UC San Diego
Other Documents

Title
3rd Annual UNT Open Access Symposium: Panel 4.

Permalink
https://escholarship.org/uc/item/9wr5x8ft

Author
Schottlaender, Brian E. C.

Publication Date
2012-05-21

Copyright Information

Subject	**Panel 4: Research Data Curation (Sayeed Choudhury (JHU), Allen Renear (UIUC))**

NOTES

Choudhury
- The key reason for preservation is re-use => in fact, preservation IS re-use
- A federation of interoperating data nodes, which tell each other what they're doing via OAIS
- If we don't think about provenance, our prospects for preservation are greatly diminished.
- Data management layers:
  - Curation
  - Preservation
  - Archiving
  - Storage
- Each layer going you the stack is necessary, but insufficient, itself, to accomplish data management
- Curation = Adding value throughout the data's lifecycle
- Data lifecycle is not linear
- "We all need to be transparent about what we're doing, how we're fumbling around. If we don't do this, we run the risk of raising false expectations The more we share, the better off we'll all be."
- "Without a deep, rigorous exploration and investigation of preservation, we run the risk of incomplete data management."

Renear
- "If we're going to develop infrastructure, tools, training, standards, practices, and policies that work, then we need to know a lot more than we do now about all aspects of data creation and management." => until we do, we're going to be in a pre-scientific period of data management.
- The challenges associated with integrating heterogeneous data [and they are] are non-trivial.
- Scientists do not have consistent notions of what constitutes a dataset => thus, we need to abandon trying to develop a coherent colloquial definition of dataset, in favor of taking a more FRBR-like approach to tha definition => cf. Systematic Assertion Model (SAM) => http://www.ideals.illinois.edu/handle/2142/30470

:: BECS :: OA2012 :: UNT ::