Where DH and Data Services Meet

Heather Tompkins
Carleton College
IASSIST 2015
Source: “"Washington Avenue at Nicollet Avenue, Minneapolis, Minnesota" by A.D. Roth made available by the Minnesota Streetcar Museum via Minnesota Reflections.
All sessions are noon – 1 PM in the Larson Meeting Room (Weitz 236), unless otherwise noted. Lunch is for 50 served beginning at 11:45.

TUESDAY JANUARY 7
Translating Experience: A Globally Engaged Curriculum

Learn more about plans and opportunities for international work and study, foreign language across the curriculum (FLAC), and advanced interdisciplinary work. Especially useful for those preparing applications for Gil summer grants (due Feb. 7). Students are welcome to attend.

Scott Carpenter, Co-Director of Global Engagement; Timothy Poitras, Professor of Political Science

TUESDAY JANUARY 28
A Coach, an Artist, and a Scientist Walk into a Circle

Three faculty members from widely different disciplines (the studio, the court, the key [blackboard]) walk into a circle. Despite the obvious superficial differences, they encounter surprising commonalities and shared challenges for their students in those vastly different environments: how failure, about spatial reasoning and visual language, about how a group can genuinely become a team. In this session, they’ll share some of these communities—and a few of the ways in which they all judge each other. Students are welcome to attend.

Kelly Connolly, Associate Professor of Art; Nolde Jesensky, Professor of Physical Education, Athletics, and Recreation; David Penney, Assistant Professor of English

TUESDAY FEBRUARY 18
Lifelogging and Other Data Visualizations: Art or Science?

Lifelogging is the practice of collecting personal data (e.g., photos, location, and activity data) and sharing it with others. In this talk, we will explore how data visualization can be used to present the “big data” collected for personal or social projects, and analyze, visualize, manipulate, and otherwise organize for public consumption. Co-sponsored by the Perman Teaching and Quirk. Students are welcome to attend.

Laurie Bradley, Director and Curator of the Perman Teaching Museum; David Lefebvre, Associate Professor of Art; Doug Ingraham, Communication and Training Coordinator; Information Technology Services; Jeff Ondr, Professor of Mathematics and Computer Science.

DIGITAL HUMANITIES
UPDATES FROM THE DH ASSOCIATES AT CARLETON COLLEGE

ABOUT US
We are the Digital Humanities Associates at Carleton College. Our job is to collaborate with faculty to develop digital humanities projects.

Staff Bios
Data, Datsets, and Statistical Resources

Librarians for Data

Reference librarians can help you find data. They can help you form a strategy for your search for data, discuss your research question, and access data sources. To reach them:

- Drop in at the Research/IT Desk
- Contact your liaison librarian for discipline-specific data searches (e.g., help finding data on the environment)

Social Science Librarians can help with more involved data search questions.

Danya Leebaw

Environmental Studies, International Relations, Middle Eastern Languages, Political Economy, Political Science, and Government Documents

- Make an appointment or chat
- Email: dlee [at] carleton [dot] edu
- Phone: x5179
- Office: Libe 465

Kristin Partlo

Economics, Mathematics, Political Economy, and Sociology & Anthropology

- Make an appointment or chat
- Email: kpartlo [at] carleton [dot] edu
- Phone: x7668
- Office: Libe 466

Kristin Partlo is the library liaison for data. Contact her with suggestions and questions about the data collection.
Stuck?

Not finding what you need? Search for articles on your topic and pay attention to where they get their data. Then look for things published online or in print by the same groups that produced the data your articles reference. Here are some great places to find articles that use numerical data.

PAIS (includes French Language Articles)

An index to books, documents, reports, and articles in periodicals concerning public policy.

Sociological Abstracts (includes French Language resources)

This is the best place to find articles about current and historical aspects of French society.

ProQuest Statistical

Index to statistics produced by federal agencies, states, and intergovernmental organizations.

The Biggies

INSEE

Official statistics from France on everything from health to trade.

Eurostat

Published by the Statistical Office of the European Commission, this covers statistics from all members of the European Union.

Specific Statistical Topics

Europa World Plus

The online version of the Europa World Year Book and the nine-volume Europa Regional Surveys of the World series. Provides up-to-date political and economic information for more than 250 countries and territories.

ProQuest Statistical

Index to statistics produced by federal agencies, states, and intergovernmental organizations.

CIA World Factbook

Information on the history, people, government, economy, geography, communications, transportation, military, and transnational issues for 266 world entities.

NationMaster

A massive central data source and a handy way to graphically compare nations. NationMaster is a vast compilation of data from such sources as the CIA World Factbook, United Nations, World Health Organization, World Bank, World Resources Institute, UNESCO, and more.
CICLO DE VIDA

A PARCEAMIENTO  HUEVECILLOS  LARVA U ORUGA  PUPAS O CRISALIDAS

“Monarch Butterfly Life Cycle” by hspauldi via Flickr, under CC BY-SA 2.0
FIVE

Formatting and Organizing Data

Data from the book, and are the result of empirical research. Research data exist in many different forms: textual data, numerical data, databases, graph data, sound, audio-visual recordings, and data presented by machines or instruments. All digital data exist in specific file formats, the form in which information is stored as a software program can read and interpret these data. A particular file format is usually linked to a specific software program. If the file is to be read by a different program it may need to be converted. Using standard and adaptable or open "standards" data formats ensure long-term usability and quality. Quality control of data is an integral part of all research and takes place at various stages during data collection, data entry or digitization, and data checking. It is vital to develop suitable procedures before data gathering starts in order to ensure the validity of any conventions, instructions, guidelines or templates that will help preserve quality and consistency across a data collection.

High quality data are well organized, structured, named and versions and the authenticity of master data identified. It is important to ensure that different versions or versions of files, files held in different formats or locations, and information that is cross-referenced between files are all subject to version control.

### Table 5.2
<table>
<thead>
<tr>
<th>Format</th>
<th>Numeration and sequencing</th>
<th>Readability and accessibility</th>
<th>Information as a result of organization</th>
<th>Preservation at storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain Text (TXT)</td>
<td>No specific characters, puts, procedures, can be read, written, modified, used in almost any kind of software</td>
<td>Readable and accessible</td>
<td>Readable as a result of organization</td>
<td>Can be read and modified in almost any software applications</td>
</tr>
<tr>
<td>Microsoft Word (DOC, DOCX)</td>
<td>Text can be formatted, page size, and font style, etc. are standardized</td>
<td>Reader-friendly, can be read by almost any software</td>
<td>Readable as a result of organization</td>
<td>Can be read and modified in almost any software applications</td>
</tr>
<tr>
<td>PDF (Portable Document Format)</td>
<td>Text can be formatted, page size, and font style, etc. are standardized</td>
<td>Reader-friendly, can be read by almost any software</td>
<td>Readable as a result of organization</td>
<td>Can be read and modified in almost any software applications</td>
</tr>
</tbody>
</table>

### Category

<table>
<thead>
<tr>
<th>Basic/Intro Questions</th>
<th>Advanced/Later Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td>Which of your works/What kinds of works would you like to include in a repository?</td>
</tr>
<tr>
<td><strong>Storage/Preservation</strong></td>
<td>Where is your work/content currently stored? What happens if you lose it?</td>
</tr>
<tr>
<td><strong>Sharing/Accessibility</strong></td>
<td>Would you like to see your work/content shared with the campuses, with the world, or not shared at all?</td>
</tr>
<tr>
<td><strong>Organization / Discoverability (Metadata)</strong></td>
<td>Do you have descriptions of your work/content?</td>
</tr>
<tr>
<td><strong>Intellectual Property</strong></td>
<td>Do you have permission to contribute your work/content? If not, does your use of someone else's work/content constitute fair use?</td>
</tr>
<tr>
<td><strong>Linked data</strong></td>
<td>Is it important that your work/content link to other resources, both within and outside the repository?</td>
</tr>
</tbody>
</table>

- **Metadata**
- **Readability**
- **Preservation at Storage**

- **Text Formatting**
- **Numeration and Sequencing**
- **Readability and Accessibility**
- **Information as a Result of Organization**
DATA DOCUMENTATION AND METADATA

Getting the basics down
- Who? Who contributed to the project (authors, researchers, etc.)?

Getting a little more in-depth
Imagine that you have to leave the project as is for a couple of months and then come back to it. What are the most important aspects of the project you’d need help remembering? Some examples:
- file handling (how are they named, how are they divided)
- processing steps (how to get from point A to B)
- field abbreviation/name glossary (how does ABC310G stand for again?)

Now imagine you had to leave the project and come back after six months or a year. What else would you add to the list?

Need Help? Download an Example Readme.txt (plain text) template that can be adapted for your data.

Standardizing your documentation
With the "raw material" documenting your project down, the next step is to standardize the formatting. The standard to use depends on the discipline and/or format of your data. A few standards are listed below. Again, this isn’t intended to be exhaustive, but rather descriptive.

<table>
<thead>
<tr>
<th>Type of Data</th>
<th>Disciplines</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social and Behavioral Sciences</td>
<td>Data Documentation Initiative (DDI)</td>
<td></td>
</tr>
<tr>
<td>Ecology</td>
<td>Ecological Metadata Language (EML)</td>
<td></td>
</tr>
<tr>
<td>Spatial</td>
<td>Content Standard for Digital Geospatial Metadata (CSDGM) FDSC/19115</td>
<td></td>
</tr>
<tr>
<td>Biodiversity</td>
<td>Life Sciences</td>
<td>Desert Core</td>
</tr>
</tbody>
</table>

Guide to writing "readme" style metadata
A readme file provides information about a data file and is intended to help ensure that the data can be correctly sharing or publishing data. Standards-based metadata is generally preferable, but where no appropriate standard is available.

This document is also available as a PDF that includes example readme files.

- Best practices
- Recommended content:
  - Introductory information
  - Methodological information
  - Data-specific information
  - Sharing/access information
- References
- Related information