MANAGE OR BE MANAGED: INTRODUCTION TO DATA MANAGEMENT

Graduate Thesis and Dissertation Conference

Kindra Morelock, Data Services Librarian

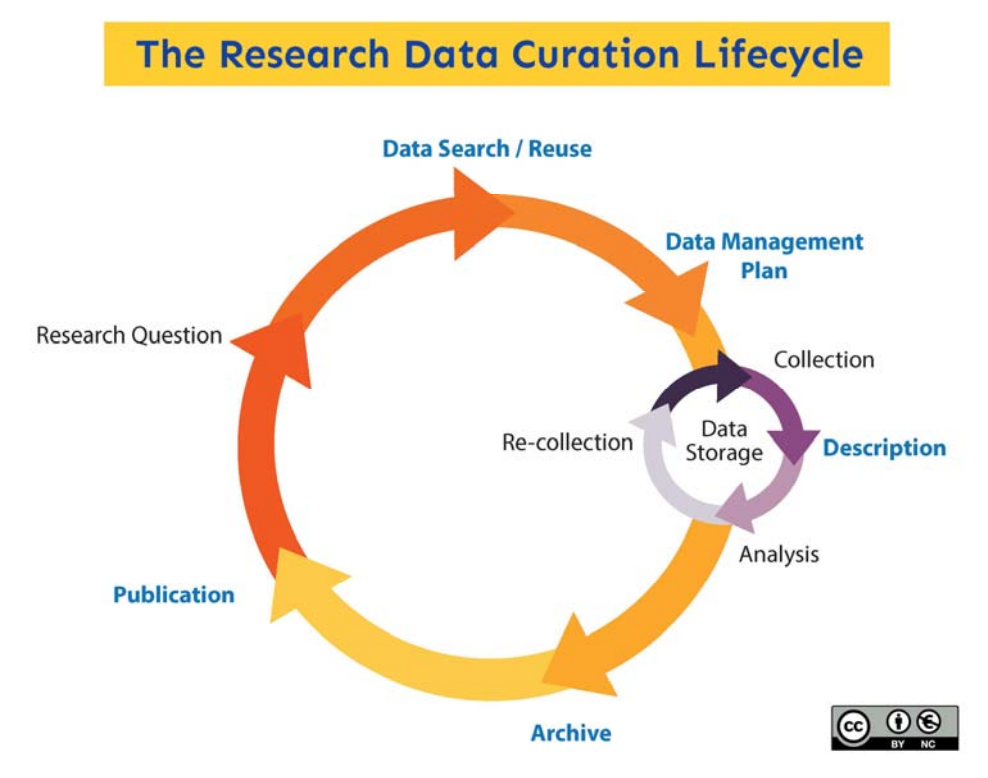
DePaul University

February 3, 2018

PATH AHEAD

- Research Data Lifecycle
- Have a Plan: Data Management Plans
- Get Your Data: Data Collection
- Clean Your Data: Data Cleaning
- Describe Your Data: Metadata Standards
- Share Your Data: Data Repositories

RESEARCH DATA LIFECYCLE



http://guides.library.ucsc.edu/datamanagement/

HAVE A PLAN: DATA MANAGEMENT PLANS (DMP)

Most federal grants and many private grants require **Data Management Plans**

Questions to Consider:

- What kind of data are you creating?
- Where are you storing this data?
- How are you planning on sharing this data?

GET YOUR DATA: COLLECTING YOUR OWN

Pros

- Can tailor to your research needs
- Collect only what you need
- Do not have to interpret someone else's data

Cons

- Time consuming
  - IRB approval
  - Designing instrument
  - Analysis
- Quarter system
- Creating metadata about the data you collected
- File versioning/management

GET YOUR DATA: FINDING AN EXISTING DATASET

Pros

- Time-saving
- Can compare similar datasets
- Metadata has already been created
- Potentially clean data

Cons

- Might have to strip out irrelevant fields
- Can sometimes be difficult to find what you need

GET YOUR DATA: MY RECOMMENDATIONS

- Try finding external datasets first:
  - ICPSR
  - City of Chicago Data Portal
  - Data.gov
  - Figshare.com
- If you are going to collect data, consider:
  - Start the IRB process ASAP
  - Create your collection instrument wisely
  - Consider file naming scheme ahead of time <initials>_<project_name>_<date>.<file extension>
    - KM_mygradschoolproject_20180203.py

CLEAN YOUR DATA: DATA CLEANING/FILE PRE-PROCESSING

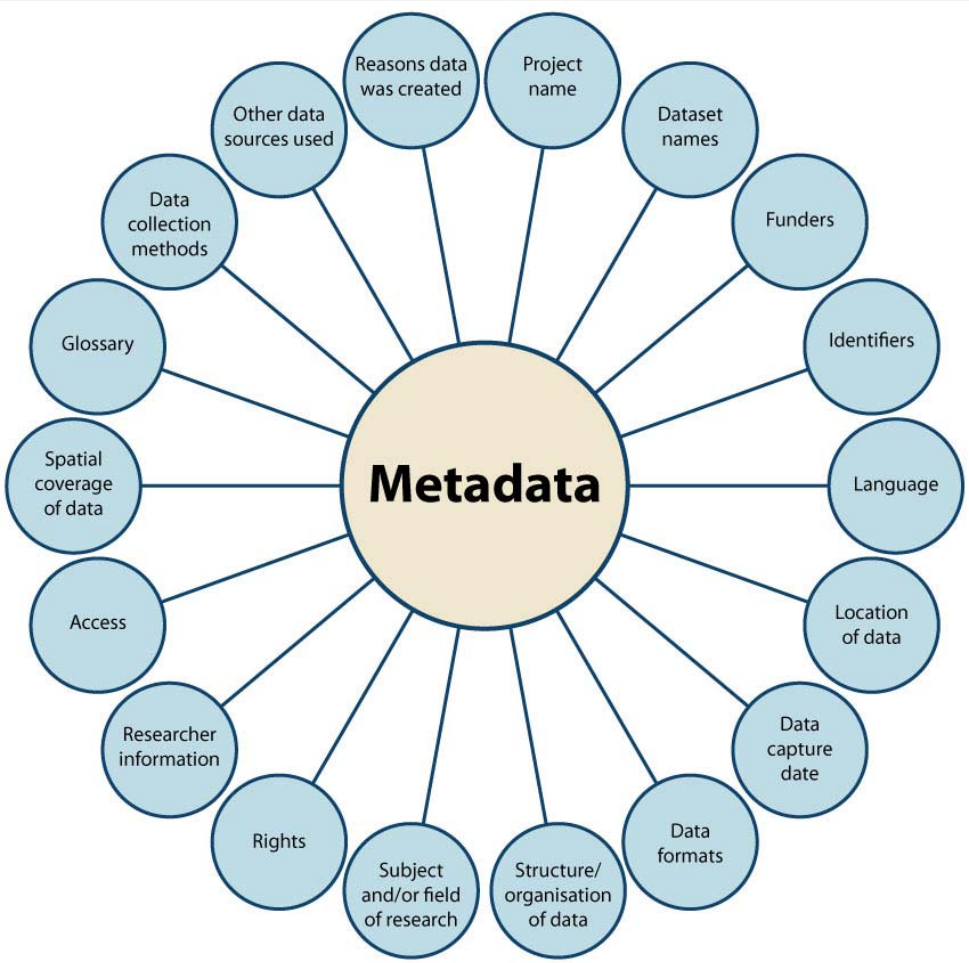
	City	Country	Population	Area	Density
r_1	New York	USA	8734520	6400	26403
r_2	Philadelphia	United States	“1,204,542”	“3,231”	NaN
r_3	New York City	USA	8734520	6400	26403

DESCRIBE YOUR DATA: METADATA

Machine-readable XML format

Discipline-specific standards for describing data:

Standard Name	Discipline
Dublin Core	All Purpose
Darwin Core	Biology
DDI	Humanities, Social Science
IEEE LOM	Education Learning Objects
CDWA	Art & Architecture



Source: http://guides.library.uwa.edu.au/c.php?g=325196&p=2178564

SHARE YOUR DATA: DATA REPOSITORIES

Benefits

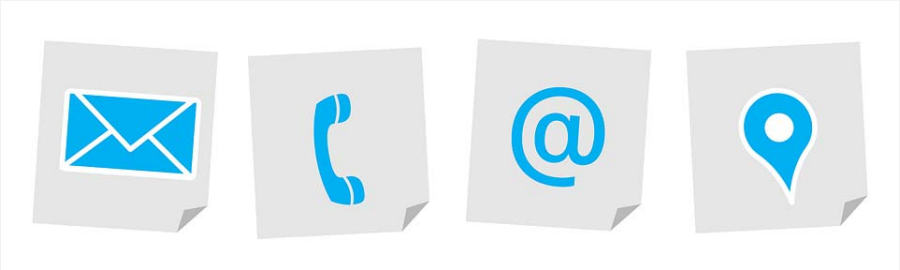
- Long-term preservation
- Discoverability
- Persistent links
- Availability

Examples

Repository	Discipline
ICPSR	Social & Political Science
DataONE	Environmental Sciences
Harvard Dataverse	Multiple
Figshare	Multiple
Via Sapientiae	DePaul's Institutional Repository

QUESTIONS?

KINDRA MORELOCK, DATA SERVICES LIBRARIAN



Contact Information

Email: kmoreloc@depaul.edu

Phone: (773) 325-4668

Campus: Lincoln Park and Loop