DIGITIZATION PROJECT OF KINDRED LANGUAGES

Jussi-Pekka Hakkarainen
Eesti Keele Instituut
23.10.2013, Tallinn

DIGITIZATION PROJECT OF KINDRED LANGUAGES

The National Library of Finland (NLF) is executing a pilot project (7/2012-12/2013) that aims for digitizing and publishing of Finno-Ugric material for the benefit of the linquistic research within the Kone Foundation Language Programme (2012-2016). Decision on the follow-up project for 2014-2016 is pending.

In addition to the pilot project for Kone Foundation Language Programme, the National Library of Finland has started to develop the Uralica portal for digitized books in and about the Uralic languages as of March 2013. The project aims for linking the metadata of digitized books and make them more visible and searchable for all the language users.

DIGITIZATION PROJECT OF KINDRED LANGUAGES

During the pilot project the NLF has produced the research infrastructure (repository and OCR editor) and takes care of co-operation with the Finnish (**Helsinki University Library**) and Russian partners (**National Library of Russia**, NLR and **National Library Resources**).

The material was digitised from the collections of the NLR. This is the first time that material published in the former Soviet Union has been made freely available for public use in the NLF (or any foreign?) data systems.

DIGITIZATION PROJECT OF KINDRED LANGUAGES

The focus is on the Finno-Ugric languages, which suddenly became socially important, at the beginning of the Soviet era in the 1920s and 1930s. No contemporary literature was digitised.

The researchers made the selection of materials for the digitization plan in autumn 2011. The selection consists of 17 000 pages of monographs in **Veps, Ingrian, Mari (Meadow and Hill Mari) and Mordvinic (Erzya and Moksha)** languages and around 25 000 pages of newspapers in Mari and Mordvinic languages. Monographs are mainly school and text books and in many cases they are translations from Russia to the local languages.

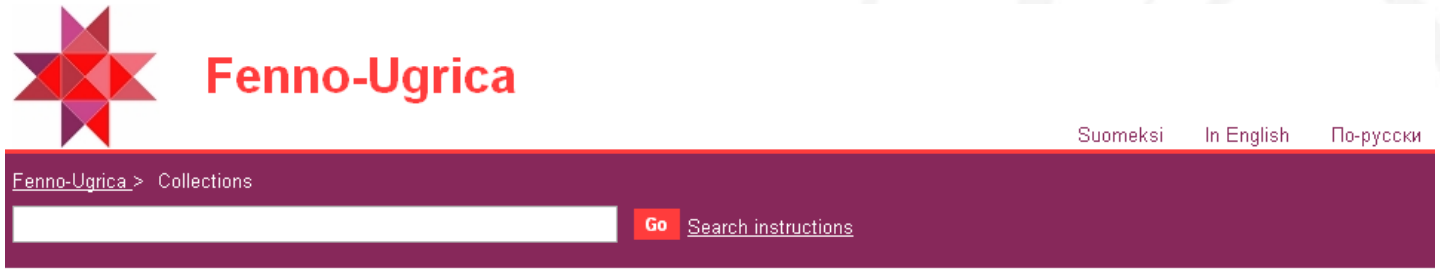
DIGITIZATION PROJECT OF KINDRED LANGUAGES

Since the Kone Foundation Language Programme has an objective of free access to the materials, NLF published the material as its own repository and provided an open access **without geographical or IP restrictions**.

In order to publish the material as public domain, the copyrights regarding material was needed to be cleared. The research on copyrights was conducted by Moscow-based **National Library Resource** in winter 2013.

Public domain allows NLF to donate the language-resources after editing to the **FIN-Clarin** for the benefit of other research (linguistic) communities.

DIGITIZATION PROJECT OF KINDRED LANGUAGES



Fenno-Ugrica

Fenno-Ugrica is the National Library of Finland's digital collection of Finno-Ugric publications. The Fenno-Ugrica collection includes monograph publications in Ingrian, Veps, Mari (Hill Mari and Meadow Mari) and Mordvinic (Erzyan and Moksha) languages and newspapers in Mari and Mordvinic languages from the 1920s and the 1930s. All in all, the collection consists of more than 120 monographs and nearly 20,000 pages of newspapers.

The material of Fenno-Ugrica has been produced by the National Library of Finland in the Digitisation Project of Kindred Languages, which is a part of Language Programme of Kone Foundation. The material Fenno-Ugrica collection belongs to the collections of the National Library of Russia (St. Petersburg), where the publications have been digitised. The digitised content of this collection is published based on the research on copyrights, which was conducted by Moscow-based copyright organization, National Library Resource.

Within the Digitisation Project of Kindred Languages, the National Library of Finland has developed an open source code OCR editor that enables the editing of machine-encoded text for the benefit of linguistic research. Permissions for the editing of the material of Fenno-Ugrica will be granted mainly for the researchers of Fenno-Ugric languages and the permissions will be administrated by the Digitisation Project of Kindred Languages. Requests and enquiries: kk-fennougrica@helsinki.fi

Collections

- Monographs [153]
- Newspapers [2092]

Search Fenno-Ugrica

- Titles
- Authors
- By Issue Date
- Subjects
- By Submit Date
- Browse by languages
- Communities & Collections

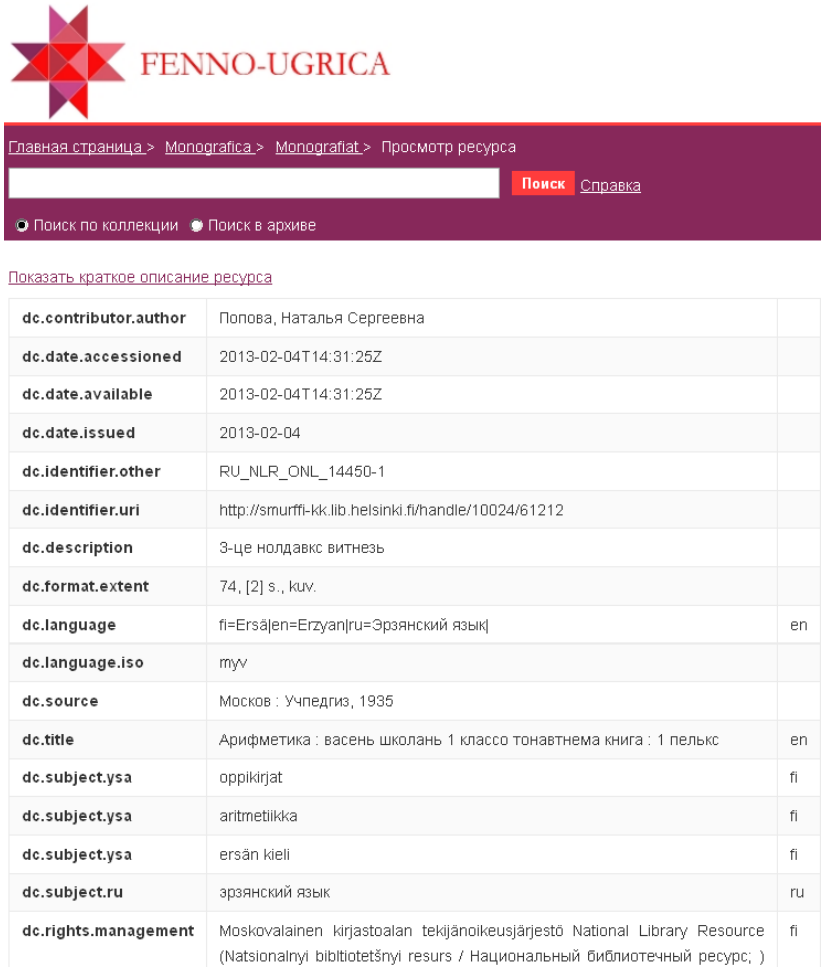
My Account

- Login
- Register

KONEEN SÄÄTIÖ

1640 KANSALLIS KIRJASTO

DIGITIZATION PROJECT OF KINDRED LANGUAGES

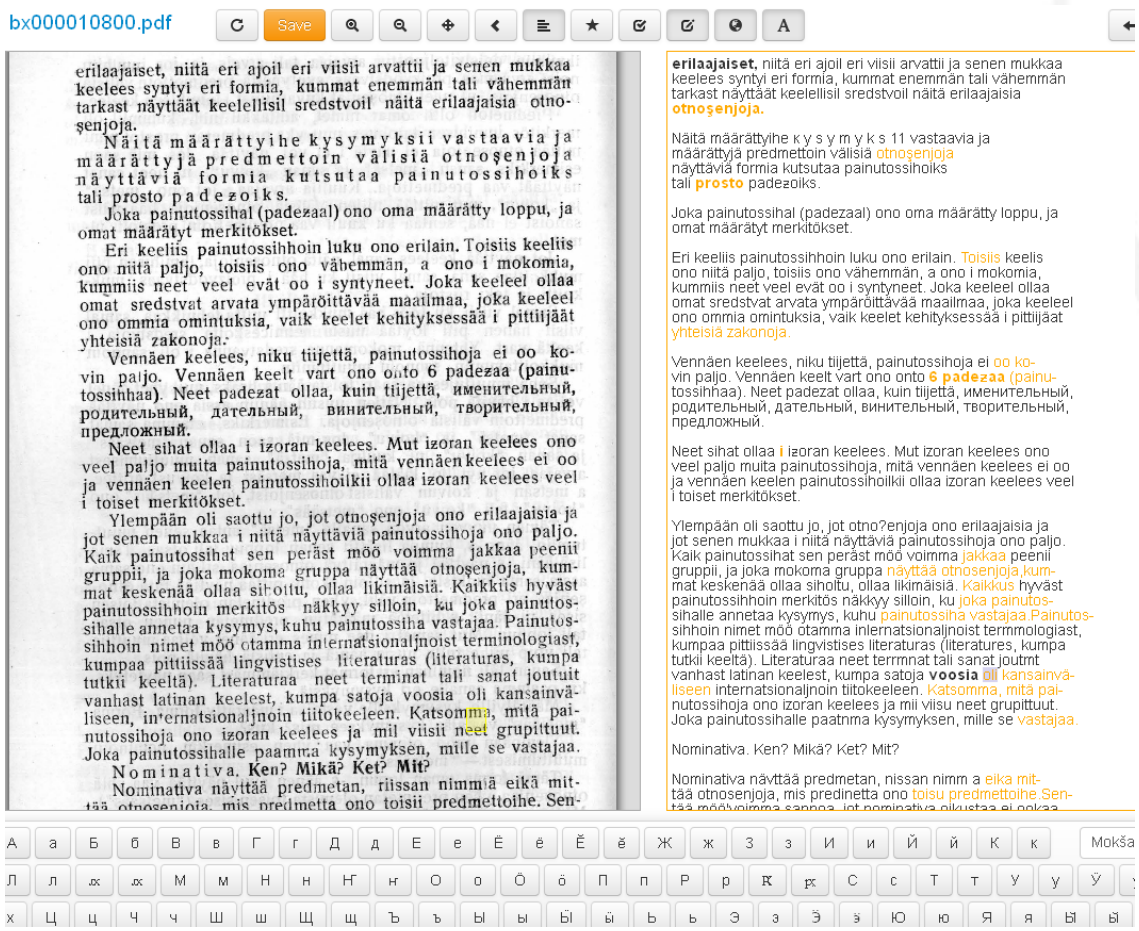


dc.contributor.author	Попова, Наталья Сергеевна	
dc.date.accessioned	2013-02-04T14:31:25Z	
dc.date.available	2013-02-04T14:31:25Z	
dc.date.issued	2013-02-04	
dc.identifier.other	RU_NLR_ONL_14450-1	
dc.identifier.uri	http://smurffi-kk.lib.helsinki.fi/handle/10024/61212	
dc.description	3-це нолдавкс витнезь	
dc.format.extent	74, [2] s., kuv.	
dc.language	fi=Ersä|en=Erzyan|ru=Эрзянский язык|	en
dc.language.iso	myv	
dc.source	Москов : Учпедгиз, 1935	
dc.title	Арифметика : васень школань 1 классо тонавтнема книга : 1 пелькс	en
dc.subject.ysa	oppikirjat	fi
dc.subject.ysa	aritmetiikka	fi
dc.subject.ysa	ersän kieli	fi
dc.subject.ru	эрзянский язык	ru
dc.rights.management	Moskovalainen kirjastoalan tekijänoikeusjärjestö National Library Resource (Natsionalnyi bibltiotetšnyi resurs / Национальный библиотечный ресурс;)	fi

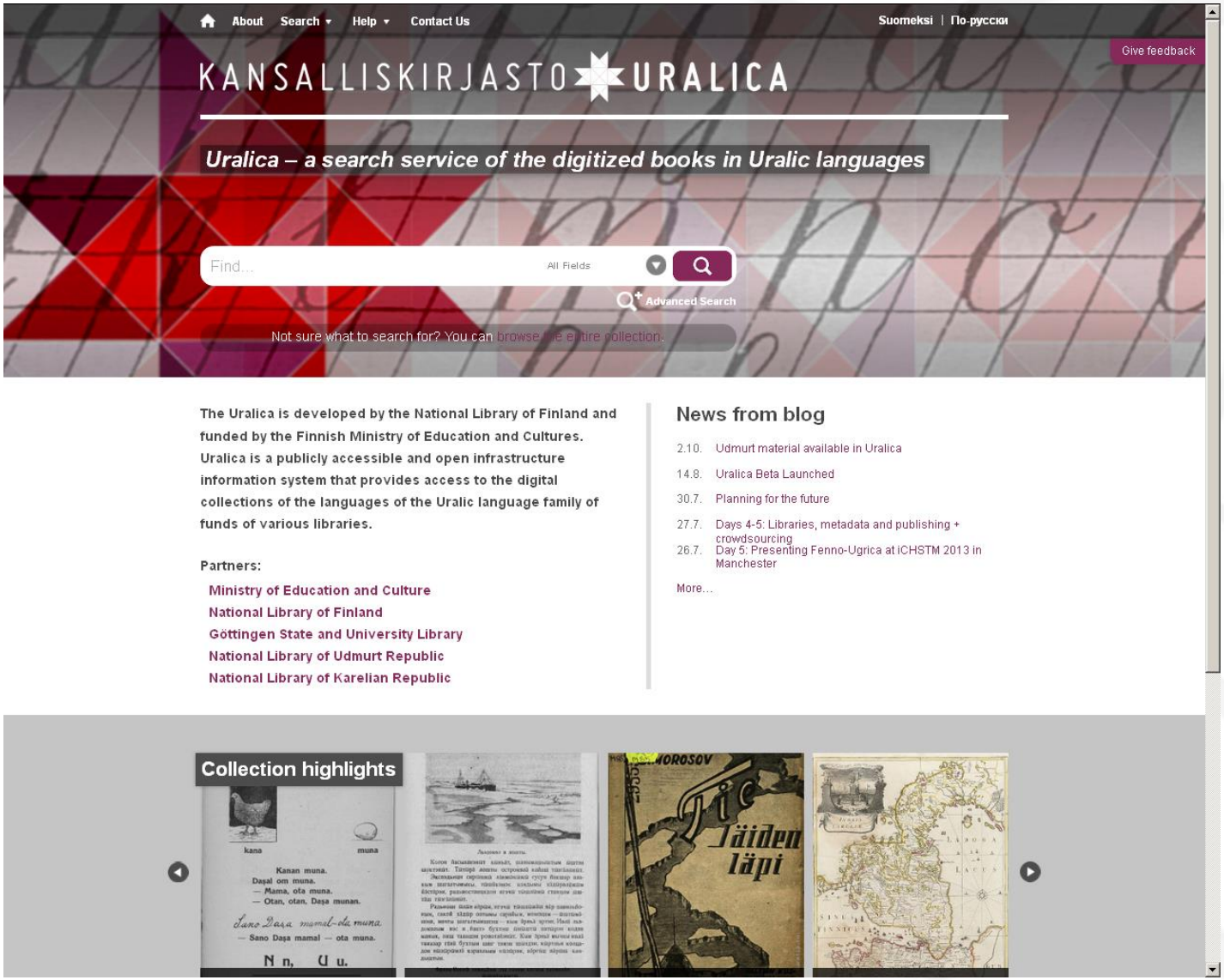
The material is catalogued directly to the repository in **Dublin Core** format, but the metadata will be linked to the local library cataloques too.

In order to ease the access to the material, the material will be linked to **Europeana** and the **National Digital Library** of Finland and it can be browsed through its interface, **Finna.**

DIGITIZATION PROJECT OF KINDRED LANGUAGES



The NLF has developed an **OCR editor** to support the research use of the material. The editor allows text that has undergone a process of machine identification to be edited for the purposes of linguistic research. Editing the characters will be done through the XML Alto files. Once the text will be corrected, the material will be re-uploaded to Fenno-Ugrica.


About
Search
Help
Contact Us
Suomeksi | По-русски
Give feedback
KANSALLISKIRJASTO URALICA
Uralica – a search service of the digitized books in Uralic languages
Find...
All Fields
Advanced Search
Not sure what to search for? You can browse the entire collection.
The Uralica is developed by the National Library of Finland and funded by the Finnish Ministry of Education and Cultures. Uralica is a publicly accessible and open infrastructure information system that provides access to the digital collections of the languages of the Uralic language family of funds of various libraries.
Partners:
Ministry of Education and Culture
National Library of Finland
Göttingen State and University Library
National Library of Udmurt Republic
National Library of Karelian Republic
News from blog
2.10. Udmurt material available in Uralica
14.8. Uralica Beta Launched
30.7. Planning for the future
27.7. Days 4-5: Libraries, metadata and publishing + crowdsourcing
26.7. Day 5: Presenting Fenno-Ugrica at iCHSTM 2013 in Manchester
More...
Collection highlights
kana
muna
Kanan muna.
Dasal om muna.
— Mama, ota muna.
— Otan, otan, Daşa munan.
— Sano Daşa mamal — ota muna.
Jäiden läpi

DIGITIZATION PROJECT OF KINDRED LANGUAGES

The digitisation of materials in Uralic languages in the national libraries operating in the Russian Federation has grown significantly during the past few years. Many libraries have established their own digital collections and given the public either full or restricted access to the materials.

At the same time, the dissemination and harmonisation of information has been overlooked in the enthusiasm for digitisation and it has become clear that there is no control over how much and which Uralic materials are getting digitised. This complicates the accessibility of materials and the systematic planning of digitisation projects and may result in redundant work, a highly inefficient use of resources.

DIGITIZATION PROJECT OF KINDRED LANGUAGES

In order to control the production of the National Library of Finland has launched a cooperation project funded by the Ministry of Education and Culture, intended to generate a shared and open information system infrastructure for the Uralic language materials digitised in different libraries.

The grant will not be used to digitise material or to produce material to be digitised. Instead, the intention is to coordinate the cooperation between partners. From the perspective of accessibility and usability, the objective of the project is to link the metadata of previously digitised material to a shared information system which will then enable the sharing of the material to third parties.

DIGITIZATION PROJECT OF KINDRED LANGUAGES

Digitised material will be made available through the open source VuFind software. The publication system also promotes the accessibility and use of the digitised material internationally, among both the international academic community and speakers of Finno-Ugrian languages. In this way the publication system complies with the target of advancing a culture of openness and interaction.

Publishing the material in Uralica also makes searching for content easier. The material in the project will be provided for open access also through Finna, the public interface of the National Digital Library and the material will also be searchable through Google and other common search engines.

DIGITIZATION PROJECT OF KINDRED LANGUAGES

The agreements on releasing the metadata and links have been made with:

- State and University Library (Göttingen)
- National Library of Russia (St. Petersburg)
- National Library of the Republic of Karelia (Petrozavodsk)
- National Library of the Udmurt Republic (Iževsk)

Coming soon?

- Institute of the Estonian Language (Tallinn)
- National Library of Mordovia (Saransk)
- National Library of Komi (Syktyvkar)

DIGITIZATION PROJECT OF KINDRED LANGUAGES

Project Web Site
www.nationallibrary.fi/services/digitaalisetkokoelmat/finnougric_en_ru.html

Fenno-Ugrica Collection
fennougrica.kansalliskirjasto.fi/

Uralica Portal
http://uralica.kansalliskirjasto.fi/

Fennio-Ugrica Blog
blogs.helsinki.fi/fennougrica/

DIGITIZATION PROJECT OF KINDRED LANGUAGES

Jussi-Pekka Hakkarainen
Project Manager
National Library of Finland
Research Library

The National Library of Finland
P.O. Box 26 (Teollisuuskatu 23)
00014 University of Helsinki
kk-fennougrica@helsinki.fi