Sensitive Data Toolkit for Researchers
Part 2: Human Participant Research Data Risk Matrix

Prepared by the Portage Network Sensitive Data Expert Group on behalf of the Canadian Association of Research Libraries (CARL)

SEPTEMBER 2020

Portage Network
Canadian Association of Research Libraries
portage@carl-abrc.ca

www.carl-abrc.ca

portage
SERVICES PARTAGÉS POUR LES DONNÉES DE RECHERCHE
SHARED STEWARDSHIP OF RESEARCH DATA

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES
ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

Introduction

The Sensitive Data Expert Group of the Portage Network has created a suite of tools for Canadian researchers. These tools have been created to help researchers understand how research data is involved in the research ethics process, and to address the evolution of research data management (RDM) practices such as data sharing and deposit in the context of existing research ethics frameworks.

This tool, the Human Participant Research Data Risk Matrix, is intended to help researchers determine risk level for human participant research data, and make decisions with respect to its management, deposit, and appropriate access/future use. Risk can be determined through consideration of three factors: 1. identifiability of the data at the time of collection and deposit; 2. vulnerability of the data subjects as individuals or as part of a community or population; and 3. sensitivity of the data in terms of its ability to cause harm - e.g., physical, psychological/emotional, social and legal. This matrix is intended to be used in concert with Part 1 of this toolkit, the Glossary of Terms for Sensitive Data used for Research Purposes.

The Portage Network's Sensitive Data Expert Group is composed of a broad membership from research communities - including research ethics professionals, representatives of funding agencies, and members of Indigenous organizations - with direct interests in sensitive research data. The group works together to develop practical guidance and tools for managing sensitive data in the Canadian landscape.

	Low Risk	Medium Risk	High Risk	Extreme Risk
Risk level definitions	Publicly available data where there is no reasonable expectation of privacy, regardless of sensitivity or identifiability. Data collected with no information that could reasonably identify individuals or groups. Data contains no confidential, private, or sensitive information. Data subjects are not vulnerable in the context of the research and would not be harmed if a breach were to occur.	All identifiers collected have been stripped so that data to be deposited has no information that could reasonably identify individuals or groups. Data may contain information originally collected as confidential, private or sensitive. Data subjects are not vulnerable in the context of the research and would not be harmed if a breach were to occur.	Identifiers remain and/or (re)-identification is possible or probable. Data contains confidential, private or sensitive information. Data subjects may be vulnerable in the context of the research and may be harmed if a breach were to occur.	Data acquired through an agreement (formal or informal) with a custodian, barring further use or retention. Identifiers remain and/or (re)-identification is possible or probable. Data contains confidential, private or sensitive information. Data subjects are vulnerable in the context of the research and would be harmed if a breach were to occur.

	Low Risk	Medium Risk	High Risk	Extreme Risk
Informed Consent	Notification that data will be made available for future use.	Notification that data will be made available for future use. Option to opt-out of deposit should be considered.	Notification that data may be made available for future use. Request for permission to share and/or deposit data clearly included in consent form or process. If possible, provide options regarding areas of future research.	Confidentiality will be maintained for as long as the data exist. Data will not be shared beyond the research team.

	Low Risk	Medium Risk	High Risk	Extreme Risk
Data Collection	Publicly available data may be found online or in public archives, or be collected through naturalistic observation. Researchers do not know the identities of research participants/data subjects. Methods should not involve direct interaction with research participants. These typically involve surveys, questionnaires and observational research. No direct or indirect identifiers are collected.	Researchers may know the identities of research participants/data subjects and may have promised confidentiality through informed consent. Methods for data collection are wide-ranging and may involve direct interaction with research participants. Direct and/or indirect identifiers may be collected. **The majority of human research will fall into this category.**	Researchers may know the identities of research participants/data subjects and may have promised confidentiality through informed consent. Methods for data collection are wide-ranging and may involve direct interaction with research participants. Direct identifiers may or may not be collected, but indirect identifiers collected may be sufficient to render participants identifiable.	Researchers may know the identities of research participants/data subjects and will have promised confidentiality through informed consent. Methods for data collection are wide-ranging and may involve direct interaction with research participants. Direct identifiers may or may not be collected, but indirect identifiers collected may be sufficient to render participants identifiable.

	Low Risk	Medium Risk	High Risk	Extreme Risk
Data Analysis/ Management	No restrictions in the analysis of data for publicly available data. Data analysis should adhere to the REB-approved protocol and informed consent document/script.	Direct identifiers should be replaced as soon as possible with a linking code (e.g., pseudonym, alpha numeric code) and separated physically and/or electronically from the master list. Consent forms should be stored separately from research data. Only members of the research team should have access to identifiable data.	Direct identifiers should be replaced as soon as possible, with a linking code, and separated physically and/or electronically from the master list. Consent forms or notes with identifiers should be stored separately from research data. Indirect identifiers should be coded, if possible. Data should not be accessed/analyzed in a public space where others could see data on a device or by other means.	Direct identifiers shall be replaced as soon as possible, with a linking code, and separated physically and/or electronically from the master list. Consent forms or notes with identifiers shall be stored separately from research data. Indirect identifiers shall be coded, if possible. Data shall only be accessed by members of the research team, as described in the approved protocol, and access/analysis shall only occur in a secure environment.

	Low Risk	Medium Risk	High Risk	Extreme Risk
Data Storage (Active Storage) and Security	All storage devices, file sharing, and cloud services are allowed, including both public and institutional cloud services. Data should be backed up in a way that is consistent with the risk level associated with these data.	Identifiable data should be stored on password-protected devices, in appropriate secure locations. If data need to be accessible through the internet, they should be encrypted. Public cloud services should not be used, unless no other options exist. If they are used, files and access should be password-protected and encrypted. Private cloud services, as supported by the research institution and/or assessed as being secure, may be used. Data should be backed up in a way that is consistent with the risk level associated with these data.	All data should be stored on password-protected encrypted devices, in appropriate secure locations. If data need to be accessible through the internet, they should be encrypted. Public cloud services are strictly prohibited. Private cloud services, as supported by the research institution and/or assessed as being secure may be used, if approved by the REB. Data should be backed up in a way that is consistent with the risk level associated with these data.	All data shall be stored on a centralized, stand-alone computer/site that is both password protected and encrypted, in appropriate secure locations. Data should be backed up in a way that is consistent with the risk level associated with these data.

	Low Risk	Medium Risk	High Risk	Extreme Risk
Data Mobility/ Sharing	Can be shared via email and all cloud services including public cloud services.	Encrypted and password-protected files can be shared via email and institution-approved cloud services or collaboration sites.	Restricted data shall only be shared with other members of the research team, as specified in the approved protocol. Files shall be encrypted and password protected.	Data restricted to a centralized, stand-alone computer/site that is password protected and encrypted. Files should not be copied or shared. Access to data shall be restricted to authorized individuals explicitly identified in the REB protocol and should involve the smallest number of individuals possible.

	Low Risk	Medium Risk	High Risk	Extreme Risk
Data Deposit and Access (including secondary use)	Data should be deposited with unrestricted access within a reasonable timeframe, taking into account publication of original papers. Secondary data use does not require REB approval.	Data from participants/data subjects who opt out should be separated from data to be deposited. De-identified data should be deposited with unrestricted access within a responsible timeframe, taking into account publication or original papers, need to replicate research and ensure appropriate shelf-life for reuse of the data. Secondary use of de-identified data currently requires REB approval.	Data from participants/data subjects who opt out should be separated from data to be deposited. De-identified data should be deposited with restricted access to be evaluated by the data custodian. Data may be separated into sets depending on potential uses that participants have agreed to through informed consent (e.g. use for this study only, only for studies in this subject area, or for any use). Secondary data use requires REB approval.	Data should not be deposited anywhere, beyond the direct storage and access needs of the research team.
Data Retention and Destruction	Data may be retained indefinitely for discovery, access, and archival purposes.	Data may be retained indefinitely for discovery, access, and archival purposes.	Data may be retained indefinitely for discovery, access, and archival purposes in accordance with the REB-approved protocol.	Data must be destroyed at the earliest opportunity, in accordance with the REB-approved protocol.