



{BnF

IFLA International Newspaper Conference

“Newspaper Digitization and Preservation.
New prospects.
Stakeholders, Practices, Users and Business Models”

11-13 April 2012
BnF, Paris

With the support of:





THE BENEFITS OF DIGITISATION: enrichment and usability

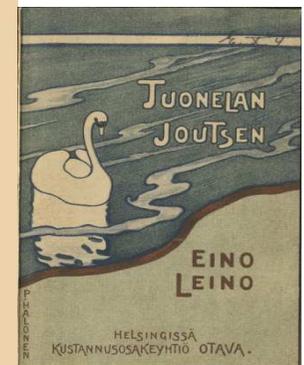
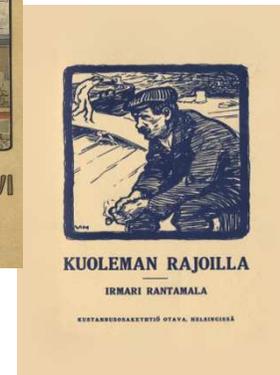
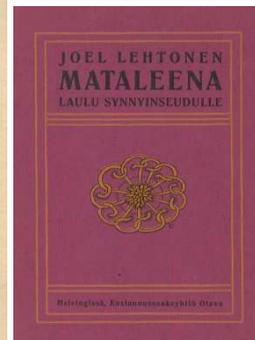
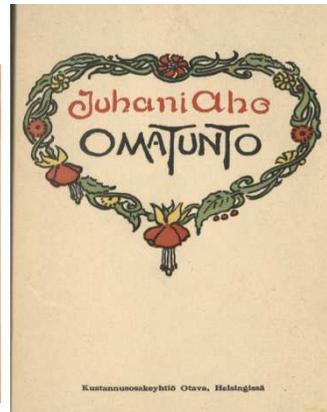
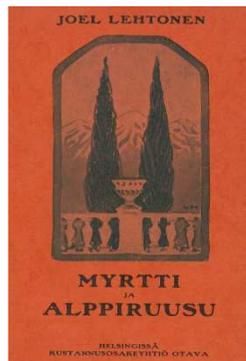
Name Majlis Bremer-Laamanen
Date 12.4.2012
Event INTERNATIONAL NEWSPAPER
CONFERENCE



”WANTED” ENRICHING CONTENT

(<http://digi.nationallibrary.fi>, <http://www.doria.fi/handle/10024/4194>)

- Newspapers - > 2 million pages, the Historical Newspaper Library
- Journals - > 3,7 million pages, free to 1910, in all legal deposit libraries to 1944
- Medieval parchments.
- Books - > travel, novels, Dissertations 17th century, Save the Book
- Ephemera - > industrial price lists
- Sound - > national sound archive, C-cassettes



CENTRE FOR PRESERVATION AND DIGITISATION

BENEFITS OF DATA AND ENRICHMENT

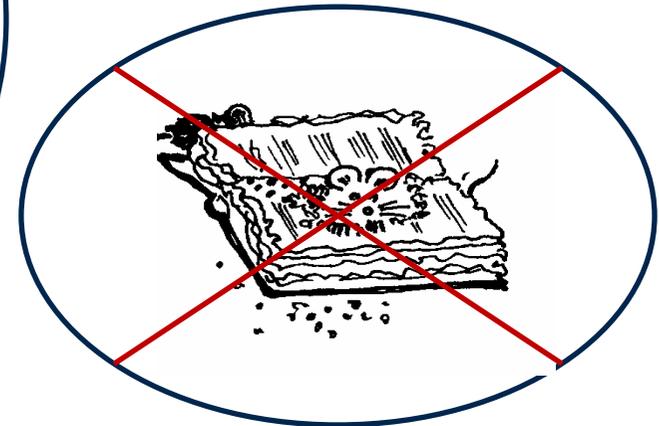
- Comprehensive digital collections in the ownership of the Nation /NLF
 - Case: Newspaper Publishers and NLF
- Use and reuse of digital content, new users and purposes
 - Community enrichment: retrospectively improve the quality through distributed efforts: Digitalkoot, Kuvatakkoot
 - Professorship to be established: Digital collections research: Helsinki University/Modern Languages/ language technology and NLF and other partners
 - Descriptive data created during digitisation is attached to the library catalogues improving searchability

CASE NEWSPAPER PUBLISHERS & NLF

- THE WIDEST POSSIBLE USE OF NEWSPAPERS
 - The NLF has had Legal Deposit Right since 1707, including all newspapers since the first one in 1771-
 - Users. the newspaper customers, researchers, students, library customers

- USER ANGLE:
 - Newspaper houses: sites with exclusive supply of newspaper content
 - The copyright society Kopiosto has the right to make extended collective licensing agreements (2012): newspapers, journals, journalists, photographs...
 - NLF: a wider use of the newspaper content: via the library sectors in Finland
 - Long term preservation: METS, National digital library (KDK)

The Extended Collective Licensing & KOPIOSTO



VISIBILITY OF THE 20 CENTURY

- Newspapers' interests:
 - Extended use, to benefit the 20th century golden age in electronic format
 - Sophisticated services via the newspaper's own site
 - Publicity and a small income via the Historical Newspaper Library / National Digital Library
 - Partnership in digitisation
 - Long term preservation
 - METS-ALTO, metadata development together with the NLF
 - Benefits of the crowdsourcing projects of the NLF

VISIBILITY OF THE 20 CENTURY (2)

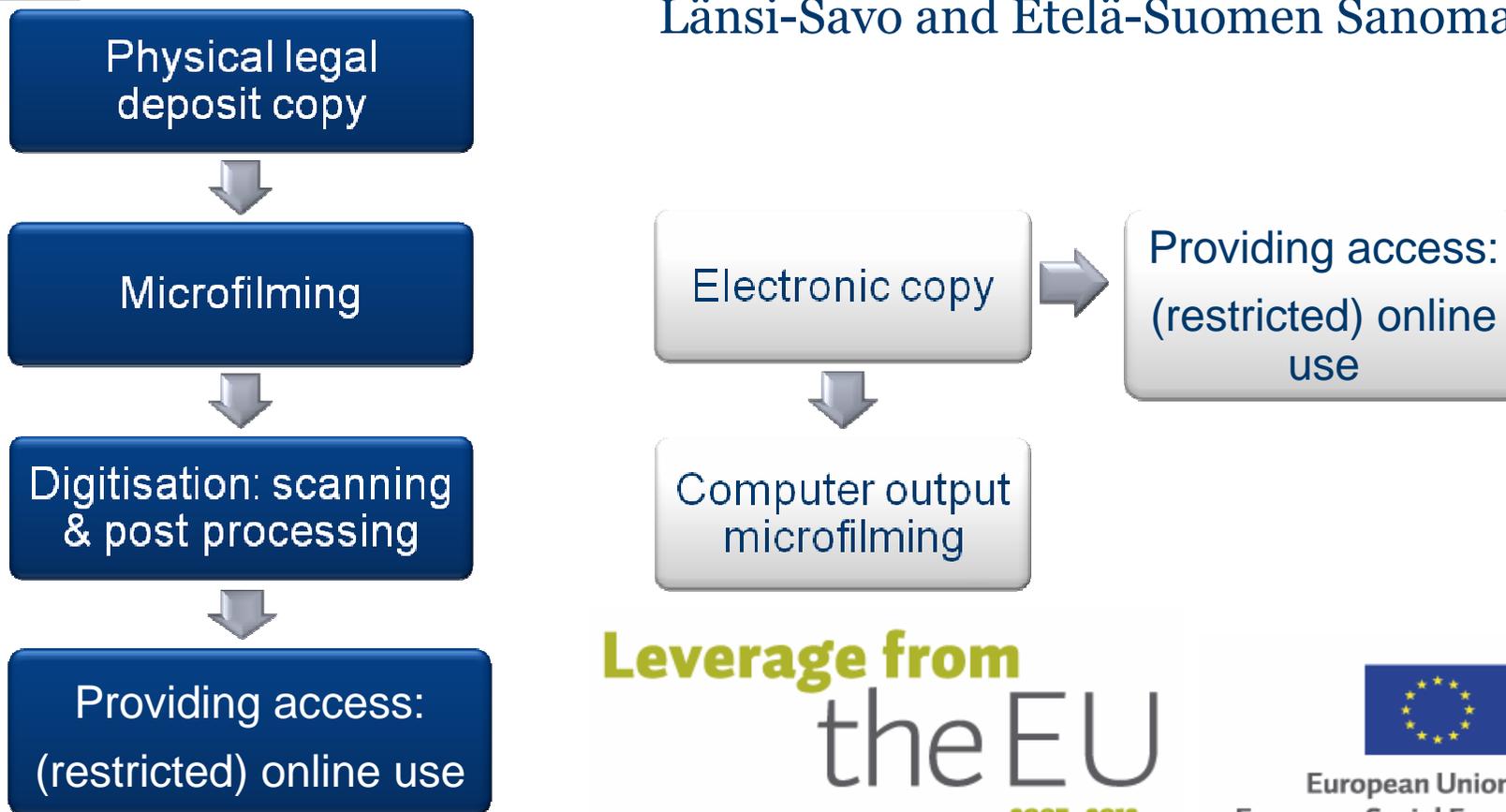
- Library:
 - Extends the supply of newspapers via the Historical Newspaper Library
 - The older part of the 20th century via the NLF:s service to the Library sectors: "the moving wall"-principle.
 - Electronic delivery of newspapers
 - Computer output microfilming
 - Rediness to develop and share: crowdsourcing, metadata- development and enrichment, Kuvatakkoot, professour-results
- Kopiosto:
 - Collects the rights form the rightholders
 - Distributes the shares to the rightholders
- Project planning for digitisation:
- National newspapers.
 - Maaseudun tulevaisuus, HBL
- with regional newspapers ongoing:
 - Etelä-Suomen Sanomat, Länsi-Savo

Leverage from
the EU
2007–2013



E-delivery: Two newspapers...to 2013

Länsi-Savo and Etelä-Suomen Sanomat



Leverage from
the EU
2007-2013



INTERWOVEN AND EXPANDING COLLABORATIONS



CENTRE FOR PRESERVATION AND DIGITISATION

USE AND RE-USE OF DIGITAL CONTENT

CASE 1 : CROWDSOURCING

- FIRST EXPERIENCE: DIGITALKOOT February, 8th 2011 ongoing, with Microtask.com for ocr-correction
- SECOND : EU-FUNDED KUVATALKOOT PROJECT 2011-2013 for marking articles and images
- DIGI= to digitise
- KUVA= here: image, photo, article
- TALKOOT=people gathering to work together voluntarily (without payment)
- METS-ALTO: makes it possible

Leverage from
the EU
2007-2013



DIGITALKOOT: Mole Bridge

DIGITALKOOT
CURRENT PROJECT: 19TH CENTURY AAMULEHTI

← BACK TO FRONT PAGE
f POST SCORE TO MY WALL

fesfusteltawafsi

IMPOSSIBLE
(TAB)

ja typeriä. Raastuwan

LEVEL 1
SCORE 0
MOLES 5
SAVED MOLES 0/1
BLOCK BUFFER

=	Na	Nu	Nä	Nä	Öö	fff	fff							
-	A	a	U	u	Ä	ä	Ö	ö	s	f	f	l	k	t
	GG	ZZ	BB	BB	Nn	WW	ww							
	S	G	T	I	P	V	B	N	n	W	M	w	m	

Current project

The current project consists of the National Library's archive of the issues of the newspaper Aamulehti from the end of the 19th century.

So far 74,581 people have visited the Digitalkoot web site. Volunteers have contributed a total of 238,263 minutes (4,989,354 microtasks) of their time.

DIGITALKOOT Facebookissa

Tykkää 4,604

DIGITALKOOT Digitalkoot CrowdsourcingDE interview from CrowdConvention 2011.

Interview Mieltinen Founder vimeo.com

Claudia Pelz (Founder Crowdsourc im Gespräch Mieltinen (C Founder Microtask.ct Rahmen der Crowdconvt in Berlin.

16. elokuuta kello 15:26

Facebookin yhteisöllittäminen

Top-6 today Mole Hunt

1. Virpi J. 322800 points	2. Vera S 131800 points
---------------------------	-------------------------

KUVATALKOOT, 2011-2013

- Extending crowdsourcing to marking articles/images
- To enable users to collect articles/images in their own interest
- To enable them to add metadata to their collections
- Might enable researchers to give people important tasks
- To share the results with the Library and the Community
- METS-ALTO: coordinates for substance



USE AND REUSE - THE NEED TO CHANGE

CASE 2: Professorship in Digital Content Research 2013-2017

- The Information is primarily to be sought in the net
- The importance of mathematical methods and electronic research content is growing
- It is important to develop and foresee new ways of use:
- to intensify the search possibilities of digitised content:
 - Data mining from large collections /automation
 - The influence of the digital recording, format, on the user experience
 - The impact of the content on search possibilities
- New ways, crowdsourcing
- Helsinki University (HU)/ Faculty of Arts, Modern Language, Language Technologies and NLF et al. Decision in May, 2012

USE AND REUSE

CASE 3: metadata input to library catalogue

- Descriptive data created during digitisation is attached to the library catalogues improving searchability
- Ephemera as a case:
 - Boxes registered as a catalogue post when sent for digitisation
 - Each item catalogued with mini-descriptive data during the digitisation process, including search terms
 - > feeded back into the library database for searchability



DATA ENRICHMENT FOR THE FUTURE

