Ownership and Data Sharing

- Ownership
- Proprietary data
- HIPAA
Ownership and Data Sharing

- Ownership—PI, Institution, Funding source
- Proprietary data—IP Ownership / Redaction / Restrictions?
- HIPAA—Privacy / De-identification procedures?
Data Types

Observational  Simulation  Derived  Compiled

Quantity?
Data Format

- Text
- Numeric
- Audiovisual
- Modeling
- Unique/specific/proprietary
- Meta data/audit trails

Image courtesy of Bing
Organization and Storage

- Fixed (No changes)
- Updated with no changes
- Updated with changes—Change Control

Risk?
Organization and Storage

- Location
- Accessibility / Security (including encryption)
- Back-up
- Retention / Archival
- Data Migration / Operating System Sustainability
Organization and Storage

Electronic Data
- Audit Trails
- Transposition Errors
- Software Compatibility
- Program Updates
  - Automatic
- Impact to significant digits

Risk?
Organization and Storage

Electronic Data

- Standard File Naming System
  - Brief, descriptive, consistent, dated
  - Plans for edits and changes

![Diagram of file naming system]

**Documents library**

- HQ.Forms.Submitted
- Name
  - SQA.apr16qc.CourseDescForm.Revised.1Dec.2015
  - SQA.apr16qc.CourseDescForm

![Diagram of file naming system]
Data Manipulation

Source data → Data Collection → Data Analysis → Data Sharing

Observations

Re-Purpose

Data Life Cycle
Data Manipulation
Mechanisms for Data Sharing

Source data → Observations → Data Collection → Data Analysis → Data Sharing → Re-Purpose → Data Life Cycle
Mechanisms for Data Sharing

- Email
- Online repositories
- Supplemental to publication
- CD / USB
- Sharing agreements
- Mixed
- Conditions / Exclusions
Mechanisms for Data Sharing

Data Sharing / Citation

- Creator(s)
- Title (version)
- Publication year (or release date)
- Publisher or Data Repository
- Identifier
### Summary of the eight standards and three levels of the TOP guidelines

Levels 1 to 3 are increasingly stringent for each standard. Level 0 offers a comparison that does not meet the standard.

<table>
<thead>
<tr>
<th>LEVEL 0</th>
<th>LEVEL 1</th>
<th>LEVEL 2</th>
<th>LEVEL 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Citation standards</strong></td>
<td>Journal encourages citation of data, code, and materials—or says nothing.</td>
<td>Journal describes citation of data in guidelines to authors with clear rules and examples.</td>
<td>Article provides appropriate citation for data and materials used, consistent with journal’s author guidelines.</td>
</tr>
<tr>
<td><strong>Data transparency</strong></td>
<td>Journal encourages data sharing—or says nothing.</td>
<td>Article states whether data are available and, if so, where to access them.</td>
<td>Data must be posted to a trusted repository. Exceptions must be identified at article submission.</td>
</tr>
<tr>
<td><strong>Analytic methods (code) transparency</strong></td>
<td>Journal encourages code sharing—or says nothing.</td>
<td>Article states whether code is available and, if so, where to access them.</td>
<td>Code must be posted to a trusted repository. Exceptions must be identified at article submission.</td>
</tr>
<tr>
<td><strong>Research materials transparency</strong></td>
<td>Journal encourages materials sharing—or says nothing.</td>
<td>Article states whether materials are available and, if so, where to access them.</td>
<td>Materials must be posted to a trusted repository. Exceptions must be identified at article submission.</td>
</tr>
<tr>
<td><strong>Design and analysis transparency</strong></td>
<td>Journal encourages design and analysis transparency or says nothing.</td>
<td>Journal articulates design transparency standards.</td>
<td>Journal requires adherence to design transparency standards for review and publication.</td>
</tr>
<tr>
<td><strong>Preregistration of studies</strong></td>
<td>Journal says nothing.</td>
<td>Journal encourages preregistration of studies and provides link in article to preregistration if it exists.</td>
<td>Journal encourages preregistration of studies and provides link in article and certification of meeting preregistration badge requirements.</td>
</tr>
<tr>
<td><strong>Preregistration of analysis plans</strong></td>
<td>Journal says nothing.</td>
<td>Journal encourages preregistration of analysis plans and provides link in article to registered analysis if it exists.</td>
<td>Journal encourages preanalysis plans and provides link in article and certification of meeting registered analysis plan badge requirements.</td>
</tr>
<tr>
<td><strong>Replication</strong></td>
<td>Journal discourages submission of replication studies—or says nothing.</td>
<td>Journal encourages submission of replication studies.</td>
<td>Journal encourages submission of replication studies and conducts blind review of results.</td>
</tr>
</tbody>
</table>
Data Management and Sharing Plan

- Data Description / Types
- Data Standards for Format and Content
- Provisions for Data Reuse and Redistribution
- Archiving / Long-term Preservation and access
- Other…
Data Quality: GDP Principles

- Attributable
- Legible
- Contemporaneous
- Original
- Accurate

ALCOA+

Complete, Consistent, Enduring, Readily Available
Part 4: Case Studies
Case Study 1—Data Sharing

Your research study will include data from approximately 500 subjects being screened for three bacterial sexually transmitted diseases (STDs) at an inner city STD clinic. The final dataset will include self-reported demographic and behavioral data from interviews with the subjects and laboratory data from urine specimens provided. Because the STDs being studied are reportable diseases, you will be collecting identifying information.

Even though the final dataset will be stripped of identifiers prior to release for sharing, there remains the possibility of deductive disclosure of subjects with unusual characteristics. Identify options (i.e., conditions) for sharing the data.
Case Study 2—Data Management

You will be conducting a single-dose drug study in a mouse model to determine if Drug X will have an impact on cholesterol values. You plan to collect weekly blood samples over a 28-day period to statistically analyze LDL and HDL levels.

Describe processes that might be in place for the storage, protection and retention of the laboratory data during and after study completion.
You are conducting a study that will generate the following electronic files. Develop a plan to organize the electronic folders and develop a standard naming system for the files that will allow for tracking of any changes. (Note: Consider if it is more useful to order the files by date, by author, or by subject, for example).

Project Number: JJ2A2016

Study conducted: January - September, 2016

- PCR data, analyzed in batches performed weekly by 3 different trained research associates (.scn files)
- Gel images, collected weekly by 2 different trained research associates (.tif files)
- Compiled blood chemistry data updated 2-3 times per week (.xlsx files)
- Compiled hematology data updated 2-3 times per week (.xlsx files)
- Statistical analysis in GraphPad Prism (.pzfx files)
Summary

- Who owns the data (resource)?
- What are your data?
- When will the data be available?
- Where will the data be retained?
- How will the data be organized, secured and shared?
References – Data Management

https://dmptool.org/dm_guidance (General guidance)
http://www.icpsr.umich.edu/files/datamanagement/DataManagementPlans-All.pdf (Guidelines)
http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/plan.html (Sample plan)
https://obamawhitehouse.archives.gov/blog/2016/02/22/increasing-access-results-federally-funded-science (White House Memorandum Archives)
http://libguides.northwestern.edu/datamanagement/federalfundingagency#s-lg-box-wrapper-7012284 (Other Federal funding agencies)
http://libraries.ucsd.edu/services/data-curation/data-management/dmp-samples.html (Example DMPs)