



# The Social Life of Data

## Data Management Challenges and Opportunities



Matthew J. Bietz  
University of Washington

# Social Life of Data\*

---

- ▶ Data sharing and collaboration
  - ▶ HIV/AIDS research
  - ▶ Marine Metagenomics
  - ▶ Various other fields
- ▶ Cyberinfrastructure, collaboratories, eScience
- ▶ Data are ***sociotechnical*** artifacts

\* Inspired by Brown & Duguid:  
The Social Life of Information (2000)

# Sociotechnical Approach

---

- ▶ “Technical” and “social” concerns are intertwined
- ▶ Key social questions:
  - ▶ Who are the stakeholders?
  - ▶ What are their motivations?
  - ▶ What are the relationships among stakeholders?
- ▶ Data play multiple simultaneous roles
  - ▶ Scientific fact-making
  - ▶ Generating reputation
  - ▶ Building communities

## Social Roles of Data Reputation

---

- ▶ “Currency” of science
  - ▶ Credit, Authorship
  - ▶ Funding, labs, awards, etc.
- ▶ Scientists known for their data
  - ▶ Quality
  - ▶ Amount
  - ▶ Innovation



# Community

---



- ▶ Community gateways
  - ▶ Who does and does not get access?
- ▶ Enabling “inbound trajectories”
  - ▶ Training students and junior scientists



# Emerging Data Sharing Challenges

---

- ▶ Changing scale
  - ▶ Small → Large
  - ▶ Local → Remote
  - ▶ Systems → Cyberinfrastructures
- ▶ “Social” becoming more important
  - ▶ Science becoming more collaborative and interdisciplinary
  - ▶ Disconnects between data creation and use



# Interdisciplinarity

---

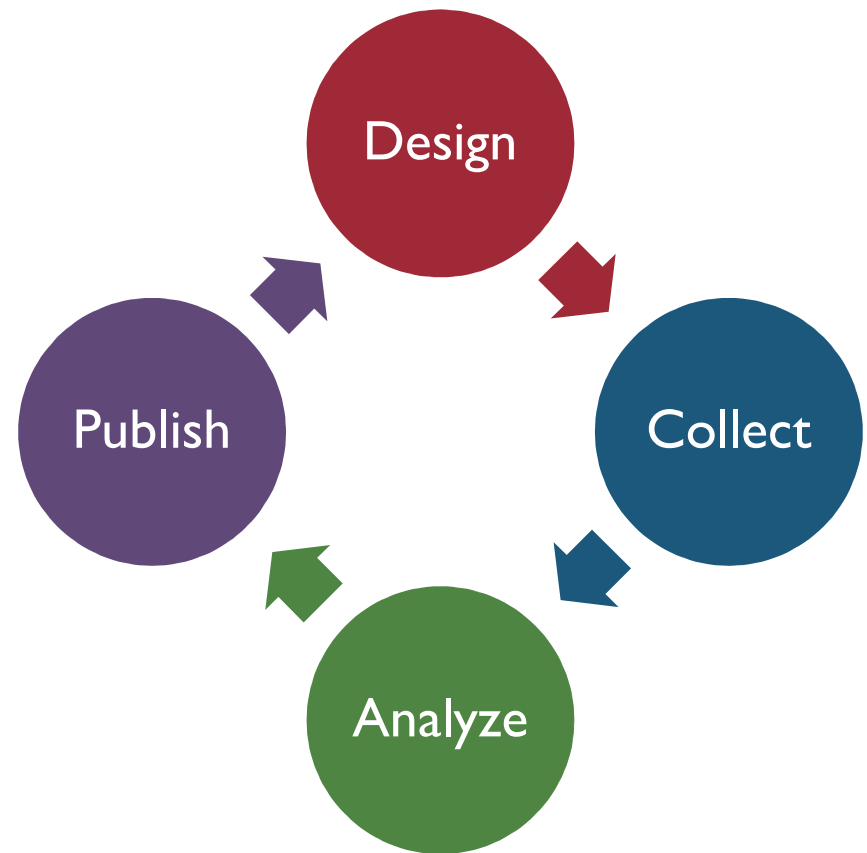
- ▶ Fact-making
  - ▶ Language, level of analysis, formats,.....
- ▶ Reputation
  - ▶ Data valued differently
  - ▶ Changing how data produces reputation
- ▶ Changing community boundaries
  - ▶ Different research questions
  - ▶ New competition



# Division of Scientific Labor

---

- ▶ Decoupling the phases of research lifecycle
- ▶ Roles are assigned different value
  - ▶ Is the Nobel awarded for data collection or analysis?
  - ▶ Is data management “science” or “tech support?”





# Sociotechnical Design

---

- ▶ Genetic Association Information Network
  - ▶ Giving credit for data production
    - ▶ (Meyer, 2009)
- ▶ Standards In Genomic Sciences
  - ▶ Populate databases with high-quality metadata
    - ▶ (Garrity et al., 2009)



# Thank You!

---

- ▶ Contact: Matthew Bietz
  - ▶ [mbietz@uw.edu](mailto:mbietz@uw.edu)
  - ▶ <http://depts.washington.edu/csclab/>
- ▶ Collaborators: Charlotte P. Lee, Gary Olson, & Jeremy Birnholtz
- ▶ Support from NSF Awards IIS-0712994 & OCI-083860
- ▶ Images: [www.lumaxart.com](http://www.lumaxart.com)

