DATA MANAGEMENT TRAINING MODULES

Module #6: DATA SHARING AND REUSE

Plans for sharing and reusing data are an integral part of the Research Data Management Planning process.

Many funders as well as journal publishers encourage, expect or require researchers to prepare and provide their data for sharing. This is particularly true of data produced through public funding. Policies in the United Kingdom and the United States are firmly in place. The Canadian government and associated funding agencies have adopted a guideline outlining a commitment to ensuring access to research results from publicly funded projects.

Examples of current data sharing and preservation policies are listed below. See “Module 2: Appendix A Funding Requirements” for details.

Canada

- Government of Canada Access to Research Results: Guiding Principles
- Canadian Institutes of Health Research (CIHR) Open Access Policy
- Social Sciences and Humanities Research Council (SSHRC) Data Archiving Policy
- Social Sciences and Humanities Research Council (SSHRC) Open Access Policy

U.S.

- National Science Foundation Data Management Plan Requirement
- National Institutes of Health (NIH) Data Sharing Policy

U.K.

- Research Councils UK (RCUK) Common Principles on Data Policy
- Economic and Social Research Council (ESRC) Research Data Policy
- Natural Environment Research Council (NERC) Data Policy
- Wellcome Trust Data Management and Sharing Policy

OECD

- OECD Declaration on Access to Research Data from Public Funding

Rationale for sharing data:

- Encourages scientific enquiry
- Promotes innovation
- Reduces duplication of research projects
- Leads to new collaborations
• Increases impact of research results
• Reduces costs of research in developing nations
• Encourages scrutiny, transparency and accountability
• Can be used in teaching

Preparing data for sharing
Preparation of data for sharing begins with the creation of a data management plan during the initial stages of the research project. Researchers should familiarize themselves with the policies of their funders as part of the planning process.

Factors to consider include:

- Legal and ethical implications
  - Will confidentiality of participants be compromised?
  - Will sensitive information be compromised?
  - Will it violate contractual agreements?
  - Will it violate licencing agreements?
  - Was sharing included in the informed consent agreement?
  - Will the data need to be anonymized prior to release?
  - Do you have consent from project partners?
  - Do you have the right to share secondary data?
- Intellectual property rights
  - Will you be commercializing or seeking patents?

The University of Guelph Library Research Enterprise and Scholarly Communication team provides guidance on data management issues and options as well as author rights. Researchers should also consult the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS) and the Office of Research for information on contractual and ethical obligations.

Obtaining consent
The Tri-Council Policy Statement (TCPS) stipulates that informed consent from project participants is necessary for the sharing and reuse of data containing identifiable information (TCPS 2 Article 3.2 and Article 5.2). To ensure that consent has been received, consideration should be paid to wording in the consent form regarding preservation, reuse and/or sharing data containing identifiable information, how this information would be protected and under what conditions the data would be shared or reused. Consent from participants is not required for sharing and reuse of anonymous or aggregated data, however, informing participants about preservation, reuse and sharing of this data is considered ethical.
Conditions for Sharing

Canadian copyright legislation does not cover raw research data although it does cover descriptions of data such as tables, graphs and databases. Sharing of data files can be controlled and protected by use of licences. Researchers, in many cases, can decide on the level of access and conditions of use related to data they are sharing or depositing in a repository. Individual repositories may have embedded licence choices within the repository platform.

Several online licencing options can be adopted for personal use:

- **Creative Commons (CC)** – allows users to combine elements of licences to create a licence for the research data in question
- **Open Data Commons (ODC)** - provides three licence options

Conditions of use should reflect the nature of the data and level of confidentiality involved.

Setting conditions of use can include:

- Requiring researcher authorization for access
- Setting access permissions for specific researcher groups
- Placing data under timed embargos
- Providing secure access to data
- Acknowledgment and attribution of original researcher

Anonymising data

Personal identifying information should never be disclosed through research findings unless explicit informed consent from participants has been provided in writing.

Researchers must ensure that a person’s identity cannot be disclosed through:

- Direct identifiers
  - Includes names, addresses, date of birth, postal code, telephone numbers, social insurance numbers, images, etc.

- Indirect identifiers
  - When linked in combination with multiple identifiers or publicly available information have the potential to reveal a participant’s identity
  - Includes workplace information, occupation, age, salary, etc.

Direct identifiers collected during the research process usually are not essential for data analysis and can be easily removed from the data. Consideration should be paid to length of time these identifiers are kept separately and securely and to the manner in which they may be
destroyed. In many cases, collection of direct identifiers can be avoided during the collection process.

Anonymising quantitative data may involve removing or aggregating variables. Techniques such as cell suppression, rounding, inference control and perturbation can be employed to anonymise data. Coding information using standard classifications at higher levels than the data has been collected is an example of a low risk technique which can be employed in the anonymising process.

Relational data requires particular attention where connections between variables may inadvertently cause identities to be revealed. Transcription of interviews may require the employment of different techniques such as the use of consistent pseudonyms, or more generalized terms to reduce risk of identification without rendering the data unusable. Retain unedited versions of your data for use within the team or in the event of errors during anonymisation. Remember to log all techniques used and instances of replacements or aggregation of variables.

Refer to “Module 5: Storing and Preserving Data” for information on long-term preservation. See “Module 7: Research Data Management Resources” for supporting tools and materials which augment information in these training modules

Acknowledgements
We would like to thank the UK Data Service UK Data Service for use of their training materials in the creation of this module.

We would also like to thank the EDINA and Data Library, University of Edinburgh for use of materials from the Research Data MANTRA [online course] in the creation of this module.