



Data Management and Data Management Plans

...good things to know when working with BC
grant applicants

Sept. 20, 2016

Enid Karr, Sr. Bibliographer for Biology, Earth & Environmental Sciences and Environmental Studies
enid.karr@bc.edu

Barbara Mento, Data/GIS Librarian, Sr. Bibliographer for Computer Science, Economics, Mathematics
barbara.mento@bc.edu

Jane Morris, Scholarly Communication Librarian
jane.morris@bc.edu

Sally Wyman, Sr. Bibliographer for Chemistry, Physics, Environmental Studies
sally.wyman@bc.edu



Why Have a Data Management Plan?

- Fits into “responsible conduct of research” compliance
- Risk of data loss for researcher and the University
- Facilitates fulfillment of requests from others to see individual researcher data
- Preserves understanding of details for later
- Shared data (“open access”) → higher citation rate!



Many grants now require this (as you know!), but it's bigger than that:

- **More scholarly publishers do, too, now:**
Nature, Science, PLoS, American Economic Review...
require that data must be
 - clearly documented
 - available for sharing
 - detailed enough to permit replication of analysis
- **New “data” journals have appeared, publishing only datasets**
- **There are even indexes that focus on data** (Data Citation Index, from Thomson-Reuters)



Elements of a “Typical” Data Management Plan (DMP)

1-2 pages describing the project and how data will be:

- **Collected (including formats, size, etc.) ... Secured ... Analyzed ... Shared ... Preserved**

Details about access/sharing

- **Potential audience(s) for the data**
- **How access will be provided and how others will find it: “Access” (freely-available) vs. “Sharing” (by request)**
- **Stipulations for privacy, confidentiality, IP or other rights**
- **Allowed re-use of the data, derivative products**

Metadata standards to be used

How long data will be retained -- archiving, long-term preservation and format migration



Key DMP Concepts

1. A “Read me” file or Code Book
2. Use of “open” (non-proprietary) file formats
3. Consistent naming practices for all files
4. Metadata
5. Back up plan
6. Long-term storage strategy
7. Data sharing
8. Plan for true “archiving”



Key DMP Concept #1

A “Readme” file or “Code Book”

This file (or document) describes the research process for collecting data, how it is stored, how it is backed-up and file formats chosen ... and more, as described, below.



Key DMP Concept #2

Use of “open” file formats, avoiding proprietary formats

Whenever possible, researchers should save data using open standards. Some examples:

- TXT, PDF/PDF Archival, not Word (doc, docx)
- ASCII, not Excel (xls,xlsx)
- MPEG-4, not Quicktime (qtff)
- TIFF or JPEG2000, not GIF or JPG
- XML or RDF, not RDBMS

Ideally, files are saved in both original format AND one of the preferred ones listed above.

Why Use Open File Formats?

- No restrictions on their use
- Open source code → future migration easier
- Propriety formats are offered by companies that may go out of business, carrying the code knowledge with them
- Facilitates sharing

Key DMP Concept #3



Consistent naming practices for all files

- **File names should be brief and unique, and might contain:**

Project acronyms, research initials, file type information, version, date, file status, like this one:

Internet Usage Study version 2, Sept. 2011, final draft, in csv format:

IUS_v02_092011_final.csv

- **Evidence of maintenance of both archival (unmodified) and updated “versioned” files (clearly labelled)**

Key DMP Concept #4



Metadata

Metadata is “data about the research data”

This “data” (subject-based terminology) helps others discover the researcher data (more about this, shortly) ...

- Metadata standards exist (per discipline, type of data, individual repository, etc.)
- Metadata is recorded in the “readme” file or code book

Data Documentation (“Metadata”)

- **Metadata captures the most critical information about a particular project. Best when captured early on... helps jog memories later ...**
- **It helps others discover the research being shared.**
- **Metadata may be required for journal publication/data deposit.**

ISO suggested Minimum Data Elements

- Title
- Creator (Principal Investigators)
- Date Created (also versions)
- Instrument and model
- Format (and software required)
- Subject
- Unique Identifier
- Description of the specific data resource
- Coverage of the data (spatial or temporal)
- Publishing Organization
- Type of Resource
- Rights
- Funding or Grant

For help, contact the subject specialist
www.bc.edu/libraries/help/askalib.html

Sample GenBank Record – example of a standard

```
LOCUS      SCU49845    5028 bp    DNA            PLN            21-JUN-1999
DEFINITION Saccharomyces cerevisiae TCP1-beta gene, partial cds, and Axl2p
            (AXL2) and Rev7p (REV7) genes, complete cds.
ACCESSION  U49845
VERSION    U49845.1  GI:1293613
KEYWORDS   .
SOURCE     Saccharomyces cerevisiae (baker's yeast)
  ORGANISM Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE  1 (bases 1 to 5028)
  AUTHORS  Torpey,L.E., Gibbs,P.E., Nelson,J. and Lawrence,C.W.
  TITLE    Cloning and sequence of REV7, a gene whose function is required for
            DNA damage-induced mutagenesis in Saccharomyces cerevisiae
  JOURNAL  Yeast 10 (11), 1503-1509 (1994)
  PUBMED   7871890
REFERENCE  2 (bases 1 to 5028)
  AUTHORS  Roemer,T., Madden,K., Chang,J. and Snyder,M.
  TITLE    Selection of axial growth sites in yeast requires Axl2p, a novel
            plasma membrane glycoprotein
  JOURNAL  Genes Dev. 10 (7), 777-793 (1996)
  PUBMED   8846915
REFERENCE  3 (bases 1 to 5028)
  AUTHORS  Roemer,T.
  TITLE    Direct Submission
```



Key DMP Concept #5

Back up Strategy

Regular, scheduled back up protects against loss

- **Back up strategy will depend on user needs:**
 - **Back up all versions of the files or certain ones?**
 - **How often to back up files?**
- **Listing at least two back up locations (so, 3 copies)**
 - **Internal (researcher computer)**
 - **External (i.e. the BC Research Data Archive or departmental servers)**
- **Assign responsibility for backing up**



Key DMP Concept #6

Long-term Storage Strategy

Plan should describe how data will be stored ... in the safest long-term locations (not a laptop or flash-drive!)

- **Local (lab computer, flash-drive):** convenient, but much less secure
- **Centralized – ITS Servers, departmental servers**
- **Remote** – disciplinary servers (GenBank, ICPSR, etc.) most tailored to disciplinary needs, but may be open (and that may be problematic for some researchers)



Grants can sometimes cover cost of long-term storage.

Key DMP Concept #7



Data Sharing – ... the ultimate goal of DMPs

Options include: personal website ... but researchers can do better:

- Journal “supplementary materials” (ACS, etc.) ... now in figshare 
- Institutional repository, e.g. eScholarship@bc, BC Dataverse
- Disciplinary (or multidisciplinary) repository 
- Or, a combination: journal-designated repository ([Nature](#), for example)



Examples of Subject Repositories

- **Biomedicine:**
 - GenBank* -- sequence data
 - RSCB Protein DataBank* -- biomolecule crystal structure coordinates, etc.
- **Chemistry:**
 - Cambridge Structural Database (CSD)*
 - PubChem (Part of NCBI Entrez, covering biological activities of small molecules)
- **Social Sciences**
 - ICPSR (Inter-university Consortium for Political and Social Research)
 - IQSS (The Institute for Quantitative Social Science)
- **Multidisciplinary: [figshare.com](https://www.figshare.com) (Open, Free)**
- *A few of the data repositories that fulfill *Science* magazine requirements for data deposition

Key DMP Concept #8



Archiving Plans

Archiving Data means not just preserving the data in the original format *but also* in a format that is non-platform reliant, using a standard that ensures that the data can be re-used in the future.



Additional Essential Elements in the DMP:



- **Ethics/privacy**
- **Data Ownership**
- **Intellectual Property/Technology Transfer**

Ethics and Privacy

- Sensitive data should be redacted before depositing in a public archive or repository.
- Access to data may be embargoed (access limited for a time) for confidentiality, legal, patentability or other reasons.
- Dark archives ensure permanent protection of confidentiality.
- Where human subjects/privacy is involved, BC's Institutional Review Board (IRB) must approve.
<http://www.bc.edu/research/oric/human.html>



Image: digitalart / FreeDigitalPhotos.net



Good DMP, then What Happens? Data is Shared, then ... Cited

Data Citation

Why is proper data citation important?

- Ensures that original producers of the data are credited in citation indexes*
- Allows researchers to locate research data used in an article
- May be required by the archive that stores the data to be repurposed

*Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3): e308.doi:10.1371/journal.pone.0000308

Citing Data Sets

Essential citation elements; style will vary:

- author or creator
- title or description
- year of publication
- publisher and/or the database/archive from which it was retrieved
- the URL or DOI if the data set is online

National Center for Biotechnology Information. PubChem Compound Database; CID=5934766, <http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=5934766> (accessed Feb. 22, 2011).

Mackey, R.A., Mackey, E.F., and O'Brien, B.A. (1990). *Lasting relationships research data archive* (eScholarship version) [Data file]. Boston College School of Social Work. <http://hdl.handle.net/2345/2228>

Questions?

Library Tools and Resources to Help

- Barbara, Enid, Jane and Sally! We are here to help.
- Research Guide for Data Management Plans
<http://libguides.bc.edu/dataplan>
- The BC DMP Tool DMPTool.org
- Dataverse <http://libguides.bc.edu/dataverse>
- E-Scholarship@bc <http://dlib.bc.edu/>
- ORCID – for BC researcher identifier <http://orcid.org>

eScholarship@BC

eScholarship@BC
MORE REACH FOR YOUR RESEARCH

[Home](#) [About](#) [Deposit](#) [My Bookmarks](#) [Help](#)

Search this repository

[Advanced Search](#)



Browse Content

[Faculty Works](#)
[Research Centers](#)
[Data Archive](#)
[Graduate Theses and Dissertations](#)
[Undergraduate Honors Theses](#)

Recent Additions

Analyzing language policy and social identification across heterogeneous scales
Clearing away assumptions through philosophy and research
Low fractal dimension cluster-dilute soot aggregates from a premixed flame

Most Viewed

1. Dramatizing the Sura of Joseph I eScholarship@BC
2. Merleau-Ponty's Phenomenology of Painting, Gestalt, and Reversibility I eScholarship@BC
3. Boston College tree inventory data archive I eScholarship@BC

eScholarship@BC is our institutional repository

- Faculty can deposit scholarly work including
 - Working papers, published articles, teaching materials, conference presentations, posters
- Reasons to deposit:
 - Compliance with **funder mandates** for open access
 - Global visibility and readership
 - Search engine harvesting
 - Eliminates economic barriers to knowledge
 - Increase citation counts
 - Get a permanent URL for the CV
 - Long-term preservation



Connecting Research
and Researchers

What is ORCID?

- ✓ Unique, persistent identifier for researchers & scholars
- ✓ Non-profit organization supporting linkages between systems

Benefits to you?

- ✓ Improves discoverability
- ✓ Connects your work
- ✓ Eliminates name ambiguity
- ✓ Stays with you throughout your career

Get your free ORCID iD at <http://orcid.org>

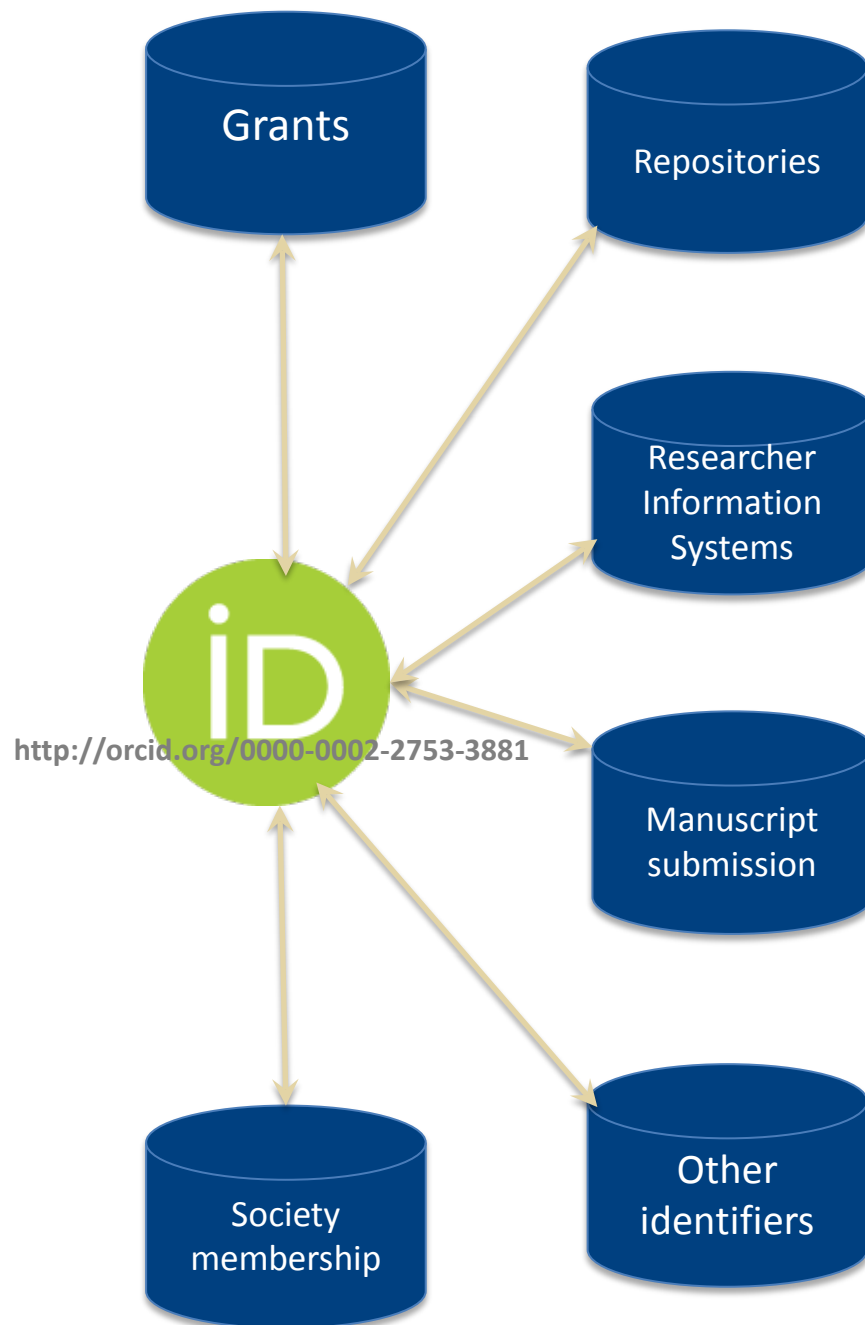
ORCID

Connecting Research
and Researchers

Use your ORCID iD!

- ✓ Manuscript submission
- ✓ Grant applications
- ✓ Professional society membership
- ✓ Link with other identifiers & profiles
- ✓ Display on your CV, web page, and more

It takes 30 seconds to claim your ORCID
iD at <http://orcid.org/>





Connecting Research
and Researchers

FOR RESEARCHERS

FOR ORGANIZATIONS

ABOUT

HELP

SIGN OUT

MY ORCID RECORD

INBOX (8)

ACCOUNT SETTINGS

DEVELOPER TOOLS

LEARN MORE

2,316,860 ORCID iDs and counting. See more...

**Kimberly C.
Kowal**

ORCID ID

orcid.org/0000-0003-4659-2307

[View public version](#)

Get a QR Code for your iD ?

Also known as

Country

Keywords

Websites

Emails

kowalk@bc.edu

Other IDs

Scopus Author ID 10240529000 🗑️

Biography



▼ Education (0)

+ Add education

⇅ Sort

You haven't added any education, [add some now](#)

▼ Employment (2)

+ Add employment

⇅ Sort

British Library: London, United Kingdom

2004 to 2015



Source: Kimberly C. Kowal

Created: 2016-02-26



Boston College: Chestnut Hill, MA, United States

to present



Associate University Librarian, Digital Initiatives & Services

Source: Kimberly C. Kowal

Created: 2016-02-26



▼ Funding (0)

+ Add funding

⇅ Sort

You haven't added any funding, [add some now](#)

▼ Works (8)

+ Add works

Bulk edit

⇅ Sort

Georeferencer: Crowdsourced Georeferencing for Map Library Collections



D-Lib Magazine

2012-11 | journal-article

DOI: [10.1045/november2012-fleet](https://doi.org/10.1045/november2012-fleet)

Source: CrossRef Metadata Search

Preferred source (of 2)



Online georeferencing for libraries: The British Library



[Search](#)
[Alerts](#)
[Lists](#)
[My Scopus](#)

[Print](#) | [E-mail](#)

Kowal, Kimberly C.

The British Library, London, United Kingdom

Author ID: 10240529000

<http://orcid.org/0000-0003-4659-2307>

[About Scopus Author Identifier](#) | [View potential author matches](#)

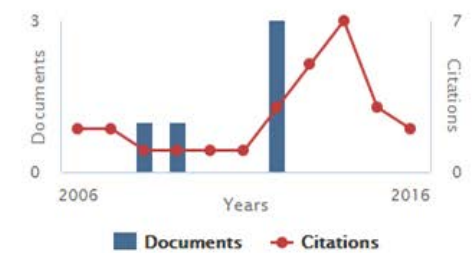
[Follow this Author](#) Receive emails when this author publishes new articles

[Get citation alerts](#)

[Add to ORCID](#)

[Request author detail corrections](#)

Documents: 8 [Analyze author output](#)
 Citations: 34 total citations by 34 documents [View citation overview](#)
 h-index: 3 [View h-graph](#)
 Co-authors: 6
 Subject area: [Social Sciences](#), [Computer Science](#) [View More](#)



[8 Documents](#) | [Cited by 34 documents](#) | [6 co-authors](#)

[8 documents](#) [View in search results format](#) Sort on: [Date](#) [Cited by](#) [...](#)

[Export all](#) | [Add all to list](#) | [Set document alert](#) | [Set document feed](#)

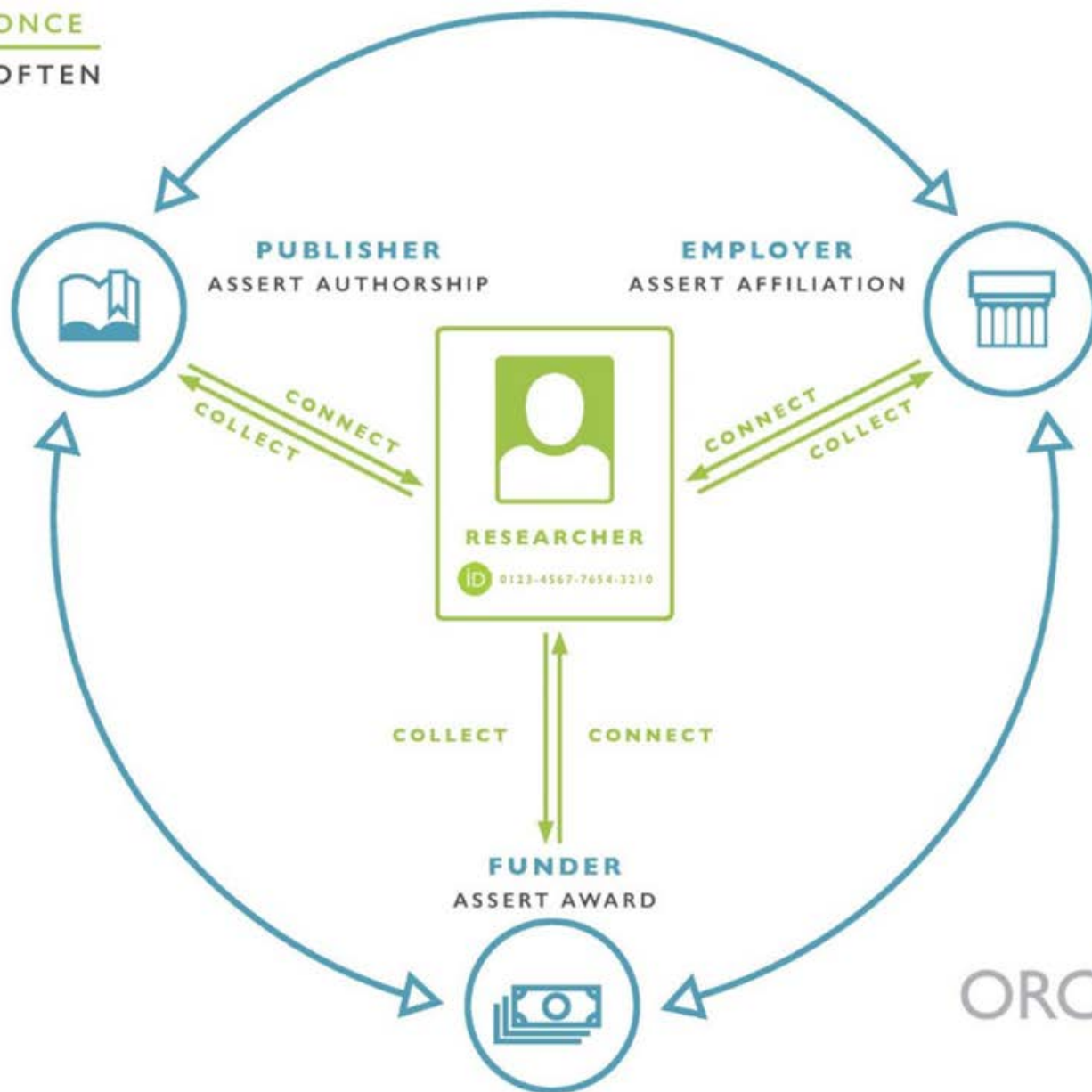
Author History

Publication range: 2002 - 2012
 References: 143
Source history:
[Library Resources and Technical Services](#) [View documents](#)
[Cartographic Journal](#) [View documents](#)
[Imago Mundi](#) [View documents](#)
[View More](#)
[Show Related Affiliations](#)

| | | | | |
|---|------------------------------------|------|--|---|
| Georeferencer: Crowdsourced georeferencing for map library collections | Fleet, C., Kowal, K.C., Pridal, P. | 2012 | D-Lib Magazine | 5 |
| Open Access | | | | |
| Online georeferencing for libraries: The british library implementation of georeferencer for spatial metadata enhancement and public engagement | Kowal, K.C., Prbreveidal, P. | 2012 | Journal of Map and Geography Libraries | 7 |
| The Pembroke Map: A Medieval Sketch | Kowal, K.C. | 2012 | Imago Mundi | 0 |
| Descriptive metadata for digitization of maps in books: a british library project | Kowal, K.C., Martyn, C. | 2009 | Library Resources and Technical Services | 3 |
| The british library's vulnerable collection items project | Kowal, K.C., Rhatigan, | 2008 | LIBER Quarterly | 0 |

INTEROPERABILITY

ENTER ONCE
REUSE OFTEN



What is BC doing?

- BC Libraries have purchased ORCID membership – gives us access to the API
- Working with ITS to develop a Create on Demand App that links to BC identity
- Data 180 will include ORCID integration
- Dissertation authors can add an ORCID to the record for their work in eScholarship@BC

Benefits to BC

- Publishers/funders increasingly require ORCID submission; profiles update automatically
- Access to reliable, accurate, timely information on scholarly output and grants
- Ability to follow graduate students' careers