Demystifying Research Data: An Introductory Session for Librarians as Researchers

Kristin Bogdan
February 23, 2016
What is research data?

- Research data is data that is collected, observed, or created, for purposes of analysis to produce original research results.

  (Boston University, http://www.bu.edu/datamanagement/background/whatisdata/)
What kinds of data have you collected/created?
Why should we share research data?

• “Data or it didn’t happen” – Kristin Briney
Why should we share research data?

• Reproducibility = better scholarship
• Prevent fraud
• Build on other research
• Encourage others to do the same
The Research Data Life-cycle

(Image from the DDI Alliance “What is DDI?”, http://ddialliance.org/what, Accessed February 21, 2016)
Concept

- DMP Assistant – not just for grants!
- ICPSR Guide to Social Science Data Preparation and Archiving
  http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf
Collection

- Research question – data considerations
  - Time
  - Geography
  - Variables
  - Format
  - Source
Collection

• Reusing Data:
  • ICPSR – Ithaka survey data
  • figshare
  • Institutional Repositories: ex. Scholar’s Portal Dataverse
  • Government and intergovernmental organizations (Census, data.gov)
  • Open Data from Library and Archives Canada: http://www.bac-lac.gc.ca/eng/discover/open-data/pages/open-data.aspx
  • Open Data Catalogue from Vancouver Public Library: http://www.vpl.ca/opendata
Collection

• Collecting New Data:
  • Quantitative or qualitative
  • Ethics: [http://research.usask.ca/for-researchers/ethics/](http://research.usask.ca/for-researchers/ethics/)
  • Social Science Research Lab (SSRL) Research and survey design services: [http://ssrl.usask.ca/](http://ssrl.usask.ca/)
    • Qualtrics
Collection

• Collecting Data (continued)
  • DMP as a Research Tool:  
    http://dmpresearch.library.oregonstate.edu/
  • Northwestern University Assessment LibGuide:  
    http://libguides.northwestern.edu/Assessment
  • iSquares: http://www.isquares.info/
Processing

• Some processing steps may include
  • Recode variables
  • Format the data for analysis
  • Look for anomalies
• Keep track of all of the steps (README file, keep code, etc.)
• Keep multiple versions of your data (including the raw)
Processing

• Processing tools
  • OpenRefine
  • Excel – use with caution
  • Databases
  • Statistical software
Distribution

• File management
  • Make a plan
  • Document naming conventions
  • Folders for working and final versions
  • MAKE BACKUPS!

• Documentation
  • Metadata – Data Documentation Initiative
  • Codebook or Data Dictionary
  • Code
Distribution

• Deposit in a repository
• Get a DOI
• Link to website, ORCID profile, online cv...
• Choose terms of use
• Create a citation
Discovery

• Related publications
• Researcher profiles
  • ORCID: http://orcid.org/
• DOIs
• Repositories
Analysis

• Replication code
• SSRL services and software
  • Qualitative: NVivo
  • Quantitative: SPSS, Stata, R
  • Spatial analysis: ArcGIS
Repurposing

- Provide data in multiple types: ISPS Data Archive
- File formats: http://data-archive.ac.uk/create-manage/format/formats-table
Archiving

• Keep the files in multiple places – don’t rely solely on a repository
• Make sure to deidentify sensitive data (ethics)
• Files may require additional processing
Other Resources

- Coursera
- MANTRA
- ICPSR Summer Program
- lynda.com
- Safari Books Online
- SSRL
- ICT Research Technologies & Services
- ownCloud
Discussion