Big Data Ethics Support Systems and Networks
Big Data Ethics Support Systems and Networks

Bonnie Tijerina & Emily F. Keller
Data & Society Research Institute
Big Data

(CC BY-NC 2.0-licensed photo by janneke staaks.)
Facebook Tinkers With Users' Emotions in News Feed Experiment, Stirring Outcry

Facebook fiasco: was Cornell’s study of 'emotional contagion' an ethics breach?

A covert experiment to influence the emotions of more than 600,000 people. A major scientific journal behaving like a rabbit in the headlights. A university in a PR tailspin.

Coming Soon: Ethics Training for Data Scientists

Harvard Researchers Used Secret Cameras to Study Attendance. Was That Unethical?

A high-tech effort to study classroom attendance at Harvard University that used secret photo surveillance is raising questions about research ethics among the institution's faculty members. The controversy heated up on Tuesday night, when a computer-science professor, Harry R. Lewis, questioned the study at a faculty meeting.

During the study, which took place in the spring of 2013, cameras in 10 Harvard classrooms recorded one image per minute, and the photographs were scanned to determine which seats were filled.

To some professors, it was an obvious intrusion into their privacy—and their students'.
Data & Fairness

The Future of Labor in a Data-Centric Society

Enabling Connected Learning

Ethics in “Big Data” Research

Intelligence and Autonomy

Data, Human Rights & Human Security
Data & Fairness

The Future of Labor in a Data-Centric Society

Enabling Connected Learning

Ethics in “Big Data” Research

Intelligence and Autonomy

Data, Human Rights & Human Security

Council for Big Data, Ethics, and Society

Ethics in Research Data Project
Ethical Concerns

• Data Collection
• Data Storage
• Data Sharing, Reuse, Replicability
• Re-identification and Consent
• Unknown and Emerging
Ethics in Research Data Project

Is there a role for libraries as part of a campus support network helping technical researchers navigate emerging ethical issues in (big) data research?
Formal Structures

- IRB
- DMPs
- Requirements
Reframing Ethics

Definitions

- Violations
- Trustworthy
- Morality
- Conscience
- Mistakes
- Transparency
- Responsibility
- Protection
- Integrity

- Anonymity
- Accountability
- Responsibility
- Teaching
- Consent
- Objectivity
- Security
Teaching Ethics

- Apprenticeship
- Required click-through training
- Embedded ethical questioning
- Lecture series
- Case studies
- For-credit courses
On-Campus Workshop
Exercise 1. Alice Exercise

- How will Alice know how to write a DMP for her work? Where will she see what others have written?
- Who will teach her about how to navigate IRB? How will she learn how to create an appropriate protocol?
- How will Alice store and secure her data? Are there services on campus that she can turn to that will support her? What will she be expected to manage?
- How will she know which private services to use in her work? (e.g., Dropbox, AWS, Azure, etc.) Who helps her figure out the security and terms of service issues that might arise?
- Who will help her negotiate a contract with ImageCompany?
- Who will help her learn how to navigate Amazon Mechanical Turk’s protocols and processes?
- Who can help her deal with best (technical and social) practices in scraping data or navigating data collection on Amazon Mechanical Turk?
- Who will help her understand how to assess the limitations and biases in her data sets?
- How can Alice get help with technical issues she’s encountered? (Consider statistics questions, machine learning questions, etc. Imagine what happens when her advisor is too busy, lacks the expertise, or she’s too afraid to ask.)
- Should her processes change depending on the social implications of her project? What if she’s consulting for ImageCompany as part of getting access to the data?
- What are the questions Alice hasn’t even thought of yet?
Exercise 2. Data Clinic Model

• Who would the Data Clinic support? Students? Faculty? What disciplines? What would be needed to support computer science faculty vs. humanities students?

• What range of problems could the Data Clinic help with? Imagine technical, logistical, and security-related support services. How much of this overlaps with the Stats Clinic?

• Who would staff the Data Clinic? What is the role of faculty, students, and support services? What kind of hires would be needed?

• What kind of funding support would be needed for this to work at your university? Who would pay for the services? What are the economic hurdles for this?

• What kind of organizational support would be needed to make this work at your university? Bottom up versus top town? Who would need to buy-in? What are the organizational barriers to making this happen?

• What kind of technical and structural support would be needed? (e.g. secure data storage, centralized company contracts) Who currently provides those services?

• What would this model not help with?

• What will get in the way of a Data Clinic working at your university?
Interlocutor

(photo source)
RDM Services in Libraries

• Triage Services
• Support of copyright and IP data issues
• Consultations on data-related skills
• Resource for questions on storage, data sharing, metadata preparation
• Discipline-specific data liaisons
Expanded Role for Libraries

• Data Sharing

• Proper use of information, copyright, and licensing

• Education & Advocacy for the Open Movement
THANK YOU.

Bonnie Tijerina
bonnie@datasociety.net
bonnietijerina.com
@bonlth

Emily F Keller
emily@datasociety.net