Research Data Management

Jeff Moon
Data & Government Information Librarian
Academic Director, Queen’s Research Data Centre
Data management: encompasses the entire data life cycle, including data collection, organization, storage, retrieval, and preservation over time.
The Data Life Cycle

Why manage & share your data?

- Promotes higher quality research and peer review
- Opens the door to further research
- Demonstrates effective stewardship of publicly-funded data
- Ensures long-term preservation of and access to data
- Builds in procedures to ensure anonymity/privacy
Other factors to consider...

• Some scientific publications require that data be available for the scientific community before the article is published

• A growing number of funding agencies require that data be deposited in a public archive

• Future researchers can go to the ‘archive’ for your data... rather than trying to find you
OK, but where do I turn?

There are many resources related to research data management

Queen’s University Library has pulled selected resources together into a guide...
Research Data Management

http://guides.library.queensu.ca/rdm

Overview

Introduction

The whole notion of developing a data management plan and depositing data may seem daunting to some. Queen’s University Library is here to support and partner with you to make sure this process is as painless as possible. See the Writing a Data Management Plan tab for more information. For assistance please contact your Subject Liaison Librarian. Or contact madgic@queensu.ca.

Why Deposit Your Data?

Depositing Data -- Advantages...

Queen’s University Library provides services for archiving, preserving and disseminating data. Both researchers and the scientific community as a whole benefit if data are archived, because:

- accessibility of data increases openness of research,
- open access enhances scientific discussion,
- many datasets contain material that have not yet been analysed,
- secondary use of publicly-funded data increases efficient use of resources,
- existing data may be used for new research questions,
- archiving will ensure long-term preservation of data,
- archiving incorporates procedures to ensure data privacy.

Support for Data Management

Contact Information:
madgic@queensu.ca

Jeff Moon
Data & Government Information Librarian
moonj@queensu.ca
(613) 533-6000 x77992

Alexandra Cooper
Data & Web Support Technician
coopera@queensu.ca
(613) 533-6000 x77481
Topics covered...

- Overview
- Writing a Data Management Plan
- Funding Agency Guidelines
- Metadata
- Data Repositories & Archives
- Citing Data
- Best Practices in Data Management
- Library as Data Partner
- QUL Research Data Archive
Data Management Workflow

RDM Planning
- RDM Plans: Help researchers with RDM planning

Data Ingest
- Receive & Store Data: DGIC staff hard drives, and DGIC working space on QShare
- Track Data: Enter ‘project’ into QRDM tracking database
- Backup Data: Ensure deposited data is safely backed up

Data Processing
- Data Assessment/Cleanup: Data verification, disclosure risk assessment, determine file formats to be archived, etc.
- Permissions: Obtain researcher permissions for dissemination, establish embargo period (if any), Copyright, security issues

Data Publishing
- Documentation: Gather, organize, and enrich documentation
- Data Markup: Describe the data file using XML standard (e.g. DDI)
- Follow-up: Conduct follow-up consultations with researcher(s) as necessary
- Data Citation: Establish DOI with DataCite Canada

Local ‘archives’
- SSDC Nesstar Discovery Systems
- QSpace ‘Data Archive’ sub-community in Library Community

External ‘archives’
- Dataverse OCUL Scholars Portal
- ODESIE OCUL Scholars Portal
- Discipline-based Archives Various
The Library as Data Partner

Writing a Data Management Plan
- Funding Agency Guidelines

Best Practices in Data Management
- Citing Data

Metadata

QUL Research Data Archive
- Data Repositories & Archives

RDM Planning
- Help researchers with RDM planning

Data Ingest
- Receive & Store Data
  - DGIC staff hard drives, and DGIC working space on QShare
- Track Data
  - Enter ‘project’ into QRDM tracking database
- Backup Data
  - Ensure deposited data is safely backed up

Data Processing
- Data Assessment/Cleanup
  - Data verification, disclosure risk assessment, determine file formats to be archived, etc.
- Permissions
  - Obtain researcher permissions for dissemination, establish embargo period (if any), Copyright, security issues
- Documentation
  - Gather, organize, and enrich documentation
- Data Markup
  - Describe the data file using XML standard (e.g., DDI)
- Follow-up
  - Conduct follow-up consultations with researcher(s) as necessary
- Data Citation
  - Establish DOI with DataCite Canada

Data Publishing
- Local ‘archives’
  - SSDC Nesstar
  - Discovery Systems
- QSpace
  - ‘Data Archive’ sub-community in Library Community
- External ‘archives’
  - Dataverse
    - OCUL Scholars Portal
  - ODESI
    - OCUL Scholars Portal
  - Discipline-based Archives
    - Various
Key things the Library can help with...

- Developing a Data Management Plan (DMP)
- Data assessment
- Finding appropriate data archive(s)
- Storing/backing up your data
- Choosing an appropriate metadata standard
Data Management Plans

Recent Tri-Council consultation paper
16 October 2013

Toward a Policy Framework for Advancing Digital Scholarship in Canada

Recent Tri-Council consultation paper

Establishing a Culture of Stewardship

a. a requirement that **all grant applications include specific data management plans** including identified costs of data collection/analysis and preservation of results and associated datasets;

b. definition of those **specific elements** of data plans that will be considered by reviewers in the assessment of funding applications;

c. guidelines indicating **which data must be preserved** and in what formats;

d. consolidated **open access policies** and guidelines (in concert with work already initiated by the TC3+);

e. guidelines for researchers in **selecting suitable data repositories**;
Coordination of Stakeholder Engagement

The Canadian research environment is characterized by a high degree of commitment to collaboration and cooperation, with individuals and organizations combining forces to address long-term planning and initiate “bottom-up” actions to shape the evolution of the digital landscape. These important actions, which bring a richness of community engagement, would benefit from development of a coordinating mechanism to ensure enhanced alignment to optimize returns on the investment of time by all parties. To help ensure the maximum contribution of key players, the TC3+ would work with other organizations and working groups to ensure ongoing consultation and coordination with all stakeholders, including the provinces, in the development of Canada’s national digital infrastructure for research.
Developing Capacity and Future Funding Parameters

To help create a forward-looking digital research environment, the parameters for the funding of coordinated national scale infrastructure should be re-examined...

a. best practices in administration, operations, policy and access;

b. enhanced networks and infrastructure

c. skills development and graduate and researcher training.
Data Management Planning

TAKEAWAY MESSAGE HERE

The Library can help with this!
Data Assessment

**Data Verification**: check dataset integrity, variable names/labels/groups, coding, missing data, etc.

**Disclosure Risk Assessment**: check for identifying variables, detailed geography (e.g. Postal Codes), etc.

**File Formats**: Assess file formats delivered, alternative formats, and conversion options
Finding the right data archive(s) for your data

Discipline-based archives
Bala Parental Alienation Study: Canada, United Kingdom, and Australia : [3]
<table>
<thead>
<tr>
<th>Preview</th>
<th>Issue Date</th>
<th>Title</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2013</td>
<td><strong>Data Files</strong></td>
<td>Bala, Nicholas; Hunt, Suzanne; McCarney, Carrie; Ashbourne, Christine; Gwynne, Erin</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td><strong>Documentation</strong></td>
<td>Bala, Nicholas; Hunt, Suzanne; McCarney, Carrie; Ashbourne, Christine; Gwynne, Erin</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td><strong>Summary of Data</strong></td>
<td>Bala, Nicholas; Hunt, Suzanne; McCarney, Carrie; Ashbourne, Christine</td>
</tr>
</tbody>
</table>
## QSpace – data file

<table>
<thead>
<tr>
<th>A1</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Title</td>
<td>Reference</td>
<td>Prov</td>
<td>Decision</td>
<td>Custody/Control Before</td>
<td>Outcome</td>
<td>Alienator</td>
<td>Expert Opinion</td>
<td>Court Ordered</td>
<td>PA Found Expert</td>
</tr>
<tr>
<td>2</td>
<td>Poliut v. Poliut (SK Crt Queen's bench)</td>
<td>[2008] CarswellSask 160</td>
<td>SK</td>
<td>Father alienated children from mother; custody changed to joint; father can't seek contact children for 3 months</td>
<td>Father</td>
<td>No remedy for PA - Support reordered</td>
<td>Male</td>
<td>Assessment</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>D. (L.) v. B. (A.) (Droit de la famille - 03229) (Cour supéérieure)</td>
<td>[2000] CarswellQue 1302</td>
<td>PQ</td>
<td>&quot;The Wife has been totally rejected by her two sons, aged 16 and 13 - to an extent that is beyond comprehension. It is obvious that they are alienated from her; but not appropriate to reduce child support in order to punish an alienating parent&quot;</td>
<td>Father</td>
<td>No remedy for PA - Support reordered</td>
<td>Male</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Flaber v. Flaber (ON SCJ)</td>
<td>[2008] CarswellOnt 6548</td>
<td>ON</td>
<td>Father alienated children from mother; loses custody and contact</td>
<td>Father</td>
<td>Custody change</td>
<td>Male</td>
<td>Assessment - OCL</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>Corby v. Corby (ON SJ)</td>
<td>[2006] CarswellOnt 5180</td>
<td>ON</td>
<td>Mutual hostility between parents, but F primary instigator. Encouraged PA. Sole custody to M.</td>
<td>Joint</td>
<td>Custody change</td>
<td>Male</td>
<td>Assessment - OCL</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>L. (J.K.) v. S. (N.C.) (ON SCJ)</td>
<td>[2006] CarswellOnt 2903</td>
<td>ON</td>
<td>Father loses custody; non sent for &quot;deprogramming&quot; in US</td>
<td>Father</td>
<td>Custody change</td>
<td>Male</td>
<td>Assessment</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>7</td>
<td>C. (A.S.) v. C. (S.) (MB Crt Queen's Bench)</td>
<td>[2008] CarswellMan 190</td>
<td>MB</td>
<td>Father alienated children from M; joint custody therapy; no sole because older children wouldn't stay; judge does not want to separate youngest from others.</td>
<td>Father</td>
<td>Custody - joint legal therapy</td>
<td>Male</td>
<td>Assessment</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>Gerth v. Gerth (BC Sup Crt)</td>
<td>[2007] CarswellBC 884</td>
<td>BC</td>
<td>M &amp; F separated; rotated one week on/one off in not home. M had affair; F wanted mail home and kids aligned with him</td>
<td>Joint - Blamed mother for breakdown</td>
<td>Custody - joint; F can't be in home for one month then resume schedule</td>
<td>Male</td>
<td>No - Assessment ordered</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>L. (R.A.) v. R. (R.D.) (AB Crt Queen's bench)</td>
<td>[2007] CarswellAlta 183</td>
<td>AB</td>
<td>Father alienated children from M. M gets sole custody; F loses access.</td>
<td>Father</td>
<td>Custody change</td>
<td>Male</td>
<td>Assessment</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
BALA PARENTAL ALIENATION STUDY: CANADA, UNITED KINGDOM, AND AUSTRALIA 1984-2012

datacite:

If you use these data, please add the following citation to your scholarly references. Why cite?

Nicolas Bala; Suzanne Hunt; Carrie McCarney; Erin Gwynne; Christine Ashborne, 2013-01-23, "BALA Parental Alienation Study: Canada, United Kingdom, and Australia 1984-2012", http://hdl.handle.net/10864/10279 Data and Government

Data Citation Details

<table>
<thead>
<tr>
<th>Study Global ID</th>
<th>hdl:10864/10279</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors</td>
<td>Nicolas Bala (Queen's University Faculty of Law); Suzanne Hunt (Research Assistant, Queen's University Faculty of Law); Carrie McCarney (Research Assistant, Queen's University Faculty of Law); Erin Gwynne (Research Assistant, Queen's University Faculty of Law); Christine Ashborne (Research Assistant, Queen's University Faculty of Law)</td>
</tr>
<tr>
<td>Producer</td>
<td>Nicolas Bala, Queen's University Faculty of Law</td>
</tr>
<tr>
<td>Production Date</td>
<td>January, 2013</td>
</tr>
<tr>
<td>Funding Agency</td>
<td>Social Science and Humanities Research Council (SSHRC) and the Law Foundation of Ontario</td>
</tr>
<tr>
<td>Distributor</td>
<td>Data and Government Information Centre (DGIC), Queen's University</td>
</tr>
<tr>
<td>Distributor Contact</td>
<td>Data and Government Information Centre (Queen's University), <a href="mailto:madgic@queensu.ca">madgic@queensu.ca</a></td>
</tr>
<tr>
<td>Distribution Date</td>
<td>January 23, 2013</td>
</tr>
<tr>
<td>Deposit Date</td>
<td>January 23, 2013</td>
</tr>
<tr>
<td>Provenance</td>
<td>Queen's University Library Research Data Archive Dataverse</td>
</tr>
</tbody>
</table>
ODESI – Data Portal

Social Surveys

- CANADA
- INTERNATIONAL

Globalization of Personal Data (GPD)

  [Seven-Country] NEW
- Metadata
- Variable Description

  [China] NEW
  [Japan] NEW
### Dataset: The Globalization of Personal Data (GPD) Project International Survey on Privacy and Surveillance, 2006 [Seven-Country]

**Type:** Column percentage

<table>
<thead>
<tr>
<th>[q6] What level of trust do you have that private companies will protect your personal information?</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very high level of trust</td>
<td>7.5</td>
<td>7.9</td>
<td>7.7</td>
</tr>
<tr>
<td>Reasonably high level of trust</td>
<td>33.6</td>
<td>37.0</td>
<td>35.4</td>
</tr>
<tr>
<td>Fairly low level of trust</td>
<td>33.2</td>
<td>32.9</td>
<td>33.0</td>
</tr>
<tr>
<td>Very low level of trust</td>
<td>25.7</td>
<td>22.2</td>
<td>23.9</td>
</tr>
<tr>
<td>Total</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>N=</td>
<td>3,213</td>
<td>3,572</td>
<td>6,785</td>
</tr>
</tbody>
</table>

Note: In many cases, it is advisable to weight analysis results before reporting them. Correct weighting requires careful consideration; please always consult the weighting procedures of the study before applying the weights. To apply weights, select the Weight icon and choose the weight variable to be used. All results need careful interpretation. The data collectors and the data producers bear no responsibility for the analysis and interpretation of the data.

[q6] What level of personal information?: Categories
[qe] Gender: Categories
Type: Column percentage

<table>
<thead>
<tr>
<th>[qe] Gender</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>[q6] What level of trust do you have that private companies will protect your personal information?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very high level of trust</td>
<td>7.6</td>
<td>7.8</td>
<td>7.7</td>
</tr>
<tr>
<td>Reasonably high level of trust</td>
<td>33.8</td>
<td>37.0</td>
<td>35.5</td>
</tr>
<tr>
<td>Fairly low level of trust</td>
<td>33.4</td>
<td>33.0</td>
<td>33.2</td>
</tr>
<tr>
<td>Very low level of trust</td>
<td>25.2</td>
<td>22.2</td>
<td>23.7</td>
</tr>
<tr>
<td>Total</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>N=</td>
<td>3,260.6</td>
<td>3,529.2</td>
<td>6,789.7</td>
</tr>
</tbody>
</table>

Weight is on
How do **YOU** store or backup your data?

[Google Logo]

**Gmail**
A Google approach to email.

www.google.com
http://www.colourbox.com/image/an-image-of-treasure-chest-on-white-background-image-3343371
“Data about data”

- Documentation in *standardized* and *structured* form
- Ideally *computer-ready*
- *Explains* your data: origin, purpose, time reference, geographic location, creator, access conditions and terms of use of a data collection.

You as the researcher are the best person to answer questions about your data!
Metadata Standards

**General**
- Dublin Core (DC)
- Metadata Object Description Schema (MODS)

**Humanities**
- Text Encoding Initiative (TEI)
- Visual Resources Association Core (VRA)

**Social Science**
- Data Documentation Initiative (DDI)

**Archives/Repositories**
- D-Space Metadata
- DataStaR Minimum Metadata

**Natural History**
- Darwin Core
- Integrated Taxonomic Information System (ITIS)

**Geographic/Geospatial**
- Federal Geographic Data Committee (FGDC)
- ISO 19115
- Geospatial Interoperability Framework (GIF)

**Earth Science**
- Directory Interchange Format (DIF)
- Standard for the Exchange of Earthquake Data (SEED)

**Ecology**
- Ecological Metadata Language (EML)
The Data Continuum

Statistics
Tables, charts, graphs...
Open to the world

Microdata
- Public-Use Microdata Files (PUMFs)
- Responses of individuals
- Anonymity of respondents is assured
- Generally free
- Under license from Statistics Canada (DLI)

Custom Tabulations
- Generated on request by Statistics Canada using unabridged Master files
- Potential for more detailed tables
- Tables vetted for anonymity
- Cost recovery from Statistics Canada $$$

Real-Time Remote Access (RTRA)
- Desktop analysis of unabridged Master files
- Potential for more detailed tables
- AUTOMATIC table vetting
- Statistics Canada charges annual fee for access $$$
  (this service is under consideration)

Research Data Centre (RDC)
- Controlled access to unabridged Master files
- Potential for more detailed tables & analysis
- Tables vetted by Statistical Analyst for anonymity
- No charge to researchers

ODESI
Dataverse
QSpace
Queen’s Nesstar
Other archives

A lot of research data fits in here... but other discipline-based archives exist

QUEEN’S UNIVERSITY LIBRARY
Examples of archived data

Research Data Management at Queen's University

Overview Writing a Data Management Plan Funding Agency Guidelines Metadata Data Repositories & Archives Citing Data

Best Practices in Data Management Library as Data Partner QUL Research Data Archive Contact

QUL Research Data Archive Comments(0) Print Page

Search:

Data Archive Projects

Bala Parental Alienation Study: Canada, United Kingdom, and Australia, 1984-2012
Using databases of judicial decisions from Canada, Australia, and the United Kingdom, from 1984-2012, searching for the terms "parental alienation", "alienated child", "alienated", etc., cases were combed through to determine whether they represented cases where the court dealt with an issue of alienation. Alienation is defined as a situation where one parent undermines the child’s relationship with the other parent resulting in the child’s rejection of that parent based not on the child’s own experiences but rather the attitude of the alienating parent. Cases found to have dealt with parental alienation were studied and analyzed.

Lead Researcher: Nicholas Bala, Faculty of Law, Queen’s University

Access: QSpace | Dataverse

Baldwin-Green Study: Canada-U.S. Census of Industry, 1867-1940
This study matches Canadian and US manufacturing industries at the 2-digit SIC code level for census years 1900 to 1940. Canadian figures start at 1870.

Lead Researchers: Alan Green, John Baldwin, Department of Economics, Queen’s University

Access: QSpace | Dataverse

Canadian Relocation Cases: Heading Towards Guidelines, 2011
Using Westlaw and Quicklaw, Canadian relocation cases (where one parent wants to relocate with a child, following separation) reported in English from 2001 to the beginning of 2011 (738 in total) were identified and analyzed.

Lead Researcher: Nicholas Bala, Faculty of Law, Queen’s University

Access: QSpace | Dataverse

http://guides.library.queensu.ca/content.php?pid=437556&sid=3637637
Some early examples...

Baldwin-Green Study - Canada-US
Census of Industry 1867-1940

1994

Historical Canadian Macroeconomic Dataset 1871 - 1994

2001
The Globalization of Personal Data Project

Studied implications of surveillance and personal data in nine countries: Canada, U.S.A., France, Spain, Hungary, Mexico, Brazil, China, and Japan.

Key element of success was a ‘data manager’ on the project.

QSpace  Dataverse  ODESI
Cranial Nonmetric Traits Database

27 separate databases in Borland Paradox format on an older vintage PC

Reviewed, refined, & enriched coding scheme

Reviewed, refined, & enriched documentation

Used “R” to stack the files.

QSpace

Dataverse
Ideally, however...

We’d like to catch you early...
The Data Life Cycle

Penitentiary study...

Archived Data & Documentation

http://larshagberg.files.wordpress.com/2012/05/lth102.jpg
Online shopping...

Archived Data & Documentation
Mental health...

Pilot data provided

Archived Data & Documentation

http://www.livingroompharmacy.ca/assets/images/contentimages/mental_health15400988XSmall-1.jpg
Contact information

Jeff Moon –
Data and Government Information Librarian
Academic Director, Queen’s Research Data Centre
moonj@queensu.ca
Ext. 77992

Alex Cooper –
Data and Web Support Assistant
coopera@queensu.ca
Ext. 77481

Acknowledgement: Some content developed by Matthew Barabash, Library Intern, Fall 2012