Humanities ‘big data’
Myths, challenges, and lessons

Amalia S. Levi
IEEE Humanities Big Data Workshop 2013
**Myth #1**

‘Big Data’: It’s a recent phenomenon

**Myth #2**

All humanities data exist in archives and museums.

**Myth #3**

Humanities resources are scarce.
There have always been ‘BIG DATA.’
(They are called History).
BUT these data

....were analog

(not computationally malleable)

....were consistently ‘appraised’

(reduced in quantity to fit shelves, boxes)
Diasporas • minorities • immigrants • refugees
PRIMARY SOURCES IN THE STUDY OF COMMUNITIES

Traditional sources

- Manuscripts
- Artifacts
- Images
- Art
- Newspapers
- Books
- Testimonies...

The ‘Memory Studies’ turn

- Oral tradition
- Individual narratives
- Rituals
- Festivals
- Performances
- Monuments
- Commemorations...
**CHALLENGES**

Nature of material

- Geographic dispersion
- Time span (centuries)
- Regimes
- Languages and alphabets
- Analog vs. digital

Institutional practices

- Metadata standards (silos)
- Descriptive practices (aggregates and not item-level)
- Interoperability (libraries, archives, museums)
NEW ‘KINDS’ OF BIG DATA

• Really big...
• Everywhere...

• What is new:
  – They are *already* digital
  – Malleable
  – Quantifiable
“information overload” (Burdick et al.)

“a cultural, technological, and scholarly phenomenon” (boyd & Crawford)

“introducing ethical dilemmas” (Hitchcock)

“altering historical methodology” (Gibbs & Owens)
“Unheard of historical abundance, ...future of record scarcity.”*

Social media content:

• Fragile
• Lack of sophisticated retrieval
• Lack of links to existing cultural heritage holdings

This reality will affect historians’ and social scientists’ ability to retrieve and use them meaningfully in the very near future.

*Roy Rosenzweig, 2003
CHALLENGES CONDUCTING ‘BIG DATA’ HUMANITIES RESEARCH
CHALLENGE:
A balancing act
CHALLENGE:
Datafication of ‘un-data’
CHALLENGE: ‘Recordness’

Are they ‘recordy’ enough?

CHALLENGE:
Catastrophic violence
(Natural disasters, wars, regimes)

Destroyed:
• 5,000 manuscripts
• 200,000 pages
• 1,200,000 books
• Card catalog
• Computers, laboratories
• Rare books and special collections

Yugoslav wars:
Sarajevo’ National Library burning, 1992
CHALLENGE: ‘Archival violence’ (J. Derrida)

Archivists appraise for

- Importance
- ‘Recordness’
- Evidentiary value

...defining what will be preserved “forever”
…the percentage of US federal government data sent to NARA to be archived
Big data in humanities scholarship:
Serendipity; semantic access (contextual); visualizing dispersed collections
Traversing the network

S = Scholar
RO = Research object
DO = Digital object
Implications for archives

• Do not control “stuff” on their way in

• Collect the haystack, not the needles

• Provide tools so that scholars can find the “needles” they want

• Blur the distinction between current and historical content

Thomas et al., 2010; Weinberger, 2008
What can humanities scholars learn from scientists?
Incorporate ‘big data’ in scholarship, tie to existing cultural heritage resources
What can scientists learn from humanities? Humility

Taming ‘un-data’ or violently reduced data is not always easy or possible.
What can both communities learn from each other?

TALK
UNDERSTAND
COLLABORATE
THANK YOU!

Amalia S. Levi
iSchool, UMD
amaliasl@gmail.com