Research Data Management:

Secure and Preserve: Best Practices to Preserve your Research Data

OCTOBER 19, 2017

CAROL PERRY, LIBRARY

A. MICHELLE EDWARDS, OAC

As you finish one project and move onto another, you want to ensure that the legacy of your project's research data is preserved. What best practices should you implement to encourage proper data preservation?

In today's research environment, your research data may be requested by other research labs around the world to replicate your data. How should your data be documented and preserved to facilitate the exchange of research data?

Objectives

Understand the basics :

- How organized is your research project?
- How well documented is your data and your research project?
- Are you ready to share/preserve your data?

Research life cycle

Research Data Management Plan

RDM Refresher

A sound strategy and best practices used to……

- Organize
- Document
- Store
- Analyze
- Secure
- Preserve/Share/Reuse

……Your data

Setting the stage

Workshop Scenario

The French Blast Research Group received an NSERC grant in 2016-2017 to continue an ongoing feeding trial. The research involves three horses from a number of different farms, where each horse is weighed(kg) at the beginning of the trial, placed on one of three feed regimens (hay, pasture, or silage), their total feed intake(kg) is measured for 2 consecutive weeks, and their weights (kg) are taken at the same time.

Additional information that is gathered includes: the name of the horse's owner, the annual income of the owner, and feed costs of the owner for their horses.

You have been hired by the French Blast Research Group to manage their data collections. They have 5 years of horse feeding trial data that was collected from a number of horse farms across Ontario since 2011. You have been hired to review, manage, and preserve their research trial data according to best practices and requirements of the NSERC grant.

Why should I care?

Can you find the data you need?

Can you understand the data you have?

Can others understand the data you have?

Are you required to handle your data in a specific way?

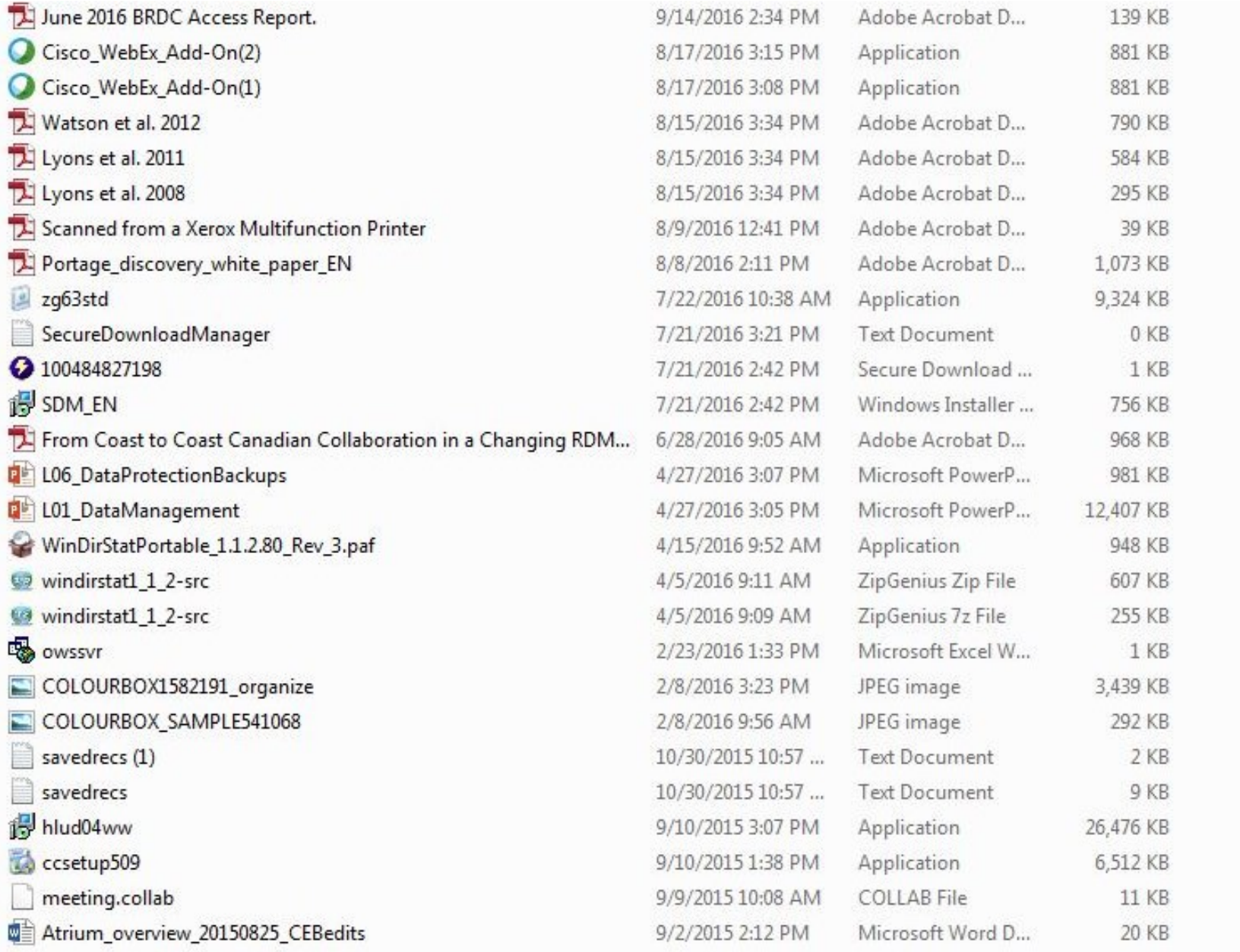
Are you required to share or preserve it?

How will you share/preserve your data?

Where will you share/preserve it?

Is this what your directory structure looks like?

Sample directory: desktop ERS folder



June 2016 BRDC Access Report.	9/14/2016 2:34 PM	Adobe Acrobat D...	139 KB
Cisco_WebEx_Add-On(2)	8/17/2016 3:15 PM	Application	881 KB
Cisco_WebEx_Add-On(1)	8/17/2016 3:08 PM	Application	881 KB
Watson et al. 2012	8/15/2016 3:34 PM	Adobe Acrobat D...	790 KB
Lyons et al. 2011	8/15/2016 3:34 PM	Adobe Acrobat D...	584 KB
Lyons et al. 2008	8/15/2016 3:34 PM	Adobe Acrobat D...	295 KB
Scanned from a Xerox Multifunction Printer	8/9/2016 12:41 PM	Adobe Acrobat D...	39 KB
Portage_discovery_white_paper_EN	8/8/2016 2:11 PM	Adobe Acrobat D...	1,073 KB
zg63std	7/22/2016 10:38 AM	Application	9,324 KB
SecureDownloadManager	7/21/2016 3:21 PM	Text Document	0 KB
100484827198	7/21/2016 2:42 PM	Secure Download ...	1 KB
SDM_EN	7/21/2016 2:42 PM	Windows Installer ...	756 KB
From Coast to Coast Canadian Collaboration in a Changing RDM...	6/28/2016 9:05 AM	Adobe Acrobat D...	968 KB
L06_DataProtectionBackups	4/27/2016 3:07 PM	Microsoft PowerP...	981 KB
L01_DataManagement	4/27/2016 3:05 PM	Microsoft PowerP...	12,407 KB
WinDirStatPortable_1.1.2.80_Rev_3.paf	4/15/2016 9:52 AM	Application	948 KB
windirstat1_1_2-src	4/5/2016 9:11 AM	ZipGenius Zip File	607 KB
windirstat1_1_2-src	4/5/2016 9:09 AM	ZipGenius 7z File	255 KB
owssvr	2/23/2016 1:33 PM	Microsoft Excel W...	1 KB
COLOURBOX1582191_organize	2/8/2016 3:23 PM	JPEG image	3,439 KB
COLOURBOX_SAMPLE541068	2/8/2016 9:56 AM	JPEG image	292 KB
savedrecs (1)	10/30/2015 10:57 ...	Text Document	2 KB
savedrecs	10/30/2015 10:57 ...	Text Document	9 KB
hlud04ww	9/10/2015 3:07 PM	Application	26,476 KB
ccsetup509	9/10/2015 1:38 PM	Application	6,512 KB
meeting.collab	9/9/2015 10:08 AM	COLLAB File	11 KB
Atrium_overview_20150825_CEBedits	9/2/2015 2:12 PM	Microsoft Word D...	20 KB

Folder structure - refresher

BCDS_2015
BCDS_2016
BCDS_2017

OR

bcds_feed_suppl
bcds_silage
bcds_hay

OR

bcds_farm_a
bcds_farm_b
bcds_farm_c

Readme file – sample

Beef Cattle Diet Survey 2016
BCDS = Beef Cattle Diet Survey

Folders:
bcds_feed_suppl = *feed containing supplements data – all farms*
bcds_hay = *hay feed data - all farms*

Files:
bcds_raw = *rawdata files – all farms*

Data collected daily stored in monthly files January 1 2016-December 2016.
Date = ISO date yyyy/mm/dd
UOM = units of measure
Note: Farm 3 data not collected January 15 2016 and January 16 2016.
Out of range values are shown as -6999

Access, sharing, reuse

Advantages to sharing

Increases impact of your research

Helps others replicate your research

Encourages further scientific enquiry

Reduces research costs by reducing duplication

Encourages transparency and accountability

Access, sharing, reuse

Review legal & ethical obligations/restrictions

- Research Ethics Board
- Funding agency obligations
- Federal/provincial regulations
  - Animal Use Data to Animal Care Services - annual reporting to the Canadian Council on Animal Care

Anonymize data where necessary

- Remove identifying information

Determine sharing criteria – understand your Intellectual Property rights

Legal & Ethical issues

REB form
- What did you agree to regarding sharing and preserving your data?

Funding policies
- Are there sharing and preservation requirements?

Regulations
- What federal/provincial policies & regulations govern your data?

Contracts/partnerships
- Are there restrictions to sharing your data due to commercial interests?

Journals
- Does your publisher require data be shared/deposited?

Anonymization

Shared data or data preserved on a repository platform should be anonymized.

Determinents:

- Can data be anonymized without rendering it unusable by others?
- Different types of data may be highly sensitive (health) and might never be shared
- Sometimes, consent or access control is a better alternative to anonymization

Methods of anonymization:

- Aggregation- cell suppression; inference control; perturbation; rounding; sampling
- Masking
- Pseudonmizination (de-identifying)

Anonymizing research data

Direct identifiers

Remove direct identifiers (or replace with pseudonyms)
e.g. names, address, institution, photo

Indirect identifiers

Reduce precision/detail through aggregation
e.g. birth year vs. date of birth, occupational categories, area rather than village

Generalize meaning of detailed text
e.g. occupational expertise

Restrict upper lower ranges to hide outliers

e.g. income, age

Anonymizing qualitative data

Remove direct identifiers, or replace with pseudonyms or replacements

Avoid blanking out

Identify replacements with [brackets]

Plan for anonymization at time of transcription

Avoid over-anonymization – avoid distorting data, making it unreliable or misleading

Keep log of anonymization actions taken

- Keep separate from anonymized data files
- Do not share log file publicly

Exercise 1

Working in groups of 2-3, anonymize the data file from our study.

10 – 15 minutes

Review

Explain the measures undertaken to anonymize the file.

Any questions?

Preservation

Preservation options

Institutional repository

Publish with results

Deposit in major data repository

Deposit in discipline-specific data repository

University of Guelph-based options

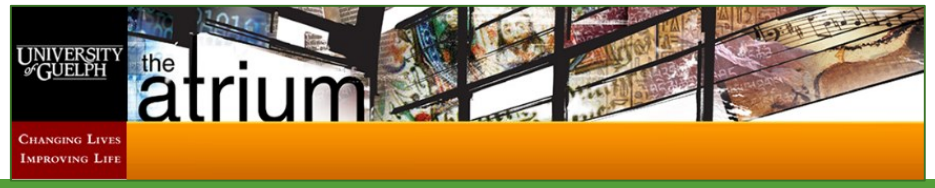
Atrium (institutional repository) - e-theses, articles, reports, videos, etc.
https://atrium.lib.uoguelph.ca/

Agri-environmental Data Repository- research data

https://dataverse.scholarsportal.info/dataverse/ugardr

University of Guelph Data Repository – research data

https://dataverse.scholarsportal.info/dataverse/ugrdr



Discipline Specific Repositories

Re3data.org – global registry of research data repositories
www.re3data.org/

Stanford University Libraries – Guide to Domain-specific Data Repositories

https://library.stanford.edu/research/data-management-services/share-and-preserve-research-data/domain-specific-data-repositories

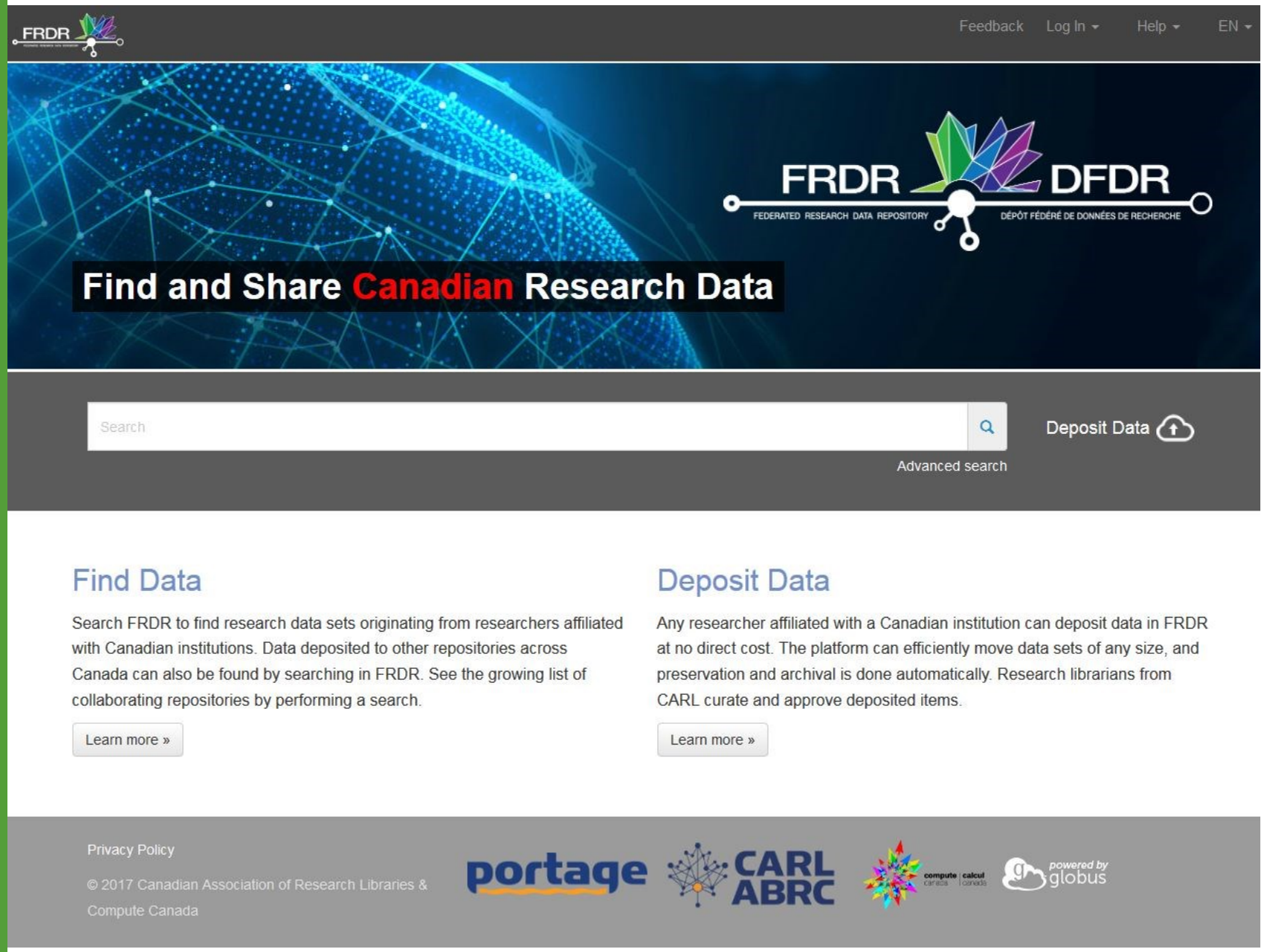
re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

National preservation & discovery system

Part of: Research Data Canada Federated Pilot Project

Data repositories will be connected through a national network and portal for discovery

https://www.frdr.ca/repo/



Metadata

Metadata = documentation that describes the project and its contents

Best practice for project management

Repositories will require some documentation to accompany your files

- Project level
- File level
- Item level

- End user terms of use
- Citation information

Metadata element	Description	Fill in information as applicable
Title	Full title of the dataset	
Author (s)	Person, corporate body, or agency responsible for the work's intellectual content. Include names, affiliations, and email addresses (e.g. [NAME], [DEPARTMENT/SCHOOL], [COLLEGE], [UNIVERSITY].	
Production Date	The date the final version of the dataset being deposited in the repository was created. This is the date that will be used in the dataset citation.	
Funding agency	The source(s) of funds for the production of the work	
Time period covered	Time period covered by the data	
Date of collection	The date(s) when the data were actually collected	
Keywords	Words or phrases that describes the data collection's content (please provide at least 5 keywords)	
Description	Summary describing the purpose and scope of the data collection, and what questions the investigators attempted to answer.	

Exercise – Metadata

Begin filling in the metadata template with information from the Project Scenario.

Preparing files

TRANSFER FILES TO NON-PROPRIETARY FORMAT

- ascii
- text
- csv
- Pdf or pdf/a
- tiff
- jpeg2000
- shp,shx,
- Mpeg4
- xml
- html

PREPARE DOCUMENTATION TO ACCOMPANY FILES

- Codebooks
- Syntax files
- Readme files
- User guides
- Scripts etc.

Sample accessible dataset

Potential for northward expansion of the American dog tick (Dermacentor variabilis, Say) range under climate change in North America

http://hdl.handle.net/10864/GSSJX

End user licensing

Determine what access or actions you want end users to be able to make

Provide clear instructions on citation requirements and contact information

Sample licence:

Restrictions

Open access.

Citation Requirements

The publishing of analysis and results from research using any of the data products is permitted in research communication such as scholarly papers, journals and the like. The authors of these communications are required to cite the primary investigators Minigan J. N. et al. (2017), as the source of the data, and to indicate that the results or views expressed are those of the author/authorized user and are not those of primary investigators.

Depositor Requirements

To provide funding agencies with essential information about use of archival resources and to facilitate the exchange of information about Agri-Environmental Research Data Repository (AERDR) participants research activities, users of AERDR data are requested to send to AERDR bibliographic citations for, or copies of, each completed manuscript or thesis abstract. Please indicate in a cover letter which data were used.

Conditions

Original publication must be cited (see Related Publications); see publication for methods.

Disclaimer

The original creators of the data, J.N. Minigan, H.A. Hager, A.S. Peregrine, and J.A. Newman, and the funding agency, the Ontario Ministry of Agriculture, Food and Rural Affairs, bear no responsibility for uses of this data set or for interpretations or inferences upon such uses

Resources

MANTRA – Research Data Management Training. University of Edinburgh

- http://mantra.edina.ac.uk/protectionrightsandaccess/

Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans - 2010

- http://www.pre.ethics.gc.ca/pdf/eng/tcps2/TCPS_2_FINAL_Web.pdf

TCPS 2 Tutorial Course on Research Ethics (CORE)

- http://www.pre.ethics.gc.ca/eng/education/tutorial-didacticiel/

UK Anonymisation Network - Resources

- http://ukanon.net/ukan-resources/

Contact information:

Michelle Edwards
OAC Statistics Consultant
oacstats@uoguelph.ca

http://www.uoguelph.ca/oac/stats

Carol Perry
Associate Librarian Research & Scholarship
lib.research@uoguelph.ca