National Library of Finland: Digitisation within Context of National Digital Library and Infrastructure Developments

The National Library of Finland is meeting digitisation challenges front on. This article will provide a broad overview of steps taken between 2007-2009 in renewing digitisation processes, production and workflows under a scheme funded by the Education Ministry for Mass Digitisations. The focus of the article is to provide a digitisation production environment perspective into National Digital Library architecture and long term infrastructure developments in Finland. Museums, Archives, Libraries and Audio-Visual memory organisations with vast physical collections to be digitised, are faced with the challenge of examining their production environments within the context of national infrastructure development. The article provides a glimpse into such a development, from the field of digitisation at the National Library of Finland.

Paradigm Shift in Digitisation Infrastructure Development: Towards Enduring, Sustainable and Open Access

Cultural and research sectors are moving towards provision of global, open and free access to digitised content. Change is driven by user community shifts in how digital content is accessed, used and re-used. These community shifts have a direct impact on how digital content creation, capture, management, dissemination and preservation is enabled by cultural organisations; shifts in how sustainable IT infrastructure is developed to cater for research and community needs.

At a recent visit to the Library of Congress in May, Deanne Marcum, Associate Librarian for Library Services (LoC) referred to infrastructure shifts in the US, towards enduring access in digitisation; a shift towards national infrastructure developments and national consortia to ensure unified, sustainable solutions to access and long term preservation. The term was coined during discussions of a federated and state based system and the need for co-ordinated approaches in the US for infrastructure developments due to tightening budgets. In Europe, the European Commission’s push (ec.europa.eu, ec.europa.ec/i2010) towards National Digital Libraries might be seen as a parallel development with the exception that each European country is responsible for its own infrastructure development. In Finland, the responsibility for the National Digital Library Initiative is shouldered by the Ministry of Education under by Secretary General Minna Karvonen and the national DL’s infrastructure development by Deputy National Librarian; Director, National Library Network Services Kristiina Hormia-Poutanen.

This article will provide an overview of how digitisation at the National Library of Finland is currently situated in 2009, into the broader context of national infrastructure developments – providing a digitisation production environment perspective to developments. Memory organisations have an opportunity to play a proactive role influencing national infrastructure decisions made and ensuring national infrastructure investments underway meet community needs in the future. Close co-operation among higher education and cultural organisations is needed to ensure open and free access to digitised content, are and remain core values of national infrastructure developments.

Current State of Affairs: National Infrastructure Developments and Digitisation

In a small country as Finland, consolidation of resources is vital for government funded service delivery; in the long run digitisation production, provision of access and preservation need to be technologically and economically sustainable. The National Digital Library architecture was developed in 2008 with wide consultation. A common understanding evolved in which the architecture was viewed as one production line between memory organisations (Museums, Archives, Libraries, Audio-Visual) and a National Digital Library (www.kdk2011.fi).
Specialization of Production Units by Format Type:

The National Digitisation Centre situated at Mikkeli, is a unit of the National Library of Finland and an integral part of DigitalMikkeli Digitisation Cluster (www.digitalmikkeli.fi). The National Digitisation Centre specialises in digitisation of paper and audio formats (digi.lib.helsinki.fi). Its production focus is in conversion of:

- Historical Newspapers (national/regional)
- Scientific Serials
- Journals
- Monographs
- Ephemera
- Maps
- Parchments
- Audio


- A public access interface for the information resources of libraries, archives and museums (2008 – 2011)
- Digitisation of the most essential information resources of libraries, archives and museums
- Development of a long-term preservation solution for cultural heritage materials (a detailed plan in 2010)
- Increasing competence in Finland
- Operate as an aggregator for the European Digital Library Europeana.

The first practice drafts for a National Digital Architecture in Finland, was depicted a National Digital Library Framework, where the National Digitisation Unit of the National Library could be best depicted as a back end, digitisation production unit. Its purpose being to serve the National library, library sectors in general, Helsinki University and customers by digitising paper and audio based material, converting these into digital format and ingesting into a uniform, access and preservation system. All other memory institutions (Museums, Archives, Libraries, AV), of which there are now 35 in the current National Digital Library Initiative, were to be seen in a similar fashion, as back end production units holding vast physical collections in paper, audio, video formats or artefacts of cultural or research value, each responsible for their own digitisation production environments, each responsible for provision of digital content for ingest into a National Digital Library. The initial architecture mirrored closely the early architecture of Europeana, the European Digital Library (www.europeana.eu), a service developed for Museums, Archives and Libraries to provide access to critical mass of European digital content. The added layer for the Finish DL architecture solution was an early notion of incorporating a long term preservation system or trusted repository in step with provision of access, thus strategically positioning to consolidate national resources.

One Production Line: Digital Library Architecture and Production Environments...

Three concurrent initiatives funded by the Education Ministry, as if by sheer osmosis, have been complementing each other since 2007.

2. Long Term Preservation System Initiative, functional requirements currently being defined and to be implemented by 2015.
The initial DL draft architecture was to focus its attention to management of digital content from ingest onwards with a clear distinction of excluding production environments as out of scope. While at the onset, this appeared a sensible proposition, the Digitisation Centre for one, as a back end production unit, brought forward during the consultation phase a view that there was a clear need for incorporating a unified metadata approach within the context of any DL solution.

There are times, when practice can inform theory – therein is the healthy synergy between the national infrastructure project and the mass digitisation production project. The dilemma of any production unit is that specifications are required at the onset of digitisation production as metadata is created, captured and added during digitisation production processes - not at point of ingest. Clear specifications were therefore needed at the onset of a production line, be it for printed, audio or digital format with regards metadata specifications and standards used to ensure interoperability, authenticity and trustworthiness of digital material ingested. A SIP package alone, as depicted in the DL architecture, is only as good as its metadata contained…. the metadata contained might not be of high quality in terms of enduring access. Investment in good quality metadata and processes is well worth pursuing.

The Digitisation Centre had been operating in the field since 1999, the early days of the newspaper digitisation from microfilm. Since 2004, the National Digitisation Centre had been using METS (Metadata Encoding and Transmission Standard), as container or wrapper for digital objects and associated metadata created in its digitisation production. In theory, the production output, METS SIP packages produced were well placed for easy ingest into a potential national DL. In practice, the mass digitisation project 2007 was to review what NLF METS packages contained and to examine quality of metadata gathered starting with monographs production.

The Digitisation Centre had been operating in the field since 1999, the early days of the newspaper digitisation from microfilm. Since 2004, the National Digitisation Centre had been using METS (Metadata Encoding and Transmission Standard), as container or wrapper for digital objects and associated metadata created in its digitisation production. In theory, the production output, METS SIP packages produced were well placed for easy ingest into a potential national DL. In practice, the mass digitisation project 2007 was to review what NLF METS packages contained and to examine quality of metadata gathered starting with monographs production.

Zooming in into One Production Line – Are We There Yet ?
Digitisation Process Modelling and Workflow Design at the National Library

I had the great pleasure of joining the National Library of Finland in early 2007, at an exiting crux in developments in digital libraries, nationally and internationally. The word buzz... is perhaps what best describes the working atmosphere at the moment, with several concurrent projects each working in unison to achieve common goals. The Ministry of Education granted the Digitisation Centre funding for 2008–2009 for developing library wide mass digitisation processes. The aim of the funding was to establish efficient logistics and workflows for digitisation and for securing enduring access to a critical mass of digitised cultural content. The project objectives were set as:
Level of Structural Analysis

- to ensure one production line; efficient, library wide logistics, processes and workflows are modelled and renewed where needed
- to ensure quality of metadata used, captured and packaged throughout the production line is adequate for access and long term preservation needs
- to ensure appropriate tools are put in place for tracking and managing workflows

As background information, the trigger for the project was a visit in late 2007, from the Digitisation Centre to the British Library Mass Book Digitisation Centre. Back then the digitisation project leader was Mr. Neil Fitzgerald and a digitisation studio had just been established on-site BL premises by CCS (Content Conversion Specialists, [www.ccs-gmbh.de](http://www.kansalliskirjasto.fi/extra/bulletin/)). The business model was BL supported by Microsoft, outsourcing its book digitisation services to CCS and CCS providing scanning services to BL on-site and post processing services remotely. The National Library of Finland had been co-operating closely with CCS since 2004, since adopting docWorks in its production. Being one of the few Digitisation Centres in Europe with full in-house digitisation production, the library had acquired substantial know-how and expertise in digitisation. A strategic position was taken to pursue developing this asset and opportunity - retaining and developing Finnish know-how in digitisation and opting out from an outsourced model. A strong argument was put forward to decision makers that it is vital a memory organisation fully understands its digitisation processes and thus production requirements - a learning process that would benefit all. This strategic decision has proved quite fruitful.

Process Modelling, Library Wide Working Groups and more Working Groups

At the kick of meeting of the mass digitisation project, Anttolan Hovi, Mikkeli March 2008, a blue print was presented internally to the library for a need to review production process in digitisation and to identify gaps in use, creation and capture of administrative, bibliographic, technical, structural and long term preservation metadata attached to digital objects created in digitisation. A high level digitisation process solution for mass digitisation of monographs was modelled on CCS solution provided for BL and modified over 2008 to reflect National Library of Finland (NLF) organisational environment for in-house production and to incorporate quality metadata capture, use and re-use throughout the production processes. A library wide, process oriented approach was taken in which collections, cataloguing, transport, digitisation, conservation and network services were seen as one production line. Through several meetings and working group sessions, the process was modelled using QPR and is currently in final draft version for documentation. The top level approach was presented for NLF top management for acceptance as the solution presented would require strategic decisions affecting library operations, some more critical than others, impacting also change management required with regards to work practices. A summary of some of the gaps identified and new measures being put to place are listed below.

Strategic Decisions Taken In Renewing Production Processes

BARCODE ID’s – Unique ID for Physical Items at Collections

- NLF decided to adopt and implement unique physical Bar Code ID’s to items from collections being sent for digitisation. The decision ensured items can be tracked throughout the digitisation process and that a 1:1 relationship could be established between an original source material, physical item and its equivalent digital representation. The Physical ID would be indicated in Fennica Catalogue and in MARCXML – MODS fields.

Minimal Bibliographic Record for non-catalogued items

- Exhaustive cataloguing to a National Bibliographic Record standard is labour intensive and a bottle neck from a mass digitisation workflow perspective slowing selection and shipment of material to digitisation. A decision was made that non-catalogued items can be entered into Fennica Catalogue at a pre-defined minimal bibliographic entry level to allow for items to be sent to digitisation and that post-cataloguing to full National Bibliographic record could be tested utilizing the digital object as a post process procedure, after digitisation.

Two Bibliographic Records will be created into Fennica

- A decision was made to use two bibliographic records in Fennica; physical and digital. Both bibliographic records will be interlinked. The two bibliographic record solutions is based on a library metadata internal working recommendation that it is better suited for FRBR cataloguing rules and future cataloguing needs.

Unique and Persistent ID’s for digital objects

- A decision was made to create persistent ID’s for digital objects and to level of structural elements defined (page, issue, article, section). NBN:URN’s would be generated and used as persistent ID’s using urn.fi resolution. Link will be used for example in bibliographic record of digital object.

Metadata Re-use and Enrichment

- Decision was made to automatically ingest bibliographic metadata directly from Fennica into docWorks ensuring more efficient processes and less human error. Conversely, a decision was made to enrich Fennica catalogue after digitisation, by ingesting back and enriching the catalogue with previously missing metadata elements 'table of contents' and 'page numbers'. This is done by creating an enriched MARC record for the printed original at digital object export step. The MARC record is then ingested back to the catalogue.

Level of Structural Analysis
Level of structural analysis would need to balance mass digitisation and volume targets with work effort required in production. Level of structural analysis would be defined according to material type; newspapers, monographs, serials, audio – or by projects (specifications) Mass digitisation will be focused towards historical Newspapers with limited structural analysis at issue level, not article level for the time being. Structural analysis of monographs will be at significant section level, mainly chapter level. Table of contents will be identified for all. All material types will have following structural elements defined: pages, footnotes, running title, tables, advertisement; image (captions, picture categories (exLibris, chart, picture, portrait, scenery, item, event, decoration, map, notes)

Metadata formats already in use by the Digitisation Centre since 2004 were formally accepted as NLF library standards in digitisation. These standards being:

- MODS and MARCXML for descriptive and bibliographical metadata (http://www.loc.gov/standards/mods/)
- MIX for image technical metadata (http://www.loc.gov/standards/mix/)
- PREMIS for preservation metadata (http://www.loc.gov/standards/premis/)
- PREMIS for rights management metadata.

A further decision was made for recommending METS and PREMIS to be incorporated into the overall National Digital Library metadata standards portfolio as SIP package containers. The fine detail with regards to the standards and elements used (MODS, MARCXML, MIX, PREMIS), and contained within the METS SIP package, was under review in the mass digitisation project, starting with printed material type - monographs.

**METS Profiles for Monograph, Newspapers, Parchments, Audio....**

By having METS profiles, NLF ensures that the digital objects and the contained metadata are well documented, interoperable and preservable. The strategic position is to develop METS Profiles for each printed material type over the course of 2009-2010 envisaging multiple production lines ingesting to a National DL/long term preservation system. In practice, implementing the task was approached systematically by defining one production line, one format, and one material type first, starting with monographs – then replicating work done with minor modifications were needed for newspapers, parchments, audio etc. In-house knowledge and practical experience in digitisation production by NLF IT staff was well harnessed and key to the profiling work underway. The planned changes will be peer-reviewed and after finalisation a METS profile for Monographs is created and submitted for METS editorial board approval.

One benefit of using an integrated tool for digitisation production is the level and detail of metadata that can be captured during the process. This applies particularly to the recording of administrative metadata of the various process steps.

The major changes in METS are:

- Inclusion of the printed and the digital bibliographic records
- New technical and provenance metadata about events and agents related to the object and individual files inside it using PREMIS schema.
- Update of used metadata schemas to their latest versions

**Wrapping Up.....**

The article aimed to provide an overview of national infrastructure developments towards a National Digital Library and an overview how National Library of Finland digitisation production environment is situated into the wider context. One production line and metadata are seen at the core of sustainable infrastructure development for enduring access. Working though the production details in the Mass Digitisation Project 2007-2009, has had the benefit of informing the national infrastructure of standardisation and metadata practices required throughout the production chain: administrative metadata, bibliographic metadata, technical metadata, structural metadata and long term preservation metadata – and only then wrapping up the fine bundle into a METS SIP package.

As project manager, I have had the great pleasure of working with a dream team. Director of Microfilm, Conservation and Digitisation, Ms Majlis Bremer-Laamanen allows innovation and creativity. Director of IT Development Juha Hakala, also chair of Technical Working Group who was roped into chairing the internal library metadata working group for his metadata expertise, the team gained a view into the national perspective and the national DL initiatives a view into a digitisation production environment – as one production line. In addition, special thank you is due to Eeva Murtomaa for her cataloguing expertise, Karo Salminen for his PREMIS expertise and Harri Ahonen from Collections for his contributions. Last but not least, Jukka Kervinen who is the technical IT guy working through the fine detail for the METS Profile, work that nobody else can do. I sincerely believe each organisation has a Jukka, and if not – they should.

Ms Tiina Ison
Senior Analyst &
Project Leader, Mass Digitisation
National Library of Finland
tiina.ison@helsinki.fi

Written and revised by Tiina Ison.