Data management training @ IUPUI

Instructional Design

Assessment

Activities

Success & Challenges
Audience

• Pilot Lab (January 2014)
  – Indiana CTSI Clinical Data Management Team
  – Research Staff from local hospitals

• Workshop Series (Spring 2014)
  – Mostly doctoral students, some master’s level
  – Dentistry, Philanthropic Studies, Sociology, Anthropology, Epidemiology, Nursing, Public & Environmental Affairs
Build awareness of digital RDM issues

Introduce methods to facilitate data integrity & address common DM issues

Build proficiency in applying these DM methods

Introduce institutional resources supporting effective DM

Build strategic skills that enable researchers to solve new DM problems
Instructional Design

• Outcome based planning
• Active learning
• Learner-centered
• Constructivist & behaviorist approaches
Why did they register?

• I’m not sure what I need to know (12 of 23 registered)
  – Learn everything I am supposed to know; effective data management techniques; direction & guidance; concrete way to manage and define data found in the field

• I need a plan! (5 of 23)
  – Avoid future stress and anxieties; setting up for dissertation and research beyond that; skills to build a robust data management plan for funding and everything else...

• Specific help needed (4 of 23)
  – Efficient management of raw data; better prepare and organize my information to write my paper

• Big picture (3 of 23)
  – Thesis; GA work; career skills
Content

- Intro to RDM
- DM Planning
- Organize Data & Files
- QA / QC
- Collection
- Entry & Coding
- Screen & Clean
- Automate
- Protection & Security
- Rights & Access
- Attribution & Citation
- Ethical & Legal Obligations
## Instructional Design

<table>
<thead>
<tr>
<th>Module</th>
<th>Topics</th>
<th>Learning Outcomes</th>
<th>Activities</th>
<th>Products Collected in Box</th>
<th>Case</th>
<th>Case Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2: Documentation &amp; Metadata</td>
<td>Organizing data &amp; files</td>
<td>Explain the role of metadata and standards</td>
<td>Identify three functions of metadata for data sharing and reuse</td>
<td>3-minute paper (Word doc)</td>
<td>Y</td>
<td>View examples from subject repositories (DataDryad, NCBI, etc.)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Develop a consistent and coherent file organization and naming convention scheme for all project files</td>
<td>Develop a scheme for folder structure and generating unique descriptive file names</td>
<td>DMP (Word doc)</td>
<td>Y</td>
<td>Examples</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Select appropriate non-proprietary hardware and software formats for storing data</td>
<td>Choose file formats appropriate for your data spreadsheet, database, hierarchical, etc.) and describe your rationale</td>
<td>DMP (Word doc)</td>
<td>N</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Create protected copies of files at crucial points in your study</td>
<td>Identify key points at which you need to create a protected copies of your data</td>
<td>N/A</td>
<td>Y</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Use versioning software or documentation for tracking changes to files over time</td>
<td>Describe the benefits of logging and versioning in provenance/audit trails</td>
<td>DMP (Word doc)</td>
<td>N</td>
<td>N/A</td>
</tr>
<tr>
<td>3: Data Quality</td>
<td>Quality assurance &amp; control</td>
<td>Develop procedures for quality assurance and quality control activities</td>
<td>Define data quality and QA/QC; Identify possible errors and protocol violations; Identify monitoring processes &amp; procedures</td>
<td>Data Quality standards for case (Word doc); Data Collection tool (Word doc)</td>
<td>Y</td>
<td>Sample chart</td>
</tr>
<tr>
<td></td>
<td>Data collection</td>
<td>Describe key considerations for selecting data collection tools</td>
<td>Describe considerations for data collection related to discussions in previous modules</td>
<td>Data Collection Tool (Word doc)</td>
<td>Y</td>
<td>Coding scheme exercise</td>
</tr>
<tr>
<td></td>
<td>Data coding</td>
<td>Use best practices for coding</td>
<td>Develop coding scheme for data collection tool</td>
<td>Coding Scheme (Word doc)</td>
<td>Y</td>
<td>Prepared sample files (Excel, SPSS, REDCap)</td>
</tr>
<tr>
<td></td>
<td>Data entry</td>
<td>Use best practices for data entry</td>
<td>Discuss data entry best practices</td>
<td>N/A</td>
<td>Y</td>
<td>Provide examples of</td>
</tr>
</tbody>
</table>
## Assessment Products

<table>
<thead>
<tr>
<th>Learning Outcomes</th>
<th>Activities</th>
<th>Products Collected in Box</th>
<th>Case</th>
<th>Worked &amp; Practiced Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explain the role of metadata and standards</td>
<td>Identify three functions of metadata for data sharing and reuse</td>
<td>3-minute paper (Word doc)</td>
<td>Y</td>
<td>View examples from subject repositories (DataDryad, NCBI, etc.)</td>
</tr>
<tr>
<td>Develop a consistent and coherent file organization and naming convention scheme for all project files</td>
<td>Develop a scheme for folder structure and generating unique descriptive file names</td>
<td>DMP (Word doc)</td>
<td></td>
<td>Examples</td>
</tr>
<tr>
<td>Select appropriate non-proprietary hardware and software formats for storing data</td>
<td>Choose file formats appropriate for your data (spreadsheet, database, hierarchical, etc.) and describe your rationale</td>
<td>N/A</td>
<td>Y</td>
<td>N/A</td>
</tr>
<tr>
<td>Create protected copies of files at crucial points in your study</td>
<td>Identify key points at which you need to create a protected copies of your data</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>Use versioning software or documentation for tracking changes to files over time</td>
<td>Describe the benefits of logging and versioning in provenance/audit trails</td>
<td>DMP (Word doc)</td>
<td>N</td>
<td>N/A</td>
</tr>
<tr>
<td>Develop procedures for quality assurance and quality control activities</td>
<td>Define data quality and QA/QC; Identify possible errors and protocol violations; Identify monitoring processes &amp; procedures</td>
<td>Data Quality standards for case (Word doc); Data Collection tool (Word doc)</td>
<td>Y</td>
<td>Sample chart</td>
</tr>
<tr>
<td>Describe key considerations for selecting data collection tools</td>
<td>Describe considerations for data collection related to discussions in previous modules</td>
<td>Data Collection Tool (Word doc)</td>
<td></td>
<td>Coding scheme exercise</td>
</tr>
<tr>
<td>Use best practices for coding</td>
<td>Develop coding scheme for data collection tool</td>
<td>Coding Scheme (Word doc)</td>
<td>Y</td>
<td>Prepared sample files (Excel, SPSS, REDCap)</td>
</tr>
<tr>
<td>Use best practices for data entry</td>
<td>Discuss data entry best practices</td>
<td>N/A</td>
<td>Y</td>
<td>Provide examples of</td>
</tr>
</tbody>
</table>
Data mapping

Outcomes mapping

Data outcomes mapping
Mapping Data Outcomes

• Part of the planning process
• Identifies desired outputs
  – Statistical descriptions
  – Images, graphs, tables, etc.
• Step-by-step planning backwards from those outputs to identify
  – Data characteristics
  – Data quality issues
  – Quality control & quality assurance processes
<table>
<thead>
<tr>
<th>Research Questions</th>
<th>Analyses</th>
<th>Variables, Data Fields</th>
<th>Source/Collection Method</th>
<th>Data Requirements &amp; Assumptions</th>
<th>Analytical Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does body image differ between racial/ethnic groups?</td>
<td>chi square</td>
<td>Q6_COMP:RACE:COMBINED-COMPUTED, Q8:YOURTHOUGHTS ONYOURBODY</td>
<td>World Health Organization, Health Behaviors of School-aged Children</td>
<td>1. Each person, item, entity is included in only one cell of the contingency table (i.e., no repeated measures); 2. Expected frequencies should be greater than 5</td>
<td>Look at row and column percentages to interpret any effects logistic</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Potential Collection Issues</th>
<th>Potential Processing Issues</th>
<th>Quality Standards</th>
<th>QA/QC Procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify response rate required for sufficient power; Does sampling method address this? Time required to complete; motivation; validity of self-reported data; Prevent missing responses</td>
<td>How to handle missing and inconsistent data? Need to identify multiple types of missing data? How to handle logical errors?</td>
<td>Valid range; Expected distribution; Expected missing cases</td>
<td>Screen with frequency distributions and check standard deviation for sample and by racial/ethnic groups; Check missing cases for variables against overall missing data rate; Run normality tests (K-S or S-W)</td>
</tr>
</tbody>
</table>
What worked

• Place learning outcomes in the big picture; help students understand why they should care about data management & preservation
• Break up lecture with frequent breaks for student activities
• Provide opportunities for students to learn from others’ diverse experiences
• Use experiential methods (e.g., scenarios, case studies)
...and what didn’t

- Start where the students are → Pre-assessment
- Demonstrate how each piece fits into the whole → Show the logical sequence of learning outcomes (i.e., outcomes map)
- Teach them to learn your material → Give students a mental model for data management
- Support novice researchers → Provide better scaffolding
- Build in more assessment opportunities & use the data to adjust course → Incorporate peer review
Ongoing challenges

• Reaching the right people at the right time (point-of-need support)

• Attrition & scheduling

• Relating recommended practices to immediate needs & priorities; keeping them engaged

• Developing materials to support diverse practices and norms of multiple disciplines

• Continued engagement after instruction sessions
Evolution of DIL instruction @ IUPUI


More Resources

See our examples & other data management curricula and programs at

http://hdl.handle.net/1805/6043