



# **An Introduction to the OCR Webinar**

## Making the Impact on Research and Society

Jussi-Pekka Hakkarainen  
Project Manager  
Digitization Project of Kindred Languages  
National Library of Finland

**OCR Webinar, 5 March 2014, Helsinki**



# Digitization Project of Kindred Languages

## Scope of the project

The **National Library of Finland** is executing the Digitization Project of Kindred Languages in 2012-2016 that aims for digitizing and publishing of material in Uralic languages for the benefit of the linguistic research within the **Kone Foundation [Language Programme](#)** (2012-2016).

By the end of 2015, [Fenno-Ugrica](#) collection will consist of ca. 115 000 pages of monographs and around 98 000 pages of newspapers and it will be the largest resource for the prints in Uralic languages online. The focus is on the languages, which suddenly became **socially important** at the beginning of the Soviet era. No contemporary literature will be digitized.



# Digitization Project of Kindred Languages

## Erzya Language as a typical example

Mordvinic language, **Erzya**, was converted into a medium of popular education, enlightenment and dissemination of information pertinent to the developing political agenda of the Soviet state. The “deluge” of popular Erzya literature, 1920s-1930s, suddenly challenged **the lexical orthographic norms** of the limited ecclesiastical publications from the 1880s.

Newspapers were written in orthographies and in word forms that the locals would understand. Schoolbooks were written to address the separate needs of both the adults and children. **New concepts** were introduced in the native language. It was the beginning of a renaissance and period of enlightenment.



# Digitization Project of Kindred Languages

## Objectives of the Language Programme

One of the key objectives of the **Language Programme** is to foster a culture of openness and interaction in linguistic research. This entails, for example, **the unlimited availability, accessibility and usability of source material** and research results via virtual library.

**The participation of the language community** is needed and the Language Programme also supports this in creating new material and enhancing the usability of old material by applying the methods of **citizen science**, which is a form of interactive research that includes the participation of researchers, students and interested citizens etc.



# Digitization Project of Kindred Languages

## Meeting the objectives

In order to meet these objectives, it is not essential ensure the availability of language materials, but also to make it easy for different users to approach and use them.

Free, open access to the material ensures that it can be used by both the communities, the **academic** and the **native-speakers**.

However, the mere digitization is not enough to meet the requirements. Something that connects the researchers and the speakers of the kindred languages needs to be discovered.



# Digitization Project of Kindred Languages

## OCR editor as a connector

**Can crowdsourcing tools serve two masters?** Into our opinion, the OCR editor can operate as a tool that connects the native-speakers, or the citizen scientists with the linguistic research.

This applies to digital humaniora in general. For example, the OCR editor is used or will be used in following projects:

- **Digitization Project of Kindred Languages** – National Library of Finland
- **Dictionary for the Old Literary Finnish** – Institute of Languages in Finland
- **Bank of Finnish Literature** – University of Helsinki



# Digitization Project of Kindred Languages

## Values for Society and Research

How to measure the impact? Will the communities change by the resource in beneficial ways that can clearly be identified?

1. **Utility Value** – I'm glad it is here for me!
2. **Existence / Prestige Value** – I never have used these books, but I am glad that its there for others!
3. **Education Value** – Wow! I didn't know that!
4. **Community Value** – That looks good, let's use it!
5. **Inheritance / Bequest Value** – I'm glad they can benefit from this too!

(Indebted to Simon Tanner for his [Balanced Value Impact Model](#))





# Digitization Project of Kindred Languages

## Impact on Society and Research

Huge impacts on society and research are expected, but we don't really know yet what the impact will be and how valuable that is.

## Get beyond the number games!

Once the digital resources and tools for enriching the data will be used, a change will take place and a **wider set of opportunities** will be available to different communities, like native-speakers and academic.





# Digitization Project of Kindred Languages

## Resources and Services Online

### Project Web Site

<http://www.nationallibrary.fi/services/digitaalisetkokoelmat/kindred.html>

### Fenno-Ugrica Collection

[fennougrica.kansalliskirjasto.fi/](http://fennougrica.kansalliskirjasto.fi/)

### Fenno-Ugrica Blog

[blogs.helsinki.fi/fennougrica/](http://blogs.helsinki.fi/fennougrica/)

### Kone Foundation Language Programme

<http://www.koneensaatio.fi/en/grants/language-programme/>



# Digitization Project of Kindred Languages

## Contact Details

Jussi-Pekka Hakkarainen  
Project Manager

The National Library of Finland  
Research Library  
Digitization Project of Kindred Languages  
P.O. Box 26 (Teollisuuskatu 23)  
00014 University of Helsinki  
[kk-fennougrica@helsinki.fi](mailto:kk-fennougrica@helsinki.fi)