Applied Data Science: Managing Research Data for Reuse  
July 23-27, 2012  
Syllabus

Instructors/Organizers:

Jake Carlson, Purdue University  
Kathleen Fear, University of Michigan  
Jared Lyle, ICPSR  
Mary Vardigan, ICPSR

This workshop is for individuals interested or actively engaged in the management and curation of research data, particularly data scientists, data managers and analysts, librarians, archivists, and data stewards and curators. As data multiply in sheer quantity and become increasingly important in the research process, the demand for knowledgeable curation rises. What are the best practices for managing research data? How does one apply them to daily practice? What tools can assist in curation efforts?

This five-day workshop will explore and apply the concepts and benefits of life cycle principles for data management, including the management of data from its inception through archiving and beyond, and how to encourage and trace data re-use. An ICPSR dataset will serve as a case study and participants will track the dataset as it makes its way through the ICPSR data pipeline. Participants will also learn about confidential data management, repository requirements and assessment, effective documentation practices, and how to create and evaluate required data management plans. In addition to presentations and panel discussions, participants will engage in small group exercises, technical demonstrations, and in-depth case studies. The workshop will draw on a diverse set of instructors representing the deep data curation expertise at ICPSR, the world's largest social science data archive; the University of Michigan, including its School of Information; and other peer institutions.

List of Relevant Readings

Monday, July 23rd  
Day 1

1:00 pm – 1:30 pm  
Introductions and Course Logistics  
Facilitators: Jake Carlson, Kathleen Fear, Jared Lyle, and Mary Vardigan

1:30 pm – 2:15 pm  
Research Life Cycle Stages  
Instructor: Jake Carlson  
Presentation
2:15 pm – 3:00 pm  Mapping Role and Services to Life Cycle Stages
Instructor: Jake Carlson
Exercise

**3:00 pm – 3:15 pm**  BREAK

3:15 pm – 3:45 pm  Identifying Information to Support Data Curation from Documentation
Instructor: Jake Carlson
Case Study Exercise

3:45 pm – 4:30 pm  Introduction of Case Study Dataset
Instructor: Jared Lyle
Case Study

4:30 pm – 5:45 pm  Reception in the Ehrlicher Room

**Tuesday, July 24th**  Day 2 - Data Management and Sharing

9:00 am – 9:45 am  Case Study Continued
Instructor: Jared Lyle
Case Study and Exercise

9:45 am – 10:30 am  Planning for Data Management and Sharing
Instructors: Jake Carlson, Alex Kanous, and Amy Pienta
Presentations

10:30 am – 11:30 am  Developing an Informed Data Management Plan
Instructors: Jake Carlson, Alex Kanous, and Amy Pienta
Exercise

**11:30 am – 11:45 am**  BREAK

11:45 am – 12:30 pm  Data Management for the Long Tail
Instructor: Margaret Hedstrom
Presentation

**12:30 pm – 1:30 pm**  LUNCH

1:30 pm – 2:15 pm  Informal Conversation with Stakeholders about Data Management and Sharing
Instructors: Chris Dick, Jacqui Hinchey, Ron Nakao
Discussion
2:15 pm – 3:00 pm  Managing and Sharing Disparate Data Types (Part 1): Video Data and Qualitative Data  Instructors: Bryan Beecher and Kaye Marz  Panel Discussion and Presentations

3:00 pm – 3:15 pm  BREAK

3:15 pm – 4:45 pm  Managing and Sharing Disparate Data Types (Part 2): Images, Geospatial Data, Relational Databases  Instructors: Paul Conway, Nicole Scholtz, Justin Noble  Panel Discussion and Presentations

Wednesday, July 25th  Day 3 - Metadata Matters

9:00 am – 10:30 am  Case Study Continued  Instructor: Jared Lyle  Case Study and Exercise

10:30 am – 10:45 am  BREAK

10:45 am – 11:30 am  Life Cycle Metadata Management and Tools  Instructor: Sanda Ionescu  Presentation

11:30 am – 12:30 pm  Creating Study- and Variable-Level Metadata for Sample Dataset  Instructors: Sanda Ionescu, Matthew Richardson, and Mary Vardigan  Exercise

12:30 pm – 1:30 pm  LUNCH

1:30 pm – 3:00 pm  Provenance Issues for Data  Instructors: George Alter, Devan Donaldson, and H.V. Jagadish  Panel Discussion and Presentations

3:00 pm – 3:15 pm  BREAK

3:15 pm – 4:00 pm  Confidential Data Management - Opportunities and Challenges; Disclosure Risk and IRBs  Instructor: George Alter  Presentation
4:00 pm – 4:45 pm  Procedures for Limiting Disclosure Risk: What to Look For
Instructor: JoAnne McFarland
Presentation

Thursday, July 26th  Day 4 - Repository Management

9:00 am – 10:30 am  Case Study Continued
Instructor: Jared Lyle
Case Study and Exercise

10:30 am – 10:45 am  BREAK

10:45 am – 11:30 am  Restricted-use Data Contracts and Handling at ICPSR – Tools
Instructors: Linda Detterman and Bryan Beecher
Demonstration

11:30 am – 12:30 pm  Digital Preservation Overview
Instructor: Beth Yakel
Presentation

12:30 pm – 1:30 pm  LUNCH

1:30 pm – 2:15 pm  Repository Management - Organizational Issues
Instructor: Beth Yakel
Presentation

2:15 pm – 3:00 pm  Repository Management - Technologies
Instructor: Bryan Beecher
Presentation

3:00 pm – 3:15 pm  BREAK

3:15 pm – 3:45 pm  Repository Requirements and Assessment -- ICPSR; Data Seal of Approval and TRAC
Instructors: Jared Lyle and Mary Vardigan
Presentation

3:45 pm – 4:15pm  Repository Requirements and Assessment – Hathitrust; TRAC
Instructor: Jeremy York

4:15 pm – 4:45 pm  Comparing Repository Assessment Methods
Instructor: Kathleen Fear
Exercise
<table>
<thead>
<tr>
<th>Time</th>
<th>Event Description</th>
<th>Instructors</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00 am – 9:45 am</td>
<td>Case Study Conclusion</td>
<td>Jared Lyle</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Case Study and Exercise</td>
</tr>
<tr>
<td>9:45 am – 10:30 am</td>
<td>Tracking Reuse</td>
<td>Elizabeth Moss</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Presentation</td>
</tr>
<tr>
<td><strong>10:30 am – 10:45 am</strong></td>
<td><strong>BREAK</strong></td>
<td></td>
</tr>
<tr>
<td>10:45 am – 12:00 pm</td>
<td>Data Curation and Reuse Across Domains</td>
<td>Ixchel Faniel, Jessica Faul, and Kathleen Fear</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Panel Discussion and Presentations</td>
</tr>
<tr>
<td>12:00 pm – 12:30 pm</td>
<td>Next Steps and Follow Up</td>
<td>Jake Carlson, Kathleen Fear, Jared Lyle, and Mary Vardigan</td>
</tr>
</tbody>
</table>