

## INTRODUCTION

The digitization of materials in Finno-Ugrian languages in the national libraries operating in the Russian Federation has grown significantly during the past few years. Many Russian libraries have established their own digital collections and given the public either full or restricted access to the materials. The assumption is that the digitization and improved accessibility to Finno-Ugrian materials both old and new promotes the language learning of native speakers of these languages. These Finno-Ugrian languages can be thought to represent a crucial resource or the worldview of their communities, cultures and traditions – and culture, traditions and the worldview are best transmitted in free use of the native language.

At the same time, the dissemination and harmonisation of information has been overlooked in the enthusiasm for digitization, and in the discussions conducted with the Russian national libraries it has become clear that there is no control over how much and which Finno-Ugrian materials are digitised in the Russian Federation. This complicates the accessibility of these materials and the systematic planning of digitization projects and may result in redundant work, a highly inefficient use of resources.

The same problem is reflected in the spotty accessibility to the material. From the perspective of the end users, learning to use and administer the many user interfaces requires not only time and IT know-how, but also skills in the dominant language. There is no system which could be used to search, browse or access all digitised Finno-Ugrian materials. The standardisation has also been insufficient, and typically no harmonised quality requirements have been set for the digitised materials. This has prevented the use of this material in, e.g., linguistic research. We have for some time hoped to receive a suitable proposal from the Russian libraries which would enable cooperation between them and the National Library of Finland in terms of both harmonising the digitised materials and the development of their quality specifications.

## THE SUB-PROJECT AND ITS GOALS

To promote the objectives described above, the National Library of Finland will launch a cooperation project funded by the Ministry of Education and Culture, intended to generate a shared and open information system infrastructure for the Finno-Ugrian language materials digitised in different libraries. The National Library considers its role in terms of combining different entities with expertise in digitization, the presentation of digitised materials and experience in international cooperation which can open a door to the international community for the Russian libraries.

This sub-project of the *Digitization project of kindred languages* is spread over several years. The first period began in March 2013 and will continue with a special grant from the Ministry of Education and Culture until the end of 2013. This first project period will generate practices and technological implementations which will be used during the infrastructure project proper during 2014–2016. In the first stage, the work will be discussed with the National Library of Russia (St. Petersburg), the National Library of the Republic of Karelia (Petrozavodsk) and the National Library of the Udmurt Republic (Izhevsk) which have the IT and cultural resources required to participate in the planning and implementation of the early stages of the project.

The *Suomalais-ugrilaiset kielet ja internet* (Finno-Ugrian languages and the Internet, SUKI) project, coordinated by the Department of Modern Languages at the University of Helsinki will also be involved in this project. SUKI is a four-year project funded by the Kone Foundation and supported by the National Library, intended to construct a prototype of an automatic system which could maintain a portal for smaller websites in the Finno-Ugrian languages. The first version of the prototype is estimated to be usable at the beginning of 2014.

The intention of this first project phase is to also enable new forms of international cooperation and to widen the cooperation, if possible, to cover other memory institutions managing and digitising Finno-Ugrian language materials, not just in Russia and Finland, but everywhere in the European Union. The infrastructure should be ready and available by the summer of 2016.

#### **SUB-PROJECT PARTICIPANTS AND ORGANISATION**

The National Library will create the infrastructure required for the presentation and compilation of the digitised Finno-Ugrian materials in the *Digitization project of kindred languages* during 2012–2016. Funded by a special grant from the Ministry of Education and Culture, this sub-project of the infrastructure project will be implemented by the Research Library Services of the National Library of Finland in cooperation with the national libraries of Russia, Karelia and the Udmurt Republic. Copyright cooperation will be conducted with both the abovementioned libraries and the National Library Resource (Moscow), a nation-wide copyright organisation in the library sector.

The National Library of Finland has a great deal of experience in cooperation with Russian libraries. During the past years the cooperation has been closest with the National Library of Russia, but the National Library of the Republic of Karelia has also served as a key partner in the Russian library sector. From the summer of 2012, this interactive cooperation network has been strengthened in the sub-project (*Digitization project of kindred languages, pilot phase 2012–2013*) of the Kone Foundation Language Programme, in the IFLA 2012 conference, and in particular in the networking event organised by the Kone Foundation and the Institute for Russia and Eastern Europe in conjunction with the Kindred Languages Programme.

In these contexts the National Library of Finland has begun independent networking and cooperation planning with the different national libraries. Cooperation with the National Library of Russia has been easy and fruitful with the ongoing pilot phase *Digitization project of kindred languages*, but future needs related to digitised materials have also been discussed with other organisations. The development of cooperation projects has gone the furthest with the National Library of the Udmurt Republic, with whom the National Library of Finland has planned the digitization and online publication of old Finno-Ugrian newspaper materials. The construction of a cooperation network has not been restricted to Russian libraries, as the National Library of Finland wishes to enter into negotiations regarding cooperation with the Göttingen State and University Library (*Niedersächsisches Staats- und Universitätsbibliothek*) too.

This cooperation is coordinated by a project manager employed by the National Library of Finland, who reports on his work directly to the leadership of Research Library Services at the National Library. The project manager is responsible for the general coordination, planning and budgeting of the first stage of the project, as well as the preparation of agreements with project participants and other organisations, and the creation of cooperation procedures between Finnish and international organisations.

#### **THE MATERIALS, THEIR QUALITY SPECIFICATIONS AND COPYRIGHTS**

The special grant from the Ministry of Education and Culture 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.

Material linked to the information system will not be provided with new metadata. The intention is for all project participants to provide the bibliographic information of the material to be digitised into the publication system in the form that they have been entered into the local databases. When the material is transferred to the publication system, its bibliographic metadata will be converted and enriched to correspond to the conventions of digital materials. After the metadata have been harmonised, the

material may be used through the user interfaces of third parties, such as Europeana, the European portal of digital culture and art, and Finna, the National Digital Library user interface.

One of the objectives of the project relates to the process and quality of digitization conducted in the Russian Federation. Through cooperation, jointly defined quality criteria could be set for the digitization of Finno-Ugrian materials, since a harmonised, standardised digitization result and metadata also enable the long-term conservation of the material which guarantees that it will be available for future generations. The National Library of Finland is an active partner in the National Digital Library's project for planning and implementing long-term conservation, and the intention is to include Finno-Ugrian digital materials in the conservation.

A large amount of printed material in the Finno-Ugrian languages falls under Russian copyright legislation, and is primarily under the same use restrictions as materials published in Finland. In the first stage of the project, only materials which are openly available and free of copyright will be linked to the information system. The participants are expected to commit to only linking digitised material which has been proven to be free of copyright or for which permission from the copyright holders has been acquired to share the material online.

If some of the material to be digitised is still within copyright, the conditions for digitising and publishing the material must be further examined together with the participants and the National Library Resource. The National Library of Finland has previously settled copyright issues relating to Finno-Ugrian materials with the National Library Resource.

#### **SYSTEMS, EQUIPMENT AND INTERFACES WITH OTHER SYSTEMS**

The careful planning of the information system infrastructure to be conducted in the *Digitization project of kindred languages* is the main priority both in terms of the project objectives and the end users. At the moment, material digitised in the Russian Federation is spread out over several separate systems, the use of which requires mastering different user interfaces. From the perspective of systems infrastructure planning it is apparent that once the materials produced by different organisations can be accessed through one environment, each individual document becomes part of a larger whole. Centralised information system services, cooperation and partnerships between different organisations and the harmonisation of the metadata of digitised material promotes the public and research use of the materials.

Digitised material and material submitted by cooperating organisations will be made available through the publication system maintained by the National Library of Finland, based on the open source VuFind software. The publication system environment also promotes the accessibility and use of the digitised material internationally, among both the international academic community and speakers of Finno-Ugrian languages. The VuFind user interface will also be available in Russian. Additional language versions may be published in cooperation with the national libraries. To enhance communications and publicity, an easily recognisable visual style will be designed for the publication system. In this way the publication system complies with the target of advancing a culture of openness and interaction.

Publishing the material through the VuFind software 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 equivalent international search services, such as Europeana. The materials will also be searchable through Google and other common search engines.

Collecting the material into the VuFind publication system would also facilitate the harmonisation of different metadata practices. The bibliographic data entries cannot be utilised as such in the publication system, because Russian libraries utilise their own metadata format, RusMarc, while VuFind uses DublinCore, a metadata format specifically developed for the description of digital publications which is directly compatible with Europeana. In conjunction with the collection of materials, the

bibliographic metadata will be converted to correspond to the DublinCore metadata specifications for digital materials and the international requirements regarding the harmonisation of metadata.

### **COMMUNICATIONS**

During the first project period, the project manager is in charge of communications in cooperation with the National Library communications unit. The project will have its own page on the National Library of Finland website. The page will include information and news about the project in Finnish, Swedish, English, Russian and other languages if possible. The project can also utilise other communication channels of the National Library of Finland, such as client magazines and social media, enabling communications with specific user groups, particularly to those Finnish public and third sector organisations which are involved in cooperation with the Finno-Ugrian peoples and support the development of their languages. The project manager is in charge of the communication content and of making sure the information is up to date.

Communications aimed at the local communities, i.e., the native language speakers, are mainly the responsibility of the National Libraries of Karelia and the Udmurt Republic, but the National Libraries of Finland and Russia will assist these organisations if this is considered necessary among the partners.

### **SCHEDULE AND FUNDING OF THE SUB-PROJECT**

The Ministry of Education and Culture awarded a special grant in March 2013 to begin the coordinating work for the sub-project. The project manager hired for the first project period began the work relating to the communication between the project partners and general planning in the beginning of March 2013. The work related to agreements, copyrights, metadata management and information systems (identifiers, permanent links, metadata formats) will be conducted in May-September 2013. At the same time, the material will be prepared for compatibility with both Europeana and the National Digital Library. The testing on the material to be linked to the information system should begin in October 2013, and the information system and its features should be ready for presentation to other organisations by the end of the year. The intention is to cover most of the costs of the project period proper in 2014–2016 by external funding.