Guidelines for considering access to human participant research data

Introduction

A fundamental purpose of human research is to generate generalizable knowledge that may be utilized to benefit communities and/or society at large. This notion is expressed by the Canadian granting agencies in their respective mandates:

“To excel, according to internationally accepted standards of scientific excellence, in the creation of new knowledge and its translation into improved health for Canadians, more effective health services and products and a strengthened Canadian health care system…” CIHR Overview

“The work SSHRC supports encourages the deepest levels of inquiry. It spurs innovative researchers to learn from one another’s disciplines, delve into multiparty collaborations and achieve common goals for the betterment of Canadian society. Research outcomes are shared with communities, businesses and governments, who use this new knowledge to innovate and improve people’s lives.” About SSHRC

“The Natural Sciences and Engineering Research Council of Canada (NSERC) helps make Canada a country of discoverers and innovators for the benefit of all Canadians.” NSERC’s Vision

For maximum utilization of research results, it is important for researchers to communicate, collaborate and share, as much as possible, particularly when the research is being funded by public money. Data sharing is therefore considered a public good, and is not only encouraged, but mandated by national and international public research funding agencies, including CIHR, SSHRC, Genome Canada, the National Institutes of Health, UK Wellcome Trust and the European Commission.

However, when personal data are involved, serious consideration must be given to one of the central principles of research ethics: respect for privacy and duty of confidentiality to research participants. This principle is recognized in most of the data sharing policies, represented as an exclusion category to the policy. However, a total exclusion of all human participant data breaches another tenet – respect for justice – which includes the right to participation, and involvement in public good.

Sharing of data in human participant research need not be an all- or none- decision, but instead, as in all research ethics situations, requires a balance of principles to reach a beneficial conclusion. The purpose of these guidelines is to assist researchers and Research Ethics Board (REB) members in understanding the issues involved in data sharing in general, and of human data, in particular. Policy requirements from large research funders of University of Toronto research will be described and put into context for human participant research data. Finally, logistical aspects will also be addressed.
**Definitions and Principles**

**What are research data and research materials?**

Research data are defined as factual records used as primary sources for research and are commonly accepted in the research community as necessary to validate research findings. Research data may include electronic data sets, interview transcripts, survey results, qualitative information in digital format, still and moving image and sound databases, and other digital objects used for analytical purposes.

Data may be **personal** – information that identifies or could potentially identify someone – and/or **confidential** – information that has been given with the expectation of secrecy and that could potentially cause damage to the source if released to third parties. The informed consent process is the time at which the researcher and participant should discuss and agree to the terms of use and disclosure of personal and/or confidential data.

Research data are useful to other investigators for replicating results and furthering new research discoveries: e.g. small molecules, organisms, viruses, cell lines, nucleic acids, purified enzymes, antibodies, reagents, source code and software, protocols, research tools for evaluation, questionnaires, interview guides, data abstraction forms, and manuals for patient services.

**What is data access?**

Data access refers to the removal of barriers – legal, financial and technical – to potential users of research data. Open access allows for broad release of data with no barriers. Controlled access requires that permissions be sought by the prospective user prior to sharing of research data by the researcher or custodian. Such permissions may take the form of online “click-wrap” agreements with terms and conditions of usage, open content licenses (e.g. Creative Commons license) or an individual project requisition process (Christian, 2009).

**What are the benefits of access to research data and materials?**

“Sharing and openness are the hallmarks of the scholarly tradition. Researchers publish their results, not for financial return, but to engage other researchers to build upon them and to contribute to the progress of knowledge in their fields” (Shearer, 2011).

Sharing of research data has been strongly supported by many reports across the world, with significant benefits for research and society considered to include:

- **Accelerating scientific progress:** The biggest example of this is arguably the Human Genome Project, which involved thousands of researchers sharing data through online repositories,
usually prior to publication, enabling sequencing of the whole human genome in less time than originally planned.

- **Avoiding duplication in research**: Researchers can use data or results that already exist and build upon these to look at related phenomena; or they may combine, or “mash” data sets to examine seemingly unrelated phenomena or drive new hypotheses.

- **Enabling replication and verification of research results**: Ensures data integrity and reliability.

- **Increasing visibility and impact of research**: by utilizing data already available, time and money may be spent to translate the findings into knowledge, activities and products that can benefit society in the developed and developing world (Borgman, 2011).

Each of these proposed benefits are good reasons for human research participants to agree to share their data. Balanced against these are the potential risks of harm for data access.

**What are the risks of access to research data?**

Most reports acknowledge the legal and ethical issues associated with research data sharing. While legal issues can exist for any type of data, the ethical issues are strongly tied to personal information.

- **Privacy and confidentiality of data**: data may enable research participants to be identified, either directly through the research data being shared, or through linking or mashing data sets.

- **Intellectual property**: research data may include information that is patentable; sharing may allow third parties to profit from the information or discoveries or products directly related to the shared data and materials.

- **Knowledge acquired from community partnerships**: communities may have “ownership” over the information and not wish that it be shared with other researchers or the public.

- **Sensitive data**: third party knowledge may cause harm to the individuals from whom the data originated (e.g., data related to national security).

All of these categories are common exemptions from data sharing policies. However, in many cases, researchers, in planning their research and using accepted research ethics guidelines (such as the Tri-Council Policy Statement) should consider these risks in context. In many cases, the arguments for sharing such research data outweigh the risks. Balancing the risks and benefits of data sharing will be discussed in the Guidance section below.
Current requirements for University of Toronto based research, by funding body


- The scope of this policy includes all research outputs that have been financially supported in whole or in part by CIHR including industry-partnered research. Research outputs include: peer-reviewed journal publications; research materials; and final research data.
- All applicants of CIHR grants must include a Research Output Access Plan, including anticipated outputs and how they may be made accessible to others, or provide reasons for any restrictions on access to research outputs.
- Research materials should be provided to recipients of not-for-profit research institutions at cost and with as few restrictions as possible. They may also be shared with the commercial (for-profit) research community following university and institution procedures regarding material transfer agreements.
- Final data sets, generally in electronic form, should be made available upon request after the publication date of a peer-reviewed publication and should be of quality and accompanied with necessary metadata (i.e., information that describes the characteristics of the data set) or codebooks.
- Original data sets arising from CIHR-funded research should be kept for a minimum of five years after the last date of the "Authority to Use Funds" period of the grant. This applies to all data, whether published or not.
- Institutions and REBs may have additional policies and practices regarding the preservation, retention, and protection of research data that must be respected.

SSHRC Research Data Archiving Policy (1990)

- The purpose of this policy is to facilitate access of SSHRC-funded research data to other researchers, thereby encouraging researchers to share research data and provide opportunities to further analyze, replicate, verify and refine research findings. It is expected to enhance progress within fields of research as well as support the expansion of inter-disciplinary research.
- The policy requires that all research data collected with the use of SSHRC funds be preserved and made available for use by others within two years of the completion of the research project for which the data were collected.
- As with CIHR’s policy, SSHRC recognizes that modifications or exemptions may be required for human participant research and REBs should be consulted for advice on appropriate action.

Genome Canada Data Release and Resource Sharing (2008)

- Genome Canada is committed to the principle of rapid data release and sharing of unique resources with the scientific community, and therefore requires that data and resources that they fund be made available quickly, with few or no restrictions.
• Data must be released and shared no later than the original publication date of the main findings from any datasets generated by that project. For large datasets that are collected over periods or phases, data may be released in phases as they become available or as main findings are published. However, at the conclusion of a project, all data must be released without restriction.
• Patentable data must be released no later than the date that patents have been filed or at the time of publication, whichever comes first. A maximum extension delay of 90 days is acceptable if there are extenuating circumstances. Permission to delay release of data must be obtained in writing from Genome Canada.
• Sharing of resources generated by the projects, such as unique biological specimens and computer programs designed to analyze datasets should also be addressed. Biological reagents such as unique strains should be deposited into repositories and computer programs designed to analyze large datasets should be made available to others through the use of license agreements that adhere to open source principles.


• NIH supports data sharing, believing that it is essential for expedited translation of research results into knowledge, products and procedures to improve human health. Final research should be shared to serve these and other important scientific goals.
• Researchers submitting an NIH application that seeks $500,000 or more in direct costs in any single year are expected to include a plan for data sharing or explain why data sharing is not possible.
• NIH recognizes that in some cases data sharing is limited by institutional policies, local IRB rules, local, state and Federal laws and regulations, including the Privacy Rule.

Wellcome Trust Policy on Data Management and Sharing (2010)

• Wellcome Trust is committed to ensuring that the research outputs that it funds, including research data, are managed and used in ways that maximize public benefit.
• Funded researchers should maximize the availability of research data with as few restrictions as possible and must submit a data management and sharing plan prior to an award being made.
• All users of research data should acknowledge the sources of their data and abide by the terms and conditions under which they accessed the original data.

While all of these public research funding bodies require open access to research data, each provide room for exceptions for human participant research data. This leeway is important, but not necessarily helpful, as it is left to the researcher and REB to determine whether or how much of the data must be excluded from access by other researchers. Keeping too much may pose risk to participants; keeping too
little may render the data useless for any potential users. The following are guidelines for assisting researchers in developing data sharing plans and REB members in reviewing them.

**Guidelines for sharing of human participant research data**

As with all matter of ethics review of human participant research, it is essential that risks and benefits of data sharing are balanced. Moreover, it is important that researchers and REBs consider the importance of data access prior to recruitment of participants, to ensure that the following four aspects are represented in the research protocol:

- Expressing the provision for data sharing when gaining informed consent
- Protecting people’s identities by de-identifying data where needed
- Considering whether to control access to data by secondary users
- Considering whether one needs to apply an appropriate licence (ANDS, 2012)

Researchers submitting for ethics review should include a data sharing plan as part of their protocol submission. To determine whether data sharing is appropriate, and in what format, researchers and REB members should consider the probability of a third party being able to identify or re-identify participants*, and, perhaps more importantly, what potential magnitude of harms may reasonably arise from possessing or using these data, once re-identified. The University of Toronto Risk Matrix, used as part of the ethics protocol submission, may be useful in this assessment.

*Identifiability categories, in increasing order from top to bottom:

- Anonymous
- De-identified with low risk of re-identification
- De-identified with high risk of re-identification
- Identifiable

These categories of identifiability are quite fluid in practice, and may be time-and situation-dependent. As the ability to link unrelated data sets together increases, so do the potential for re-identification and possibility that sensitive information may be included.
Balanced against these present and future risks are the benefits of data sharing. These include potential benefits to the populations and/or communities represented by the data, as data sharing may add knowledge and lead to treatments or policies that may benefit them directly. Benefits to the greater public good, from understanding phenomena and making decisions based on that knowledge should also be considered.

Whenever data sharing is a possibility, participants should be informed of these requirements and intentions, and whether the data will be identifiable or potentially identifiable (with the probability of re-identifiability quantified and explained). Researchers must also determine at the planning stage, whether all research data will be shared, or whether participants can opt out of being part of the shared data set. If the latter, informed consent may be layered, so that individuals may agree to participate, but refuse data sharing or limit data sharing to research in the same area.

If possible, data sharing should be reaffirmed at the end of the study in the case of identifiable data. However, in some situations, data may not be removed from the data set for sharing beyond a certain point in the collection or analysis, or at all. If so, it may be necessary to exclude participants from being involved in the study if they are not willing to let other researchers have access to their data.

In cases where data exist and consent has not been sought for sharing beyond the original research study, TCPS2 articles on secondary use of data and data linkage (Articles 2.4, 5.5 – 5.7) should be considered. Researchers should seek the opinion of their REB before providing user access to data where consent was not explicitly sought.

**Practical tips and considerations for researchers considering data sharing**

1. **Study design:** Consider what types of data you intend on collecting and whether such data may be useful beyond the proposed research study. If the data do have utility beyond the study, consider how future uses could be optimized by current practices or mechanisms for data collection (format – paper versus electronic), storage (physical storage, coding, level of identification) and protection.

2. **Informed consent:** The possibility of using or sharing data beyond the current research study should be part of the informed consent process. Consider probable secondary uses for the data and whether they may be useful or appropriate for other studies within the same or similar research area, or for more general use. A step-wise consent, where participants may opt in for some uses and opt out of others may be helpful. Consent to re-contact at a later date may be warranted in circumstances where future uses are uncertain. See Appendix 1 for example.

3. **Data collection and analysis:** Methods of data collection and maintenance during analysis can have implications on their future utility. Consider the format for data collection, preferably choosing the gold standard program or system for the relevant discipline. Determine what identifiers or variables should be collected or not (and why), coded or stripped, and at what point in time. Consider whether the information collected, in the chosen format, may promote
or detract from sharing in the future. In some cases, sharing is not desired and it may be strategic to render data unusable, if accessed.

4. Risks and benefits: What are the potential risks of individual and/or community harm from legitimate access and/or wrongful access? Do the potential benefits of sharing, (e.g. extension of the research to advance knowledge or to develop treatments, processes or policies), outweigh potential risks? Have these risks and benefits been considered and discussed through the informed consent process? If not, should re-consent be obtained from research participants who have agreed to share their data?

5. Ability to withdraw consent and/or data: Consider if participants who agree to have data shared can change their minds and withdraw consent. At what point, if at all, can their data be removed from the repository, or destroyed altogether? All of these issues should be articulated through the informed consent process.

6. Storage, retention and archiving: Where should the data be stored in the long term? What actions should be taken to prepare the data for storage, whether in electronic or hard copy format? Determine how easily data may be accessed, whether legitimately or not. Do you have access to a centralized repository, either physically on-site or web-based?

7. Sharing agreements: Consider who should have access to your data and what the process should look like. If low sensitivity or identifiability, perhaps the data may be made available in a publicly accessible forum. If higher sensitivity or risk of re-identification, perhaps a permission process should be developed. Is it likely that intellectual property and/or commercialization issues may arise in the future? Would participants be concerned if industry (e.g. insurance companies) or government researchers obtain access to the data? Licensing arrangements or a sharing strategy should be developed early, to reduce chances of misunderstandings and unauthorized use of data to occur later on.

8. Consult with members from the appropriate REB or staff from the Office of Research Ethics (ORE) or the Freedom of Information and Protection of Privacy (FIPP) Office. Data protection and access are important issues in the protection of research participants and the researcher. Balancing the risks and benefits of data sharing is important, and ORE staff can provide helpful insight into how best to promote research, while protecting all involved.

**Logistics of Data Access**

While the funding agencies require that research data be accessible, there is no current standard of how and where such data should be kept. The ORE will work with the chairs of the U of T based REBs and the Office of the Vice-President, Research and Innovations to develop procedures to assist with the logistical aspects of access, keeping in mind the special requirements needed for human participant research data. Researchers should engage their REB members and the ORE for any questions that they may have regarding data sharing plans and data access.
References


Appendix 1. Example of Tiered Consent Language

The following informed consent form is from the National Cancer Institute’s Informed Consent Template for Cancer Treatment Trials (2011) and demonstrates tiered consent, whereby the research participant may agree to be involved (directly, or indirectly through banked tissue samples) in some aspects of future studies and/or be re-contacted for subsequent consent at a later date. This example could be adapted for other types of research, whether involving data or biospecimens.

********

Consent Form for Use of Tissue for Research

About Using Tissue for Research
You are going to have a biopsy (or surgery) to see if you have cancer. Your doctor will remove some body tissue to do some tests. The results of these tests will be given to you by your doctor and will be used to plan your care.

We would like to keep some of the tissue that is left over for future research. If you agree, this tissue will be kept and may be used in research to learn more about cancer and other diseases. Please read the information sheet called "How is Tissue Used for Research" to learn more about tissue research.

Your tissue may be helpful for research whether you do or do not have cancer. The research that may be done with your tissue is not designed specifically to help you. It might help people who have cancer and other diseases in the future.

Reports about research done with your tissue will not be given to you or your doctor. These reports will not be put in your health record. The research will not have an effect on your care.

Things to Think About

The choice to let us keep the left over tissue for future research is up to you. No matter what you decide to do, it will not affect your care.

If you decide now that your tissue can be kept for research, you can change your mind at any time. Just contact us and let us know that you do not want us to use your tissue. Then any tissue that remains will no longer be used for research.
In the future, people who do research may need to know more about your health. While the xyz may give them reports about your health, it will not give them your name, address, phone number, or any other information that will let the researchers know who you are.

Sometimes tissue is used for genetic research (about diseases that are passed on in families). Even if your tissue is used for this kind of research, the results will not be put in your health records.

Your tissue will be used only for research and will not be sold. The research done with your tissue may help to develop new products in the future.

**Benefits**
The benefits of research using tissue include learning more about what causes cancer and other diseases, how to prevent them, and how to treat them.

**Risks**
The greatest risk to you is the release of information from your health records. We will do our best to make sure that your personal information will be kept private. The chance that this information will be given to someone else is very small.

**Making Your Choice**
Please read each sentence below and think about your choice. After reading each sentence, circle "Yes" or "No". If you have any questions, please talk to your doctor or nurse, or call our research review board at IRB's phone number.

No matter what you decide to do, it will not affect your care.

1. My tissue may be kept for use in research to learn about, prevent, or treat cancer.
   Yes  No

2. My tissue may be kept for use in research to learn about, prevent or treat other health problems (for example: diabetes, Alzheimer's disease, or heart disease).
   Yes  No

3. Someone may contact me in the future to ask me to take part in more research.
   Yes  No