

Wu, P. H. J., & Heok, A. K. H. (2006). Is web archives a misnomer- How web archives can become digital archives?

*Presented at the Asia-Pacific Conference on Library & Information Education & Practice 2006 (A-LIEP 2006), Singapore, 3-6 April 2006, Nanyang Technological University.*

## **Is Web Archives a Misnomer – How Web Archives Can Become Digital Archives?**

Paul Wu and Adrian Heok  
Division of Information Studies  
School of Communication & Information  
Nanyang Technological University

1

## **Traditional Divide**

- Libraries
  - Focus on Access
  - Collection facilitates search/browse
  - Based on precision in assigning classmark/ subject categorization
- Archives
  - Focus on preserving context
  - Collection retains evidence
  - Based on Provenance & Original Order

2



## National Web Archiving Initiatives

- Taken up by libraries instead of archives
- Collection resembles a digital library

3



## Why the fuss?

- Environment = exploding information
- Libraries saddled with mission to provide access to insurmountable data, information & knowledge
- Difficult to distinguish materials with value

4

## Archival Principles

- **Provenance** = helps draw boundaries of content
- **Macro Appraisal** = helps define what evidential & informational values a website may contain

=> Don't get drawn into vortex of examining selection criteria at the item level

5

## National Library of Australia

- **PANDORA** project  
“investigates strategies for the storage, preservation and access to digital data in the context of the creation of an electronic archive of library materials”

<http://www.nla.gov.au/policy/plan/pandora.html>



## Issues of Concern

- Long-term access
- Permanent archiving of such materials that are of great concern in academic and research communities throughout the world
- Limited resources = believes not possible to archive everything

<http://www.nla.gov.au/policy/plan/pandora.html>



## Assessment

- Legacy of print journal & manuscript projects extended into the web domain
- Includes online publication – journals & books
- Highly selective approach = collection segmented, patchy information, taken out of context
- Complemented by whole domain archive

8



## netarchive.dk\* (Denmark)

- Legal deposit law, from 1697 by royal order
- 1997 - Extended to cover all publication regardless of medium
- Practically, covers only static sites – dynamic ones not included (e.g databases)

\*Consortium of The Royal Library, Copenhagen & Statsbiblioteket, State & University Library, Aarhus 9



## Assessment

- Intent: Comprehensive rather than selective – legal deposit supported by snapshot archiving
- Still contains a lot of gaps in collection
- Awareness of law low and enforcement difficult
- Reliance on laws inadequate


10



## What we propose

- Adoption of archival principles to capture contextual information
- Reveal relationships between the items in the collection
- Collection kept as a whole and not re-arranged according to subject matter or themes

11



## Web Sphere Analysis (Assumptions)

- WWW an evolving set of structures that support online action
- Structures allow the production, inscriptions and experience of cyberculture
- Acknowledges difficulty in identifying and demarcating units of analysis in the hyperlinked and multi-level nature of the web
- Nature of the web = jointly produced by multiple actors, who create features and content

12

## Web Sphere Analysis

- Analyses communicative actions and relations between web producers and users developmentally over time
- **Boundaries** will change when new websites are found and added
- **Recursive** nature of a web sphere makes it dynamic

13

## Findings

- **Cannot** arbitrarily **limit** a collection to just those within a country's domain
- Making the boundary **rigid** would render the information incomplete and patchy
- **Relevant** information to a central event, concept or theme must be included
- **Hyperlinks** that bind the various websites should not be broken to preserve the integrity of the collection.

14



## The Arizona Model

- Manages materials as a **hierarchy of aggregates**
- Does not mix documents from one source with another
- All documents are kept in the order that the creator used to manage the materials
- Highlights **similarity** between 'archival' (record series) and 'web' collection

15

## Synthesis

- **Record series** : Used for organization of materials within fonds
- Study record series in relation to similar records series from another fond
- The **collaborative actions** or communication between two actors or websites is the bond that ties two separate fonds online together.
- **Linking patterns** are not just convenient configurations but actually functions that reveal, and are driven by, a need to collaborate.

16



## A Working Solution?

- Taking periodic snap shots of emerging web sphere = how collaborative actions are established online
- Record growth and spot trends
- National libraries should collaborate and link all their resources together

17

## A Working Solution?

- Document **separately** records description and administrative context
- This documentation of a separate records description and administration context can only be **tailor made** after some analysis of the materials in the websphere.
- See [Asian Tsunami Web Archive](#)

18