Digitisation at NLF and the National Digital Library

Karo Salminen
March 15th 2012
Sofia
Overview

- National Library of Finland as a organisation
- Digitisation at NLF
- Joint approach of libraries, museums and archives to access and preservation for digital material – the National Digital Library
Many roles of NLF

- A research library – part of University of Helsinki
- A national library:
  - Receiving legal deposits since 1707: all written publications deposited to NLF
  - Collecting web sites and electronic publications since new law came into force in 2007
- A rather new role: A national service and development centre for the Finnish library sector
Centre for Preservation and Digitisation

- Rather large organisation unit of NLF
  - Majority of the permanent employees carrying out tasks on microfilming of newspapers and magazines, and conservation
  - All the Finnish newspapers are microfilmed
  - Investigating new possibilities for replacing the traditional workflow:
    - Piloting digital depositing of newspapers with various publishers

- In-house digitisation
  - Better control over quality and handling of the physical originals
  - Structural markup is labour intensive
Digitisation workflow – from and back to Collections
Tools for mass digitisation

- An automatic scanner
- Software (docWorks by CCS) for postprocessing:
  - layout analysis
  - text recognition (OCR)
  - structure analysis
  - quality control
- Operators for controlling quality and marking up document structure
Mass digitisation in numbers

- Mainly newspapers and magazines
- In 2011 majority of magazines published between 1810-1944 had been digitised
- Possible to increase the rate of annual digitisation to 4 million pages – only (long-term) funding is missing
Challenges in the digitisation workflow

- Distance between locations
  - Logistics and item tracking
- Cooperation over organizational boundaries
Digitisation and preservation

- Digitising for access and to protect the physical originals, not for preservation per se
- Original copy is the target of preservation
- However, why not treat the digital copy as a target of preservation?
  - In the end, only digital copy can, in principle, last forever
  - Practical reasons: avoid redigitisation
- Not forgetting born-digital material
Criteria for digitisation - policies

- Critical masses – coherent collections
- Protection of the physical original
- Demand
- Content: national or international value

- Authentic and high-quality digital surrogates for research use and public reuse
- Enhancing visibility, accessability and usability of unique collections
Digitisation planning

- Public sector is the primary source of funding; cooperation when possible to produce larger collections and to reduce overall costs
- Digitising volume is dependent on resources, possibility to cooperate and level of quality
- A three-year plan, focus on:
  - Newspapers & magazines (fragile & widely used source material, but difficult to provide access to original items)
  - Cassette tapes (brittle tape)
  - Ephemera (popular)
User access

- Currently, two independent solutions exist:
  - One especially for newspapers and magazines (bespoke software)
  - Another for monographs (DSpace with a browser module developed in-house)
LAATOKKA
Sanomia Sortavalasta ja Itä-Karjalasta.

No 24.
Torstaina 15. 5. Keijutlaa.

1882.
Of the temperature of the Air, and soil of Lapland.

CHAP. III.

Of the temperature of the Air, and soil of Lapland.

We have seen how Lapland is situated; let us now proceed to other particulars of it. This is very near the Pole, and appears from its latitude, to be quite in the same degree as the summer heat of the sun's rays never set; and on the contrary, in winter it is never quite still, which Kirkenes lies is but forty days, and the three hours in the night the body of it is something darkened. So that its rays appear not; yet there is so much light, that they continue their work all the day. Indeed the sun's rays is not to be taken out of the whole country, since part of it lies nearer, and part farther distant from the Pole, and of these too some parts are more vast, and some more to the West, from whence its that with some of them the sun's course above the horizon for so many days as it prevents. And in the summer it never sets, and goes below the earth, yet neither does it rise above the same, as if it were killed and gently slips along the horizon for the most part like an in the winter when lowest it is not so much beth in which is the reason that the sun has one night for some months, yet every day the sun comes to hear, that it makes a kind of twilight. The moon rises that in the absence of the sun there are two twilights, one in the morning, the other in the evening, in which these poor remainders of night provide that the night should not be utterly destructive. And by how much the Sun is farther away, the light of the Moon is diminished. Hence Peter Planci says that when the Moon shines they go a fishing, and dispart all other occupations that are to be done without doors and when it does not, if the air be clear, even the light of the stars much brighter, and the darkness, the horizon of the night is much lengthened, and there is light enough for the dispatch of all businesses, which is further illuminated by the whiteness of the snow. The Air of Lapland is cold, but steady and clear, and consequently very wholesome, being much justified by the winds which are here very frequent and violent. It has the aspect to me by eye-witnesses, that there rises a certain wind out of the sea, which beginning to blow musts presently thick and dark clouds even in the midst of Summer, that they utterly hinder the light, and in the winter drives the snow with such force and violence, that if any one...
Preservation of digital objects

- Preservation copies stored securely and independently of the access copies
- Relying on the national effort for a more complete solution
Crowdsourcing some of the manual work

- Structural analysis and other post-processing tasks require manual work
- OCR’d text is not proofread by anyone in the postprocessing process, however
- Launched a pilot in early 2011, in attempt to crowdsource proofreading, with a game developed by Microtask
DISH Award 2011 – Best crowdsourcing project: Digitalkoot

In Mole Bridge you write the words where the computer failed. Watch the tutorial!

In Mole Hunt you try to identify words the computer has misread. Watch the tutorial!
Crowdsourcing: next steps

- Engage researches and individuals to fix the OCR’d text while reading the actual articles
- Sequel to Digitalkoot project: Kuvatalkoot, funded by European Social Fund 2011-2013
  - Marking of images and/or articles
  - Indexing them
Collaboration on national level: National Digital Library
Libraries, archives and museums together

- Funded by Ministry of Education for years 2008-2011 and 2011-2013; the most extensive cooperation project between libraries, museums and archives so far in Finland

- Goals:
  - Single point of access to all digital material, including catalogues – a joint public interface (in operation 2011/2012)
  - Digitisation of key materials in libraries, archives and museums
  - Digital preservation solution for digital cultural material
  - Central aggregator for the European Digital Library Europeana
NDL Governance 2011-2013

Executive group / appointed and chaired by the Ministry of Education and Culture

Steering group / appointed and chaired by the Ministry of Education and Culture

User Interface: maintenance and development: the National Library

Consortium of libraries, archives and museums / chaired by the NL

Support Group of libraries, archives and museums / chaired by CSC

Long-term preservation Project / CSC (IT Center for Science)
Single point of access

- Bringing the wealth of information sources to one, single public interface
- Services of libraries, museums and archives integrated into one system
  - Renew loans, buy pictures and order materials
  - Central indexing of the databases and information sources provides quick and versatile searches
  - Emphasis on usability
- Suffered from major delays, but being currently implemented
What has been digitised?

- During 2008-2011 16 million objects were digitised, for 2011-2013 goal is to digitise 7 million objects
- Photos and maps, newspapers, parish registers, diaries, museum objects, herbariums
- Within NDL several seminars on digitising audio, photographs, and digital preservation have been held
Preserving the digital heritage

- A national, shared solution for preserving digital heritage
- What’s been done:
  - Prerequisites for member organisations
  - Process modelling
  - Specifications of metadata
  - File types based on media type
  - Identifiers
  - In OAIS terminology: SIP and AIP specifications
Lessons in cooperation

- At first, conceptual boundaries between institutions hindered efficient cooperation
- Discovered needs to upgrade and existing systems and standardise information (especially in archive and museum sectors)
- In the process, growing interoperability between systems and improving information coherence
Further information

- Digitisation Policy of NLF
- Preservation Policy of NLF
- NLF’s digitised material online (not subject to copyright):
  - http://www.doria.fi/handle/10024/69173
  - http://digi.nationallibrary.fi
- Crowdsourcing:
  - http://www.digitalkoot.fi
- National Digital Library: www.kdk.fi/en