Roundtable: Preservation of Web-Based Materials
National Library of Australia's Experience

WIPO International Workshop on Digital Preservation
and Copyright
Geneva, 15 July 2008

Paul Koerbin
Manager, Web Archiving
National Library of Australia

Web archiving at the NLA

- Collecting (acquisition)

- Preservation management

- Access

- ## Collecting issues and risks:
  - Statutory mandate (1960) not supported by copyright law (1968) for electronic material
  - No Legal Deposit for electronic materials
  - Seek licence under Copyright Act 1968 for permission to 'archive'
  - Licence explicit to copy and provide access
    - Preservation actions? Copyright Amendment (Digital Agenda) 2000
  - Practicalities of obtaining specific permissions
    - Many publishers, user-contributors, third parties, ad hoc publishing models
    - Need to make it possible for publishers to agree

- **Collecting issues and risks:**
  - Permissions never seem fully resolved given the licence we obtain – 'grey areas'
    - A publisher' concession they can revoke
  - Legal advice is risk adverse
    - Some reluctance to seek it?
  - Scoping of harvests
    - Collateral harvesting 'time bombs'
    - Understanding/knowledge/skill of curators
    - Does not fit within jurisdictions
  - Need to be (and seen to be) responsible vs. risk taking
  - Need to meet NLA collecting objectives efficiently

4

- ## Preservation issues and risks:
  - ### Dealing with legal issues in archival context
    - Can be more complicated than in the 'live' web
  - ### We don't assume rights over content
    - But have to deal with issues that the content generates
  - ### Multiple copies
    - For preservation purposes
    - From collecting schedules
    - Harvesting collects changed content but also duplication
  - ### Potentially changing objects
    - Issues of provenance, authenticity, integrity
  - ### Losing context and content
    - An archive is not the same as the 'live' web

- ## Preservation issues and risks
  - ### Not doing enough?
    - Complex tasks
    - What will be the 'big crunch' and when?
    - What will we actually do?
    - What are legal implications of preservation actions?
  - ### Knowing what we actually have and the risks they present
  - ### Not our biggest concern
    - Because we are less advanced?

- # Access issues and risks:
  - ## Access is the purpose of web archiving and preservation
  - ## Legal Deposit and copyright changes will not resolve all issues
    - ### Efficiencies the main objective
  - ## Other legal issues can be more of a problem than copyright
    - ### Privacy, defamation
  - ## Status of 'archive' as publisher/host
    - ### Super ISP?
  - ## Are we responsible for archival content we host?
    - ### Not the creator
    - ### Don't necessarily know what we actually have

- **Access issues and risks:**
  - Access exposes archive to problems
    - Users interact; 3rd parties emerge
    - Content problems are revealed
  - Adding value in archive context
    - Library business is to add value for users
    - Risk of being seen as profiteering?
  - Limited access unsatisfactory
    - Single point access is worse than access to books!
  - Responsive access control systems
    - Granular, curator-friendly, automated
    - Don't compromise access to the archive generally

- **Lessons from experience:**
  - Much that we have to respond to may seem petty (to us)
    - Individuals 'cleaning up' their 'Web-prints'
  - Intellectual rights issues emerge as access expands (contributors, creators)
  - Non-copyright issues complex to deal with
  - Library's legal responsibilities in respect to born-digital archival content is unclear
  - Balance between risk and responsible actions
  - Does not stop progress on web archiving but is an impediment to efficiencies (adds cost)
  - Efficient access restriction mechanisms are critical