Web Curation within Institutions: Dealing With Researchers

Mark Phillips
UNT Libraries
IIPC General Assembly 2014
NDIIPP – Web-At-Risk Project
Archiving Web-Published Materials: A Needs Assessment of Librarians, Researchers, and Content Providers

2008

Kathleen Murray & Inga Hsieh

http://digital.library.unt.edu/ark:/67531/metadc29322/
Classification Of The End-Of-Term Archive: Extending Collection Development Practices To Web Archives

2013

http://digital.library.unt.edu/ark:/67531/metadc152437/
Web Archives are often built out of a specific mandate or collection policy
Some Web archives are more straightforward than others to explain
Mirror physical or existing collection focuses
They can be seen as an extension of a physical collection
Others are daunting to understand because of their size, scope, or specifics around the collection.
Researchers are also often consulted after a collection has begun.
Engaging researchers in using Web archives is challenging in the academic context
Issues cited include:
- Scope
- Frequency
- Depth
- Access Restrictions
- Complexity

- Lack of understanding of structure
- Transfer of question to a Web corpus:
  - Size
  - Complexity
Once we've enticed a new researcher with the wonders of a Web archive
We still have challenges:
Not being able to communicate what is held within a Web archive with researchers
Dealing with noisy, messy data
Needing tools and assistance in interacting with the Web archive
Understanding the rights issues associated with reuse within the Web archive
Terminology: SURT, seed, domain, hops, “levels”, ARC, WARC, CDX, Memento
It is often challenging to extract and give a researcher a copy of the data they are requesting.
Opportunities for collection specialists
Working with researchers to understand the kinds of information that may be helpful to capture on their behalf
Work with researchers to engage them with existing Web archives that could be leveraged for their research.
Bridge the gap between researchers and technologists to help execute research project
Thank You.

Mark Phillips

mark.phillips@unt.edu