences.

In respect of your privacy, participant lists will not be shared outside of this session, nor will question or chat transcripts.

Questions asked via the chat box will be read by the facilitator without identifying you. Note that you may be identifiable when asking a question during the session in an audio or visual format.

Code of Conduct

The Sherman Centre and the McMaster University Library are committed to fostering a supportive and inclusive environment for its presenters and participants.

As a participant in this session, you agree to support and help cultivate an experience that is collaborative, respectful, and inclusive, as well as free of harassment, discrimination, and oppression. We reserve the right to remove participants who exhibit harassing, malicious, or persistently disruptive behaviour.

Please refer to our code of conduct webpage for more information:

scds.ca/events/code-of-conduct/

Ice Breaker

What would you do?

Scenario:

You are investigating the feasibility of a research project, and find out that some of the data you would need for the project has already been collected by the Registrar’s Office.

Are there factors to consider in this scenario?

How likely are you to obtain access to this data?

What is Data Privacy?

Defining Privacy

1. The The Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2) defines privacy as "an individual's right to be free from intrusion or interference by others. It covers, among other things, an individual's body, personal information, expressed thoughts and opinions, personal communications with others, and the spaces they occupy."
2. The Information and Privacy Commissioner of Ontario defines privacy as "a fundamental right of every Ontarian. In order to protect that right, Ontario public institutions are required by law to protect your personal information, and to follow strict rules when collecting, using and disclosing your personal information"

Ethics & Privacy Risks

Privacy Protection Management Considerations

- Ethics concerns regarding privacy related to the degree to which information can be associated with a particular individual, the sensitivity of the information, and the extent to which access, use or disclosure may harm an individual or group.
- Privacy risks specifically focus on the potential harms that participants, or the groups to which they belong, may experience from the collection, use, and disclosure of personal information for research purposes.

Researchers should be aware of these risks and concerns, and plan data collection, use, retention and disclosure intentionally and purposefully.

Confidentiality

Safeguarding Sensitive and Personal Information

- The ethical duty of confidentiality refers to “the obligation of an individual or organization to safeguard entrusted information”. This duty is essential to the **integrity** of the research and to the **relationship of trust** between researchers and participants. Researchers must not misuse or wrongfully disclose information entrusted to them. This includes information derived from human biological materials. In addition to this ethical duty of confidentiality, researchers at McMaster are also bound by federal and provincial or territorial legislation that protects privacy rights.
- As part of their duty of confidentiality, researchers must let participants know what information they will be collecting, who will have access to it, how it will be protected, and how it will be used. This is captured throughout the consent process.

Legal Compliance for Privacy Protection

- Privacy laws in Ontario and Canada apply to the collection, use and disclosure of personal information

- Privacy laws are based on the Canadian Standards Association (CSA) Model Code for the Protection of Personal Information

- McMaster Privacy Office is responsible for overseeing privacy compliance, and providing support and guidance for privacy best practices

- The federal and provincial Privacy Commissioners oversee compliance with the privacy laws in Canada. Together, the Commissioners focus on protecting the privacy interests of Canadians, educating the public and organizations on privacy best practices, investigating complaints, and assisting individuals and organizations recover from privacy breaches

Personal Information

Defining Terms

Personal Information (FIPPA)

- information relating to the race, national or ethnic origin, colour, religion, age, sex, sexual orientation or marital or family status of the individual,
- information relating to the education or the medical, psychiatric, psychological, criminal or employment history of the individual or information relating to financial transactions in which the individual has been involved,
- any identifying number, symbol or other particular assigned to the individual,
- the address, telephone number, fingerprints or blood type of the individual,
- the personal opinions or views of the individual except where they relate to another individual,
- correspondence sent to an institution by the individual that is implicitly or explicitly of a private or confidential nature, and replies to that correspondence that would reveal the contents of the original correspondence,
- the views or opinions of another individual about the individual, and
- the individual's name where it appears with other personal information relating to the individual or where the disclosure of the name would reveal other personal information about the individual.

Personal Health Information (PHIPA)

- Information relating to the physical or mental health of the individual, including information that consists of the health history of the individual's family,
- information relating to the providing of health care to the individual, including the identification of a person as a provider of health care to the individual,
- is a plan of service within the meaning of the Home Care and Community Services Act, 1994 for the individual,
- is a plan that sets out the home and community care services for the individual to be provided by a health service provider or Ontario Health Team pursuant to funding under section 21 of the Connecting Care Act, 2019,
- information relating to payments or eligibility for health care, or eligibility for coverage for health care, in respect of the individual,
- Information relating to the donation by the individual of any body part or bodily substance of the individual or is derived from the testing or examination of any such body part or bodily substance,
- is the individual's health number, or
- identifies an individual's substitute decision-maker.

Principle-Informed Practices

Connecting Compliance with Best Practice

- The principles of information privacy provide researchers with measures to promote transparency, openness and accountability.
- Principle-Informed privacy practices help to build trust in relationships between researcher and participant(s).
- When researchers know that their use of personal and confidential information is subject to scrutiny, they are more likely to be sensitive to and respectful of other people’s privacy and more aware of their own privacy.

Fair Information Practices

Office of the Information and Privacy Commissioner (Canada)

There are 10 privacy principles, including:

1) accountability
2) identifying purposes
3) consent
4) limiting collection
5) limiting use, disclosure and retention
6) accuracy
7) safeguards
8) openness
9) individual access
10) challenging compliance

Best practices

Data Management

- Intentional management of data through a Data Management Plan, including:
  - Building in data privacy protocols
    - Access scope
    - Storage protocols
    - Retention policy
    - Document disclosure and deletion/destruction
    - Information security elements
  - REB certification requirement

Technology and Privacy

Managing Privacy and Information Security

- Information Security - Measures taken to protect information. It includes physical, administrative, and technical safeguards.
- Privacy Impact Assessment/Information Security Assessment
  - Data ownership
  - Access Protocols
  - Dedicated privacy staff
  - Privacy Incident Management
  - Challenges with long term retention

Privacy Protection

Protective Measures - Overview

- The type of information to be collected;
- The purpose for which the information will be used, and the purpose of any secondary use of identifiable information;
- Limits on the use, disclosure and retention of the information;
- Risks to participants should the security of the data be breached, including risks of re-identification of individuals;
- Appropriate security safeguards for the full life cycle of information;
- Any recording of observations (e.g., photographs, videos, sound recordings) in the research that may allow **identification** of particular participants;
- Any anticipated uses of personal information from the research; and
- Any anticipated linkage of data gathered in the research with other data about participants, whether those data are contained in public or personal records

Best Practice

Range of Consent

Consent – An indication of agreement by an individual to become a participant in a research project. Consent is a fundamental process for respecting individual autonomy. It must be voluntary, informed, and ongoing.

Express Consent – specific authorization is required when collecting sensitive PI (e.g., health, financial)

Opt-Out Consent – the ability to request that PI not be used for the identified purposes. Opt-out can be used for secondary purposes such as fundraising or marketing

Implied Consent – assuming consent based on an individual's actions or inactions. It must be so obvious that the person would consent if asked

Best Practice

Informed Consent

In order to obtain informed consent, researchers need to provide:

- The purpose for the collection of information, including potential secondary uses.
- How the information will be stored/protected
- Who (roles) will have access to the information
- How the information will be disseminated/published
- How long the information will be retained

Ongoing Consent:

- Should the purpose change, or an additional secondary use is intended, researchers should communicate these with participants to confirm ongoing consent.

Best Practice

Access and Data Sharing

- Research integrity requires researchers to establish access protocols
  - Levels of access – identifiable data, de-identified data, anonymous data
  - Identified roles (not individuals at each level
- Is it FOI-able?
  - Research information is excluded from the *Freedom of Information and Protection of Privacy Act*, but must be for a specific and identifiable research project.
    - However, funding information may be responsive for a Freedom of Information request.
  - Published, or publicly available information is also excluded from FIPPA.

Personal Information & Identifiability

Presenting Data 1

Where researchers will be collecting, using, sharing or accessing information (i.e., derived from data or human biological materials) that may potentially identify an individual, that information must be protected. For this reason, it is necessary to distinguish between identifiable and non-identifiable information in the context of the research being considered.

- Directly Identifiable information – Information that may be reasonably expected to identify an individual, alone or in combination with other available information. Sometimes referred to as "personally identifiable information" (PII).
- Indirectly identifying information – the information can reasonably be expected to identify an individual through a combination of indirect identifiers

Personal Information & Identifiability

Presenting Data 2

- Coded information - direct identifiers are removed from the information and replaced with a code. Depending on access to the code, it may be possible to re-identify specific participants (e.g., the principal investigator retains a list that links the participants' code names with their actual names so data can be re-linked if necessary). The process is called **De-identification**.
- Anonymized information - the information is irrevocably stripped of direct identifiers, a code is not kept to allow future re-linkage, and risk of re-identification of individuals from remaining indirect identifiers is low or very low.
- Anonymous information - the information never had identifiers associated with it (e.g., anonymous surveys) and risk of identification of individuals is low or very low.

Secondary Use of Information

When consent may be needed

- Secondary use research relies on information originally collected for a purpose other than the current research purpose. There are several reasons why a researcher may want to use such information, for example, to corroborate or criticize the conclusions of the original project, or to confirm the authenticity of the original data. From the perspective of participants, one important reason to base a research study on secondary use is to avoid duplicating unnecessarily the inconvenience and risks of research. For example, if one biological sample can serve two (or more) purposes, this reduces the burden on participants.
- Research that relies on secondary use of information may not always be necessary to seek the consent of the participants. In the case of research involving secondary use of identifiable information, a researcher must satisfy the REB that justifies not seeking consent. Where the researcher is relying exclusively on secondary use of non-identifiable information, participant consent is not required.
- Research that relies exclusively on the secondary use of anonymous information does not same level of rigor in protection as identifiable information.

Privacy Challenges

Longitudinal Studies – Long Retention Period

- Key to informed consent for participants
- Ongoing engagement and consent confirmation
- Access management
  - Managing roles over time
  - Managing data preservation
    - Navigating technology changes
    - File formats – conversion
    - Storage – migration

Privacy Challenges

Aggregated Data – Small Cell Challenge

- Aggregate data refers to numerical or non-numerical information that is (1) collected from multiple sources and/or on multiple measures, variables, or individuals and (2) compiled into data summaries or summary reports, typically for the purposes of public reporting or statistical analysis—i.e., examining trends, making comparisons, or revealing information and insights that would not be observable when data elements are viewed in isolation.
- Indirectly identifying information – The information can reasonably be expected to identify an individual through a combination of indirect identifiers (e.g., date of birth, place of residence, or unique personal characteristic).
    - Identifiability through small cell inclusion – enabling identifiability.

Privacy Challenges

Managing Privacy and Big Data

- Privacy for large data collections (big data) involves properly managing the data to minimize risk and protect personal and sensitive data. Because big data comprises large and complex data sets, many traditional privacy processes are less effective in handling the scale and velocity (data generation) required.
- To safeguard big data and ensure it can be used for analytics, researchers need to create a framework for privacy protection that can handle the volume, velocity, variety, and value of data as it is moved between environments, processed, analyzed, and disseminated.
- Big data presents potential opportunities for analysis, and identifying new secondary uses for data (triggering REB and consent activities).
- With data sets stored across multiple locations, the need for greater control increases, as well as the risk of a privacy breach.

Resources

Resources to Bookmark!

- TCPS 2: CORE-2022 (Course on Research Ethics)
- Freedom of Information and Protection of Privacy Act, Personal Health Information Protection Act
- General Data Protection Regulation (EU)
- McMaster Privacy Office
- IPC resources
- International Association for Privacy Professionals
- McPherson Institute resources

Trudi Wright
Privacy & Records
Management Specialist
privacy@mcmaster.ca