USE WITH CAUTION: A MULTI-DISCIPLINARY ANALYSIS OF DATA USE AND ACCESS CONDITIONS

Tiffany C. Chao
Graduate School of Library and Information Science
University of Illinois
IAASSIST 2013
ROADMAP

• Introduction
• Background: sharing practices in the sciences
• Methodological Design
• Preliminary results & Discussion
• Future Directions
ACCESS TO RESEARCH DATA

Increased expectations by US federal agencies and offices for open access to both the publications and data generated through research

Growing number of journals that require authors to provide access to data as part of publication


Emerging Institution-level support and services for research data management throughout their lifecycle

- DMP Tool (https://dmp.cdlib.org/)
- Data Curation Profiles Toolkit (http://datacurationprofiles.org)
Disciplinary norms and research cultures influence sharing practice:

**What data are shared**
- Digital data from sensors more likely to be shared than hand-processed data in habitat ecology research (Borgman et al., 2007); “raw” data are not typically shared or deposited within agricultural sciences (Diekmann, 2012).

**When are data shared**
- Scientists opt for post-publication or prescribed embargo period to release data to the public; very rarely are data shared prior to publication (Cragin et al., 2010)

**How data are made available**
- Greater instances of Informal (i.e. personal contact, website) sharing compared to formal (i.e. domain repository, archive) mechanisms in the Social Sciences (Pienta et al., 2010); the norm for genomics researchers is to submit to a data repository (i.e. GenBank) (Swan & Brown, 2008).

**Who has access to the data**
- Some researchers make data available only to those with whom they have worked closely with (i.e. research group, collaborators); those in physical science were more committed to open data sharing than other disciplines (PARSE, 2009)

Across disciplines, scientists are willing to share some of their data with third party researchers only with conditions in place (Tenopir et al., 2011)
CONTINUUM FOR DATA ACCESS

Data curation continuum model for “access” (Treloar et al., 2007)

**Closed Access**
- Sensitive information
- Commercial/industry connections
- Intentional withholding (Campbell et al., 2002; Blumenthal et al., 2006)

**Open Access**
- “Open access to research data from public funding should be easy, timely, user-friendly and preferably Internet-based.” (OECD, 2007)
- Available after embargo period/post-publication
- No restrictions
CONTINUUM FOR DATA ACCESS

Data curation continuum model for “access” (Treloar et al., 2007)

Closed access
- Sensitive information
- Commercial/industry connections
- Intentional withholding (Campbell et al., 2002; Blumenthal et al., 2006)

Open Access
- No restrictions
- Available after embargo period/post-publication
- “Open access to research data from public funding should be easy, timely, user-friendly and preferably Internet-based.” (OECD, 2007)
THE SPACE BETWEEN: CONDITIONS FOR DATA ACCESS

Closed Access

• collaboration opportunity;
• mandatory reprints provided;
• co-authorship;
• results of analyses need data providers’ approval prior to dissemination;
• cost recovery;
• legal permissions obtained

Open Access
THE SPACE BETWEEN: CONDITIONS FOR DATA ACCESS

Closed Access

• collaboration opportunity;
• mandatory reprints provided;
• co-authorship;
• results of analyses need data providers’ approval prior to dissemination
• cost recovery;
• legal permissions obtained

Open Access

Environmental Science & Ecology

Physical Sciences

Social Sciences

Conditions for sharing data (Tenopir et al, 2011)
THE SPACE BETWEEN: CONDITIONS FOR DATA ACCESS

Conditions for sharing data (Tenopir et al, 2011)

- collaboration opportunity;
- mandatory reprints provided;
- co-authorship;
- results of analyses need data providers’ approval prior to dissemination;
- cost recovery;
- legal permissions obtained
PROPOSED INQUIRY

While researchers are willing to share their data, it is not clear how intended conditions are presented to third party researchers interested in reuse

• What conditions for use and access of data are made visible through metadata description?

• What similarities and differences exist for use and access conditions of data across disciplines?
METHODOLOGICAL DESIGN OVERVIEW

Sample Scope
(3) disciplines/domain areas:
- Geochemistry
- Environmental Impacts
- Population Studies

Source
Dataset metadata records from GCMD (Global Change Master Directory); follows DIF format (Directory Interchange Format)
Collected in Fall 2012.
Includes fields for <access constraints> and <use constraints>

Approach
Extracted descriptions for each dataset record.
Applied analytic framework for available <use> and <access> constraints.
Reviewed initial framework codes and recoded descriptions with new list.
DISCIPLINES/ DOMAIN AREAS

Geochemistry

• Involves the study of the chemical composition along with physical and chemical processes at work in the formation of the Earth

• Examples of data content: biogeochemistry, hydration, ion exchange

Environmental Science: Human Dimensions (ESHD)

• Relates the study of interactions and interdependencies of humans and their physical and social environments

• Examples of data content: GIS, air quality, chemical traces, vegetation species

Demography/ Population Studies (DPS)

• Concerns the number of individuals in a specified area that constitute a particular race, class, or group

• Examples of data content: mortality rate, population distribution, population density
Access Constraints Data Accessible 2 Years After First Collected

Use Constraints Her Majesty the Queen in Right of Canada 2004. For more information, please contact the Copyright Officer at the following address: ESS Copyright Officer Business Development Office Earth Sciences Sector 615 Booth St., Ottawa, Ontario K1A 0E9
DIF DEFINITIONS OF ACCESS AND USE CONSTRAINTS

The <Access_Constraints> field allows the author to provide information about any constraints for accessing the data set. This includes any special restrictions, legal prerequisites, limitations and/or warnings on obtaining the data set. Some words that may be used in this field include: Public, In-house, Limited, Additional. Detailed instructions on how to access the data can be entered in this field.

The <Use_Constraints> field allows the author to describe how the data may or may not be used after access is granted to assure the protection of privacy or intellectual property. This includes any special restrictions, legal prerequisites, terms and conditions, and/or limitations on using the data set. Data providers may request acknowledgement of the data from users and claim no responsibility for quality and completeness of data.

DDI DESCRIPTION FOR USE AND ACCESS CONDITIONS

<Conditions>

Indicates any additional information that will assist the user in understanding the access and use conditions of the data collection.

Example:

The data are available without restriction. Potential users of these datasets are advised, however, to contact the original principal investigator Dr. J. Smith (Institute for Social Research, The University of Michigan, Box 1248, Ann Arbor, MI 48106), about their intended uses of the data. Dr. Smith would also appreciate receiving copies of reports based on the datasets.
1.7 Access Constraints

**Definition:** Restrictions and legal prerequisites for accessing the dataset. These include any access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the dataset.

**Example:**
- *Access Constraints:* None
- *Access Constraints:* CIESIN offers unrestricted access and use of data without charge, unless specified in the documentation for particular data. All other rights are reserved.

1.8 Use Constraints

**Definition:** Restrictions and legal prerequisites for using the dataset after access is granted. These include any use constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on using the dataset.

**Example:**
- *Use Constraints:* None
- *Use Constraints:* The Wildlife Conservation Society (WCS) and Trustees of Columbia University in the City of New York hold the copyright of this dataset. Users are prohibited from any commercial, non-free resale, or redistribution without explicit written permission from WCS or CIESIN. Users should acknowledge WCS and CIESIN as the source used in the creation of any reports, publications, new datasets, derived products, or services resulting from the use of this dataset. WCS or CIESIN also request reprints of any publications and notification of any redistributing efforts.
ANALYTICAL FRAMEWORK

ACCESS CONDITIONS

• special restrictions, legal prerequisites, limitations and/or warnings

• terms to use: Public; In-house; Limited

• instructions on how to access the data

USE CONDITIONS

• special restrictions, legal prerequisites, terms and conditions, and/or limitations

• data providers may request acknowledgement of the data from users

• data providers may claim no responsibility for quality and completeness of data
PRELIMINARY RESULTS & DISCUSSION

• Observed characterizations of Access conditions/ Use conditions
• Comparative analysis of conditions across disciplines
• Relationship between conditions and sharing practices and perceptions
DISTRIBUTION OF AVAILABLE METADATA RECORDS WITH INFORMATION ON ACCESS AND/OR USE CONDITIONS

% of metadata records

Geochemistry  DPS  ESHD

No information  Only Use  Only Access  Both Access & Use
TYPES OF ACCESS CONDITIONS

• **Users**: access limited to certain personnel

• **Permission**: contact investigator and request access data

• **Fees**: account subscriptions for data download

• **License agreement**: acknowledge and agree to conditions

• **Retrieval**: reference to where and how data can be accessed

• **Embargo period**: not all data are publically available yet
TYPES OF USE CONDITIONS

• **Proper acknowledgement**: includes specific citation to use or attribution statement
• **Disclaimer**: statements regarding accuracy of available data; providers cannot “guarantee” quality of data
• **Sensitivity**: respect for subject confidentiality
• **Copyright**: data cannot be used for commercial purposes without consent of provider; must secure written permission
• **Publication restrictions**: states who can publish
• **Fees**: assessed for data but not the metadata, which was freely available
## Top 3 Conditions for Each Discipline/ Domain Area

<table>
<thead>
<tr>
<th>Discipline/ Domain Area</th>
<th>Access Conditions</th>
<th>Use Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demography/ Population Studies</td>
<td>None (80.65%)</td>
<td>None (61.11%)</td>
</tr>
<tr>
<td></td>
<td>Retrieval (9.68%)</td>
<td>Copyright/Disclaimer (22.22%)</td>
</tr>
<tr>
<td></td>
<td>Limited use (6.45%)</td>
<td>Proper acknowledgment (11.11%)</td>
</tr>
<tr>
<td>Environmental Science: Human Dimensions</td>
<td>None (70%)</td>
<td>Copyright (34.78%)</td>
</tr>
<tr>
<td></td>
<td>Retrieval (15%)</td>
<td>License (26.09%)</td>
</tr>
<tr>
<td></td>
<td>Embargo (10%)</td>
<td>None (13.04%)</td>
</tr>
<tr>
<td>Geochemistry</td>
<td>None (40.76%)</td>
<td>Proper acknowledgment (31.67%)</td>
</tr>
<tr>
<td></td>
<td>Limited use (16.92%)</td>
<td>None (26.67%)</td>
</tr>
<tr>
<td></td>
<td>Retrieval (13.86%)</td>
<td>License (17.5%)</td>
</tr>
</tbody>
</table>
COMPARISON OF CONDITIONS WITH KNOWN SHARING PERCEPTIONS

Sharing conditions from Tenopir et al. (2011)

- collaboration opportunity;
- mandatory reprints provided;
- co-authorship;
- legal permissions obtained

From metadata records:

- Geochemistry
  - No Access conditions; Use conditions center on proper acknowledgement
- Environmental Science: Human dimensions
  - No Access conditions; Use conditions focus on copyright and license agreements
- Demography/Population Studies
  - No Access or Use conditions
CONTINUUM OF DATA ACCESS: RE-VISITED

**Closed Access**
- Data only available to certain personnel

**Open Access**
- No conditions or constraints

---

**Major Conditions:**
- Permission from data provider; Fees/license agreement

**Minor Conditions:**
- Attribution to data provider; Retrieval from specified space; Embargo period
FUTURE DIRECTIONS

1) Identify how conditions are conveyed in informal sharing venues

2) Relation between data type and conditions
   - Preliminary analysis of records from the 3 disciplines indicate PDF most prevalent format for distribution of data
   - No pattern in the type of data associated with particular constraints

3) Assess relevance of condition information to decisions regarding reuse of data
SUMMARY & CONCLUSIONS

• Current metadata schema recognize information for data access and use conditions, yet the actual inclusion of conditions vary.

• There are multiple types of conditions for access and use of data, with some more prevalent across different disciplines than others.

• Potential implications for design of data repository system and services to accommodate different conditions and facilitate sharing.
Thank you!

Contact: Tiffany Chao; tchao at illinois.edu

Acknowledgments:
Special thanks to Brittany Smith and Peter Sheh
REFERENCES


