Big Data, Little Data, or No Data? Sustaining Access to Research Data

Christine L. Borgman
PROTECTING CLIMATE DATA IN TIMES OF POLITICAL TURMOIL

UCLA Information Studies

January 20th 2017
Schedule

• 9:00 - 10:30 Data Rescue Panel
• 10:30 - 11:00 orientation for break out sessions, Morgan Currie (IS 111)
• 11:00 - 12:00 break out sessions (IS 111, IS Lab, IS 228, IS 121 – see handout for building map)
• 12:00 - 12:30 lunch break (IS 111)
• 12:30 - 3:00 group work
• 3:00 - 4:00 reconvene for presentations (IS 111)
Data Rescue Panel

- 9-9:05 introductions
- 9:05-9:15 Christine Borgman, UCLA: 
  *Big Data, Little Data, or No Data? Sustaining Access to Research Data*
- 9:15-9:25 Steve Diggs, Scripps Institute: 
  *As the oceans go, so goes the earth*
- 9:30-9:40 Katie Mika, UCLA IOES: 
  *On the space between science and policy*
- 9:45-9:55 Jason Scott, Internet Archive: 
  *Everything is Fine*
- 10:00-10:10 Joan Donovan, UCLA Society & Genetics: 
  *Today Won't Be Like Yesterday: Mining, Archiving, and Decentralizing Data to Preserve Scientific Futures*
- 10:10-10:30 general discussion
Acknowledgements

Funding and support
– UCLA Department of Information Studies
– UCLA Center for Knowledge Infrastructures

Organizers
– Morgan Currie, Irene Pasquetto, Britt Paris, Jen Pierre, Milena Golshan, Seth Erickson, Renee Rottner, Michael Hucka, Yvonne Lee, Jillian Wallis, Joan Donovan, Jesse Ruskin

Additional Support
– Lisa Snyder, Peter Broadwell, Andy Rutkowski, Albert Kochaphum
– #DataRefuge, EDGI, Internet Archive
Big Data, Little Data, or No Data? Sustaining Access to Research Data

Christine L. Borgman
Distinguished Professor & Presidential Chair in Information Studies
University of California, Los Angeles
Christine.Borgman@UCLA.edu

Center for Knowledge Infrastructures
https://knowledgeinfrastructures.gseis.ucla.edu/

Protecting Climate Data in Times of Political Turmoil: Teach-in and Hackathon
UCLA, January 20, 2017
Big Data

Volume

Data Size

Data Complexity

Speed of Change

Data Sources

Variety

Long tail of data

Volume of data

Number of researchers

Slide: The Institute for Empowering Long Tail Research
Big Science <-> Little Science

- Large instruments
- High cost
- Long duration
- Many collaborators
- Distributed work
- Centralized data collection

- Small instruments
- Low cost
- Short duration
- Small teams
- Local work
- Decentralized data collection

Sensor networks for science

Sloan Digital Sky Survey
Why sustain access to data?

• Purposes
  – Record of observations
  – Reference
  – Reproducibility of research
  – Aggregation from multiple sources

• Users
  – Investigator
  – Collaborators
  – Unaffiliated or unknown others

• Time frame
  – Months
  – Years
  – Decades
  – Centuries

http://chandra.harvard.edu/photo/2013/kepler/kepler_525.jpg
How to sustain data?

- Identify the form and content
- Identify related objects
- Interpret
- Evaluate
- Open
- Read
- Compute upon
- Reuse
- Combine
- Describe
- Annotate...

Image from Soumitri Varadarajan blog. Iceberg image © Ralph A. Clevenger. Flickr photo
When to invest in data?

http://www.lib.uci.edu/dss/images/lifecycle.jpg
Data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship.

2016's New Temperature Record Proves Climate Change Is Moving Disturbingly Fast

George Dvorsky

Yesterday 3:10pm · Filed to: CLIMATE CHANGE

Global temperature anomalies for 2016. (Image: NASA/NOAA)
Rising temperature

The difference between each year’s global average temperature and the average for the 20th century has been rising steadily since the 1950s.

2016 $+1.69^\circ$ above average

20th century average: $57^\circ$

Source: NOAA

NOAA data; Los Angeles Times graphic January 19, 2017
Policy and Practice

• Scientific data are global assets
• Good policy requires good data
• Administrations are short; science is long-term
• Data often are unique and opportunistic
• Curating data is a continuous process
• Once data are lost, they are lost forever