Digital Library at the National Library of Sweden - Today and Tomorrow

Eva Müller
Sr. Director
Digital Library Department
National Library of Sweden
eva.muller@kb.se

Overall view

- Digital Library – a definition
- Digital Library at the National Library of Sweden (NL) – the mission
- Digital content
- Collections development
- Development of supporting infrastructure
- Development of services
- Cooperation and some of the missing pieces
What is a digital library?

Many definitions…..
Important characteristics are:

collection of information is managed
there are services associated with this collection
the information is stored in digital formats
the information is accessible over a network to its users

= a managed collection of information, with associated services, where the information is stored in digital formats and accessible over a network to its users
Digital content is any content created, accessed, shared, used or preserved in a digital format.

Source: Creating A Digital New Zealand: New Zealand’s Digital Content Strategy
Digital Library – the mission

The tasks of the National Library are:

- to collect, catalogue and preserve all publications produced in Sweden (this activity has been based on the Legal Deposit Act since 1661, currently work is underway on an extension of this act to include online publications)
- to enable access to materials by the general public and other stakeholders
- to serve as a research library within arts and humanities

In order to support these tasks in the digital environment following strategies are used:

- Development of collections by
  - harvesting the Swedish Web
  - delivery, based on agreement, of born-digital materials from producers
  - digitization
- Development of supporting infrastructure
- Development of services
Collections development – Web Harvesting

Harvesting of the Swedish Web
• Over 10 years
• Strategies:
  – 2 – 3 snapshots annually
  – Swedish newspapers online on a daily basis
• Problems with harvesting password-protected web spaces, dynamic/interactive content etc.
• Restrictions on access http://www.kb.se/kw3/
• 2007- 700 millions files (approx 50 millions websites) = 40 Tb,
  approx 1,500 various MIME types (format) (97 % html, gif, jpeg or pdf)
During this fall
  – a test of selective harvesting approach (by subject, event, content type …)
  – new storage format – WARC (ISO)
  – upgrading of tools
  – from tape to disk
  – development and documentation of strategies, policies and best practices
• Challenges – harvesting of audio visual content, access to the archive, search services
Collections Development – Delivery from Content Providers

No legal deposit act covering digital information distributed in network, hopefully by 2010

Cooperation with universities since 2000; since 2007 – new agreements with newspapers publishers
  • Basic infrastructure in place
  • Tools supporting ingest and creation of high quality packages are tested
  • At The National Library, integration within the typical library processes (librarians deeply involved)

During this fall:
  • Approaching more content providers
  • Improving tools
  • Implementation of automated processes
  • Improving the current infrastructure
  • development and documentation of strategies, policies and best practices
Collections Development - digitization

Past:
Focus on high quality output only
(approx 1000 objects in the repository, 100’ digitized)

Lessons learned from previous projects:
• High quality/low quantity? Low quality/high quantity?
  What is good enough?
  – A range of production lines for various material types
  – Separate development and production
  – Reduce dependency by dividing digitization process into several separate processes (production of images, cataloguing, publishing)
  – Implement automated processes
  – Re-use metadata when possible
  – Plan staffing carefully!
  – Plan incorporation of project results in good time
Some figures and plans in the short term

- Digitization plan and planning for 2008
  All Swedish prints up to 1600, Collijn bibliography and a number of medieval manuscripts, images and maps
- Implement production lines for a range of quality level outputs
- Production lines for mass digitization of newspapers and journals – test in house target: 200 000 images to be delivered to Tel plus project
  50 millions newspaper pages to be digitized – funding is an open issue

<table>
<thead>
<tr>
<th>Category</th>
<th>Volume/Number</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Books (including &quot;Rare books&quot;)</td>
<td>Volumes</td>
<td>3300000</td>
</tr>
<tr>
<td>Newspapers</td>
<td>Volumes</td>
<td>75000</td>
</tr>
<tr>
<td>Journals and other serials</td>
<td>Volumes</td>
<td>73500</td>
</tr>
<tr>
<td>Government publications</td>
<td>Volumes</td>
<td>120000</td>
</tr>
<tr>
<td>Other type printed material not classified</td>
<td>Number</td>
<td>10500000</td>
</tr>
<tr>
<td>Manuscripts</td>
<td>Number</td>
<td>4000 sh m *</td>
</tr>
<tr>
<td>Maps</td>
<td>Number</td>
<td>310000</td>
</tr>
<tr>
<td>Photographs</td>
<td>Number</td>
<td>500000</td>
</tr>
<tr>
<td>Engravings</td>
<td>Number</td>
<td>150000</td>
</tr>
<tr>
<td>Drawings</td>
<td>Number</td>
<td>10000</td>
</tr>
<tr>
<td>Posters</td>
<td>Number</td>
<td>600000</td>
</tr>
<tr>
<td>Postcards</td>
<td>Number</td>
<td>600000</td>
</tr>
<tr>
<td>Sheet music</td>
<td>Scores</td>
<td>210 sh m **</td>
</tr>
</tbody>
</table>
Wider access to rare materials; tricky to expose...
protecting the original material from risks of harm
Digitization as an action of preservation
Development of supporting infrastructure

Why?
To manage content-rich digital library
  Support: creation, delivery, ingest, storage, metadata-description, preservation…

Essential parts of infrastructure are:
• A preservation platform
  – Original objects, limited access
• An institutional repository and a platform for digital library collections
  – Derivates of original objects in end-user friendly formats to be disseminated, accessed and delivered to services

Focus on:
Cost effective management of digital content
Flexible solutions (support for standards, metadata management, relationships between objects, integration with other services, version management…)
Software solution for multiple applications
Strategic priority: open source
The content of the archive

- A broad concept of the archive
  - Archive of Harvested Web content
  - Deposited objects based on agreement (from 2009? according to legislation)
  - Outputs from digitization projects (newspapers, journals, manuscripts etc.)
  - Audiovisual material (in cooperation with The Swedish National Archive of Recorded Sound and Moving Images)
Why an open source solution?

• Gives libraries more control over their software and stored content
  – it eliminates dependence on a vendor
• Benefits from the quite small digital library community – facilitates collaboration
• Makes it possible to customize and build upon already developed software
  – no need to build everything from scratch
• Makes it possible to implement fast new standards and best practices
  – digital content space is changing rapidly

Using of Open Source in public sector is supported by national agencies – for example Verva - Swedish Administrative Development Agency
Why Fedora?

Flexible Extensible Digital Object Repository Architecture (Fedora)

Open source

Strong community – both developers & users

Users – wide variety of institutions and organizations – public sector, private companies, national libraries

Variety of implementations and usage of Fedora (digital preservation platform, institutional repositories, parts of commercial library products, digital library collections, distributed digital libraries … )
Development of services

Interfaces:
Simple interfaces in place (ingest, search, metadata dissemination and reuse, browsing..)

Current work on:
Metadata management tool
   Non print materials – manuscripts, images, graphical materials, maps

End-user discovery services
   Search and browsing interface
   Relations between existing services (union library catalogue Libris, local library catalogue Regina) and digital library services

Dissemination services
   OAI interface
   Contribution to other services (World Digital Library, European Digital Library)

Challenges:
Integrated services between cooperating institutions
Cooperation – building shared solutions

National level
LMA sector
- Standards and best practices
- Technical solutions
- Digitization plans
- Long-term preservation (LDB-centrum)

International level
Open Source communities
- ex. Fedora commons
- EDL (CENL)
- TEL & TEL plus (ocr-red images)
- Nordbib

Bilateral cooperation agreements
What is missing? A cross-sector point of view

- Resources for
  - Mass digitization of materials at risk (hard to reallocate resources from current budget)
  - Digital preservation

- Skills
  - Digital library issues are technically intensive and create demands on new skills
    New professions are emerging – digital content manager/ digital curator?

- Solutions
  - Digital preservation

- Widely adopted data and metadata models and standards to manage some of the collections typical for other sectors than libraries (manuscripts, maps, images, objects)

- Authority name files

- Rights management & copyright issues

- National level policies and strategies
Thank you for your attention!

Photos:
Anna Guldager, Maja Kristin Nylander, Claes Jansson, Andrea Davis-Kronlund
Istvan Borbas, Olav Nyhus, Gabriela Muj-Lindroos, Jens Östman