Experience of Internet Archiving in the National Library of Latvia

Baiba Zvaigzne
National Library of Latvia
baiba.zvaigzne@lnb.lv

Aggregation and Management of Audiovisual Content in the Digital Space, October 4-7, Vilnius

© LNB
Contents

- Why?
- How?
- Today
- Future
Why?
Legal Deposit Law

• Latvian Legal Deposit Law defines responsibility and duty of the National Library of Latvia (NLL) to collect and archive nationwide publications, including electronic resources
Strange electronic resources

- Homepages
- Blogs
- Websites

are publications, defined by domain name
Reasons

Content

• Important resources:
  – science; education; culture;
  – society in general.

• Web materials are characterized by:
  – relatively short lifespan;
  – highly changeable.
How?
Archiving approach

- Selective archiving:
  - Collection development policy of the NLL;
  - Complete quality control of archived websites;
  - Not limited in national domain.

- Thematic, event-based collections
Tools

• Software:
  – The Web Curator Tool (WCT) / Heritrix;
  – WERA / NutchWAX;
  – Wayback Machine.
Web Curator Tool
Workflow

Identification → Analysis and WCT target definition → WCT task processing

- Repeated according to time table
- Result quality control:
  - Yes → Archiving
  - No

Archiving
Selection criteria

- Inclusion
  - Latvian culture;
  - Education in Latvia;
  - Science in Latvia;
  - Innovations in Latvia;
  - Professional organizations and associations in Latvia;
  - Latvian society;
  - Latvian politics and government.
Selection criteria

• Exclusion:
  – Discussion lists, chat rooms, bulletin boards and news groups;
  – Drafts and works in progress;
  – Games;
  – Online shops;
  – News sites and portals;
  – Online daily newspapers;
  – Promotional sites and advertising.
Today
At present

- 3 employees from different departments of NLL
- Started at year 2005, partly project funded
- Grown from 30 to 3000 websites
Problems

- Technical issues
- Management issues
- Legal issues
Results

• Harvested and archived approximately 3000 websites (~ 1 TB)
• 4 special collections:
  – Websites of political parties before parliament (Saeima) election in 2007
  – Website of the President of Latvia before and after Presidential election in 2007
  – The Latvian Song and Dance Festival 2008
  – European Parliament election and Municipal election 2009
Future
Future

- Building up new thematic collections on nationwide events
- Monitoring of target websites
- Maintaining of archived material
- Making archived material publicly available
Thank you for your attention!

baiba.zvaigzne@lnb.lv