Title
3rd Annual UNT Open Access Symposium: Concluding Remarks.

Permalink
https://escholarship.org/uc/item/1dm4v96q

Author
Schottlaender, Brian E. C.

Publication Date
2012-05-21

License
CC BY-NC-SA 4.0
WHAT I HEARD
DATA

PRESERVATION
• The key reason for preservation is re-use => in fact, preservation IS re-use. [Choudhury]

• "Without a deep, rigorous exploration and investigation of preservation, we run the risk of incomplete data management." [Choudhury]
DATA
RE-USE
• Science is being transformed from something you observe in real time into something you accumulate over time and analyze later [Gutman]

• Data archives are central to astronomy today => no longer "second-class science" [Hanisch]

• Facilitating research with data isn't just about more data => it's about effective use of data [Droegemeier]
DATA
AGGREGATION
• Data increasingly need to integrate with other kinds of data [Gutman]
• Science is being transformed from something you observe in real time into something you accumulate over time and analyze later [Gutman]
• The challenges associated with integrating heterogeneous data [and they are] are non-trivial. [Renear]
• EZID => helps generate citations, with which you can get credit [Kunze]
• Working to enrich data sharing and publishing infrastructure [Hanisch]
• Need a system--and more importantly, a philosophy--for giving credit to faculty for generating, maintaining, and provisioning data [Droegemeier]
DATA

ECOLOGY
• Preservation and reuse of data for scholarship is a mosaic of partnerships and roles [Gutman]
• We need to both cooperate AND specialize [Gutman]
• Objective: insert themselves into the scientist's workflow [Strasser]
• NSB: Broad array of stakeholders [Griffiths]
W/ thanks to

McKenzie Smith
Curation Ecology: technology view

1. Storage layer
   - iRODS, S3, Palimpsest

2. Data management layer
   - IRs, ICPSR, UK Data Archive

3. Linking (or Semantic) layer
   - SFX, Semantic Web

4. Discovery layer
   - Google/Google Scholar, ICPSR UI

5. Delivery layer
   - Content interaction tools, e.g. Ajax widgets like MIT Exhibit

6. Social layer
   - myGrid/Taverna, Kepler, VREs, VIVO
Curation Ecology: functional view

1. **Storage layer**
   Bit-level persistence

2. **Data management layer**
   Metadata, policies, preservation strategies

3. **Linking (or Semantic) layer**
   Identifiers, RDF, ORE encoding

4. **Discovery layer**
   Library catalogs, Web search engines, federated search

5. **Delivery layer**
   Ebook readers, visualization tools, streaming media servers, security and ethics

6. **Social layer**
   Collaboration tools, social networking tools, VLEs and VREs

7. **Business layer**
   Cost recovery, legal/policy frameworks, virtual organizations
Curation Ecology: organizational view

1. Research Groups (individual faculty, labs, Labs and Centers)
   knowledge producers/consumers (social layer)

2. Professional Societies
   knowledge aggregators (linking layer)

3. Data Centers
   system, data storage expertise (storage, data management layers)

4. Libraries and archives
   content/data management, data linking expertise (data management layer)

5. Businesses (Publishers, IT companies)
   discovery, delivery layers

6. Universities, Funders
   business, policy layers
• Curation Ecology: data view
  – eScience
  – Data-driven research (experimental data)
  – Data-intensive research (hybrid data) [Griffin]
EXPECTATION-SETTING/
STOCK-TAKING
• ... a deep, rigorous exploration and investigation of preservation [Choudhury]
• The challenges associated with integrating heterogeneous data [and they are] are non-trivial. [Choudhury]
• ... we need to know a lot more than we do now about all aspects of data creation and management. [Renear]
WHAT I DIDN’T HEAR MUCH ABOUT
PEER REVIEW
• How is quality assured and who determines it? [Droegemeier]
• “Tread softly on the question of data quality.” [Hanisch]
DISCOVERY and DELIVERY
DATA

GOVERNANCE
• Global Research Council just organized => G20 + OECD [Droegemeier]

• Key questions …
  – Who owns it and when, and who decides? [Droegemeier]
    [There’s trouble in Whoville …]
EXHORTATIONS TO LIBRARIANS
• "When you don't have subscriptions, what've you got? What do you do?" [Gutman]
• Librarians/Archivists/Curators: "How do I organize and manage data to help students and faculty answer their research questions, and do so going into the future?“ [Gutman]
• Priority for support by libraries, by data type (low to high):
  – eScience simulation data
  – Experimental data 1 (automated collection/preparation)
  – Experimental data 2 (human involvement)
  – Higher-order computed data objects [Griffin]
• Libraries can lead the effort to create new models for scholarly communication [Griffin]
• It's your [i.e., librarians'] responsibility to get the word out about the tools available to scientists. [Hulsey]

• Librarians need to become comfortable with data [Griffiths]

• ... a deep, rigorous exploration and investigation of preservation [Choudhury]

• ... we need to know a lot more than we do now about all aspects of data creation and management. [Renear]
“THE SCHOLARLY RECORD”

Scholarly Publishing (e.g., journal articles)

Libraries

Trusted Third Parties (e.g., JSTOR, Portico)

Stable
Scholarly Raw Material (e.g., archives, data)

Archives
Data Centers
[Some in Libraries; Some Not]

Less Stable

Infrastructures largely self-contained

Scholarly Publishing (e.g., journal articles)

Libraries
Trusted Third Parties (e.g., JSTOR, Portico)

Stable

“THE SCHOLARLY RECORD”

“Scholarly” = Scholarly & Scientific
Scholarly Raw Material (e.g., archives, data)

Archives
Data Centers
[Some in Libraries; Some Not]

Scholarly Raw Material
(largely self-contained)

Scholarly Inquiry/Discourse
(e.g., blogs, wikis, open notebooks)

Very unstable
Emergent

Scholarly Publishing (e.g., journal articles)

Libraries
Trusted Third Parties
(e.g., JSTOR, Portico)

Stable

"THE SCHOLARLY RECORD"

"Scholarly" = Scholarly & Scientific

INPUTS

OPERATORS

OUTPUTS

Less Stable

???

Very unstable

Emergent

Stable
Thank You

BECS@ucsd.edu
@UCSDbeecs