TEACHING AND LEARNING
ETHICAL DATA MANAGEMENT

Xiaofeng “Denver” Tang (xut2@psu.edu)
Statement about data management _1

From a graduate student:

I am listed as a co-author of a paper for which I helped with the literature review. As I am not involved in the collection and analysis of research data, I bear no responsibility for verifying the results published in the paper, which has multiple co-authors.
From a graduate student:

I am listed as a co-author of a paper for which I helped with the literature review. As I am not involved in the collection and analysis of research data, I bear no responsibility for verifying the results published in the paper, which has multiple co-authors.

Not understanding the shared responsibility for reporting verified, reproducible results in research publications.
From a junior faculty member:

My research is not supported by any extramural funding; therefore, I have no obligation to maintain procedural data as long as my results are valid.
From a junior faculty member:

My research is not supported by any extramural funding; therefore, I have no obligation to maintain procedural data as long as my results are valid.

Deviating from commonly expected research practices (research norms).

Neglecting the communal (reciprocal) aspect of research.
Statement about data management

From a PI:
I have thoroughly informed my graduate students, lab technicians, and statisticians about the proper ways of handling data; therefore, I should not be held accountable if issues about the credibility of research data arise in our co-authored publications.
From a PI:

I have thoroughly informed my graduate students, lab technicians, and statisticians about the proper ways of handling data; therefore, I should not be held accountable if issues about the credibility of research data arise in our co-authored publications.

The responsibility of leadership?
Effective communication/monitoring?
Collective review?
From a PI:
The volunteers who participated in my research have been paid of an amount they agreed. They should not have a say on how I proceed with my research based on the data I have collected from them.
From a PI:

The volunteers who participated in my research have been paid of an amount they agreed. They should not have a say on how I proceed with my research based on the data I have collected from them.

The **rights** of research participants?
Agreement on data ownership?
Respect for participants?
From a new graduate student:

I repeated a part of the experiments my group have conducted and published before. My results included some “outliers” which I knew resulted from my lack of experience. Thus I deleted these obvious outliers from the final results I submitted to my advisor, I did mention the existence of outliers during the group research meeting.
Statement about data management_5

From a new graduate student:
I repeated a part of the experiments my group have conducted and published before. My results included some “outliers” which I knew resulted from my lack of experience. Thus I deleted these obvious outliers from the final results I submitted to my advisor, I did mention the existence of outliers during the group research meeting.

Effective/ineffective communication?
Lack of proper training?

Created by Xiaofeng "Denver" Tang, Penn State, 2016
From a computer engineer:

I developed a browser plugin that makes “suggestions” to the users based on their search history. I made sure that the algorithm would blind any human actors, including our developer team, to the users’ search history.
From a computer engineer:

I developed a browser plugin that makes “suggestions” to the users based on their search history. I made sure that the algorithm would blind any human actors, including our developer team, to the users’ search history.

Proper (adequate) protection?
Informed consent (the right to collect data)?
Overall Lessons

- The ecology and lifecycle of data
- Research as a systematic and institutional activity
- Ethical concepts
“This new approach, which I call information ecology, emphasizes an organization’s entire information environment. It addresses all of a firm’s values and beliefs about information (culture); how people actually use information and what they do with it (behavior and work processes); the pitfalls that can interfere with information sharing (politics); and what information systems are already in place (yes, finally, technology).”

The Ecology of Data: Elements

1. Evolving System
2. Constraining Conditions
3. Selection Pressures
4. Info-diversity/Data diversity
5. Types/Groups/Forms
6. Data Niches
7. Resources
8. Interactions
9. Co-evolution/Co-adaptation

Adapted from Treloar, A. 2012. “Conceptualising Collaboration and Competition in the Changing Ecology of Research Data.”

Created by Xiaofeng "Denver" Tang, Penn State, 2016
# Ontology/States of Data

<table>
<thead>
<tr>
<th>Data</th>
<th>Structured Collections</th>
<th>Actors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmanaged</td>
<td>Managed</td>
<td>Individuals</td>
</tr>
<tr>
<td>Disconnected</td>
<td>Connected</td>
<td>Groups</td>
</tr>
<tr>
<td>Invisible</td>
<td>Findable</td>
<td>Research Community</td>
</tr>
<tr>
<td>Single use</td>
<td>Reusable</td>
<td>Society</td>
</tr>
</tbody>
</table>

Adapted from Treloar, A. 2012. “Conceptualising Collaboration and Competition in the Changing Ecology of Research Data.”

Created by Xiaofeng "Denver" Tang, Penn State, 2016
THE ECOLOGY OF DATA

Mega data

Created by Xiaofeng "Denver" Tang, Penn State, 2016
Planning Generating Processing
Using, sharing, and preserving
Life Cycle of Data Management
Mega data

Created by Xiaofeng "Denver" Tang, Penn State, 2016
Planning
Generating
Processing
Using, sharing, and preserving

Life Cycle of Data Management

Created by Xiaofeng "Denver" Tang, Penn State, 2016
Planning

Generating

Processing

Using, sharing, and preserving

Life Cycle of Data Management

• Understanding Data and Research
  ➢ Definitions of data
  ➢ From “data” to “facts”
  ➢ The system of data
• The life cycle of data management
  ➢ The research process and actors
  ➢ Ethical concepts
  ➢ Overall strategies

Created by Xiaofeng "Denver" Tang, Penn State, 2016
Planning

Generating

Processing

Using, sharing, and preserving

Life Cycle of Data Management

• Sources of data
  • Generating and collecting data
    ➢ Protocols and training
    ➢ Lab notebook
    ➢ Poor data practice
    ➢ Collective review
  • Human and animal subjects
    ➢ Laws
    ➢ Agencies
    ➢ Consent and rights

Created by Xiaofeng "Denver" Tang, Penn State, 2016
Planning

Generating

Processing

Using, sharing, and preserving

Life Cycle of Data Management

- Statistical tools and outliers
- Representing and interpreting
  - Interest, uncertainty, and bias
  - Conflict of interest
- Falsification
- Fishing or trimming
- Image manipulation

Created by Xiaofeng "Denver" Tang, Penn State, 2016
• Data usage
  ➢ Communal purposes of research
  ➢ Principles
• Data ownership and access
  ➢ IP
  ➢ Transparency
  ➢ Data policies
  ➢ Open access
• Data storage and protection
  ➢ Policies of journals and funding agencies
• Data storage and protection
  ➢ Methods
  ➢ Period
  ➢ Security

Created by Xiaofeng "Denver" Tang, Penn State, 2016
Mega data

- Big data
  - Meta data and data mining
- Cloud computing

Life Cycle of Data Management
- Planning
- Generating
- Processing
- Using, sharing, and preserving
Mega data

Planning

Generating

Processing

Using, sharing, and preserving

Life Cycle of Data Management

- Understanding Data and Research
- The life cycle of data management: an overview

Sources of data
- Generating and collecting data
- Human and animal subjects

- Big data
- Cloud computing

- Data usage
- Data ownership and access
- Data storage and protection

- Statistic tools
- Representing and interpreting
- Falsification

Created by Xiaofeng "Denver" Tang, Penn State, 2016
Resources for Teaching and Learning (1)

- **Handbooks**
  - On Being A Scientist
  - Guidelines for Responsible Data Management in Scientific Research

- **Statistics**
  - PSU STAT 500

- **Lab Notebooks**
  - Exemplar lab notebook

- **Data Policies**
  - NSF ENG Data Management Plan Requirements
  - NIH Data Sharing Policy
  - Nature. “Availability of data, material and methods”
  - Science. “Editorial policies”
  - Please suggest important journals and funding agencies in your field.

Created by Xiaofeng "Denver" Tang, Penn State, 2016
Resources for Teaching and Learning (2)

- Ethical theories and concepts
- Bibliography
- Case studies and analysis
  - The Online Ethics Center (http://www.onlineethics.org/)
    - “All in the Interpretation”
    - “A pHish Tale”
  - The STAP cells case
  - Data mining, national security, and privacy case
  - Potential materials for creating case studies
    - Volkswagen “defeat device” for cheating on emission tests
    - Mitsubishi cheating on fuel economy tests
    - Who owns your gene?

Created by Xiaofeng "Denver" Tang, Penn State, 2016
What are the ethical issues?

- Stakeholders
- Facts
- Values
- Consequences
- Rights/duties
- Virtues
- Relationships

Ethical Integrity

Problem Definition

Reflect

Gathering Information

Possible and best options

Generating, evaluating, and taking actions

Applying Guiding Principles

Prevent future problems

Created by Xiaofeng "Denver" Tang, Penn State, 2016
Method for Ethical Case Analysis

1. Problem identification: what are the ethical issues?
2. Gathering information
   - Stakeholders
   - Facts
   - Values
3. Applying guiding principles
   - Consequences
   - Rights/duties
   - Virtues
   - Relationships
4. Generating, evaluating, and choosing options: possible and best actions
5. Reflecting: prevent problems in the future

Created by Xiaofeng "Denver" Tang, Penn State, 2016
The STAP Cells

Haruko Obokata led a lab for the Riken Center for Developmental Biology. In 2014, Obokata and her collaborators published two articles in *Nature* claiming that they had discovered a simple way of making STAP cells. However, readers found some images in the articles had been inappropriately manipulated and some texts were copied from previous articles. Worse still, no other lab was able to replicate Obokata’s results following the method she had reported. After a failed attempt to repeat her own experiments in a monitored environment, Obokata resigned from Riken. An investigation conducted by Riken concluded that the “STAP cells” claimed in Obokata’s papers were embryonic stem cells taken from elsewhere.
Data Mining, National Security, and Privacy

http://www.wnyc.org/story/terrorism-algorithm-legal/

Created by Xiaofeng "Denver" Tang, Penn State, 2016
Teaching Ethical Data Management

- Case analysis
- Discussion
  - News
  - Research practices
  - Cultural images
- Review and creating policies
  - Protocols
  - Communication maps
- Teaching and modeling ethical data practice in research settings
  - Reviewing data in research meetings
  - Asking questions about data handling during seminars

Created by Xiaofeng "Denver" Tang, Penn State, 2016