



## **Implementation of the Data Seal of Approval**

The Data Seal of Approval board hereby confirms that the Trusted Digital repository The Language Bank of Finland complies with the guidelines version 2014-2017 set by the Data Seal of Approval Board.

The afore-mentioned repository has therefore acquired the Data Seal of Approval of 2013 on May 1, 2015.

The Trusted Digital repository is allowed to place an image of the Data Seal of Approval logo corresponding to the guidelines version date on their website. This image must link to this file which is hosted on the Data Seal of Approval website.

Yours sincerely,

The Data Seal of Approval Board

## Assessment Information

Guidelines Version: 2014-2017 | July 19, 2013  
Guidelines Information Booklet: [DSA-booklet\\_2014-2017.pdf](#)  
All Guidelines Documentation: [Documentation](#)

Repository: The Language Bank of Finland  
Seal Acquiry Date: May. 01, 2015

For the latest version of the awarded DSA for this repository please visit our website: <http://assessment.datasealofapproval.org/seals/>

Previously Acquired Seals: Seal date: May 1, 2015  
Guidelines version: 2014-2017 | July 19, 2013

This repository is owned by: **CSC – IT Center for Science**

Finland

T +358 9 4572001  
F +358 9 4572302  
E [contact@csc.fi](mailto:contact@csc.fi)  
W <https://www.csc.fi/>

# Assessment

## 0. Repository Context

### Applicant Entry

*Self-assessment statement:*

The Language Bank of Finland is a language research service at CSC – IT Center for Science. CSC is a non-profit, state-owned company administered by the Ministry of Education and Culture. CSC maintains and develops the state-owned centralised IT infrastructure and uses it to provide nationwide IT services for research, libraries, archives, museums and culture as well as information, education and research management. CSC is responsible for the Language Bank's infrastructure, maintenance and security.

CSC is applying for the Data Seal of Approval for three of the Language Bank's services:

- Language Bank Rights (lbr.csc.fi)
- Language Archive Tools (lat.csc.fi)
- Korp (korp.csc.fi)

Language Bank Rights (LBR) is the Language Bank's electronic user application system. It is based on technology developed at CSC. LBR allows users to apply for corpus and software access rights in the Language Bank's various interfaces (including LAT and Korp) as well as on CSC's software servers. It also provides the Language Bank's content providers with means for monitoring and controlling their resources.

Language Archive Tools (LAT) is an audiovisual corpus interface developed at Max Planck Institute in Nijmegen, Netherlands. LAT contains a selection of audio and video corpora with different levels of access varying from completely public resources to corpora for which each user requires personal approval from the content provider.

Korp is a concordance search tool developed at the Swedish Language Bank in Gothenburg, Sweden. It is used for depositing and studying text corpora. Korp also features different levels of access for its content.

In developing the Language Bank, CSC collaborates with the national FIN-CLARIN consortium. The other members of the consortium are University of Helsinki, Institute for the Languages of Finland, University of Eastern Finland, University of Oulu, University of Tampere, University of Jyväskylä, University of Turku, Aalto University, University of Vaasa, and Åbo Akademi University. FIN-CLARIN collaborates with and aims to become a member of the Europe-wide CLARIN ERIC consortium.

Most of the Language Bank's documentation is produced and maintained in collaboration with CSC's partners in the FIN-CLARIN consortium.

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

The Language Bank of Finland

<http://www.kielipankki.fi>

The Language Bank portal, the principal showcase website of the Finnish language research community. Presently under a thorough reconstruction; the new portal will be launched in 2015.

CSC – IT Center for Science:

<http://www.csc.fi>

The website of the CSC, the Finnish supercomputing center, the organization applying for DSA.

Language Bank Rights:

<https://lbr.csc.fi>

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

CSC's electronic user application system.

FIN-CLARIN:

<http://www.helsinki.fi/finclarin>

The national consortium collaborating with CSC and applying for membership in CLARIN ERIC.

CLARIN ERIC:

<http://clarin.eu>

International language research service federation.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

**1. The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data, and compliance with disciplinary and ethical norms.**

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

**Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

FIN-CLARIN, the national consortium coordinating the Language Bank of Finland, employs a corpus production line framework for preparing, depositing and enriching language resources. FIN-CLARIN's experts provide depositors with support and consultation during the different phases of a language resource's lifespan, including:

- preparing for collecting language data
  
- writing contracts with informants so that the resulting corpora will be as accessible as possible, and other legal advice
  
- tools for producing, maintaining and developing corpora
  
- file formats
  
- metadata

- collaboration between national and international bodies, projects and initiatives

Instructions for producing language resources:

<http://urn.fi/urn:nbn:fi:lb-201412021>

Step-by-step instructions for researchers, including how to properly collect language material for the corpus, how to make sure there will be no legal issues during or after the production process and that the resulting material will be in correct formats. These instructions aim at assuring the Language Bank's contents' high quality and usability.

Overview of the corpus production line:

<http://urn.fi/urn:nbn:fi:lb-201412022>

A visual representation of the life cycle of a corpus in the Language Bank of Finland. Produced by the University of Helsinki for instructing existing and future users and content providers. The corpora produced according to the production line are deposited in the Language Bank's repositories, including LAT and Korp.



Instructions for annotating corpora:

<http://urn.fi/urn:nbn:fi:lb-201412023>

How to properly annotate text, audio and multimedia corpora. Information about relevant formats and software. These instructions aim at assuring the Language Bank's contents' high quality and usability.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 2. The data producer provides the data in formats recommended by the data repository.

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The supported and recommended formats vary from service to service. The Language Bank of Finland provides general instructions as well as platform-specific detailed ones. Besides the recommended formats, versions of the resources in other formats may additionally also be deposited, if considered advantageous. Suboptimal formats may also be accepted in case of particularly endangered material. When needed, files can also be converted.

Instructions for selecting formats for content providers:

<http://urn.fi/urn:nbn:fi:lb-201412024>

What are the preferred file formats in the Language Bank. The requirements are service-specific. These instructions aim at assuring the files deposited in the Language Bank are of the highest possible quality and usability.

More detailed instructions about multimedia files:

<http://urn.fi/urn:nbn:fi:lb-201412025>

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

How to convert audio and video files into the preferred formats. These instructions aim at assuring the files deposited in the Language Bank are of the highest possible quality and usability.

More detailed instructions concerning the Korp corpus interface:

<http://urn.fi/urn:nbn:fi:lb-201412026>

What are the format requirements for the Korp interface in more detail. These instructions aim at assuring the files deposited in Korp are of the highest possible quality and usability.

Instructions for transferring language resources to the Language Bank of Finland:

<http://urn.fi/urn:nbn:fi:lb-201412027>

How to deliver a language resource to the Language Bank. These instructions assist the content providers in publishing their data.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

Accept

*Comments:*

### **3. The data producer provides the data together with the metadata requested by the data repository.**

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

#### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The metadata of the language resources in the Language Bank of Finland are stored in the META-SHARE facility. META-SHARE is an open, integrated, secure and interoperable sharing and exchange facility devoted to the sustainable sharing and dissemination of language resources and designed as a network of distributed repositories.

META-SHARE is developed by the META-NET consortium. University of Helsinki, CSC – IT Center for Science's partner in the Finnish FIN-CLARIN consortium, is a member of META-NET. There are presently (2015) 29 META-SHARE nodes deployed worldwide, including the one at CSC.

FIN-CLARIN takes care of entering the metadata of new resources into META-SHARE as well as maintaining them up to date. Content providers can also create META-SHARE accounts and edit the metadata themselves.

The Language Bank of Finland's META-SHARE node:

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

<http://metashare.csc.fi>

The national META-SHARE node deployed in the Language Bank of Finland, serving as the central metadata repository of the Language Bank. CSC is responsible for the Finnish META-SHARE node's operation. The metadata is entered, curated and maintained in collaboration with FIN-CLARIN.

META-SHARE user manual:

<http://urn.fi/urn:nbn:fi:lb-201412028>

How to use the current version of META-SHARE.

Instructions for creating and providing metadata:

<http://urn.fi/urn:nbn:fi:lb-201412029>

How to compile and present metadata for language resources deposited in the Language Bank and what kind of assistance is available for content providers.

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

CLARIN metadata definitions and instructions:

<http://urn.fi/urn:nbn:fi:lb-2014120210>

Information about the metadata infrastructure of the international CLARIN consortium. This is relevant because the Language Bank wants to be compatible with other centers within the consortium.

Text corpus XML annotation specification:

<http://urn.fi/urn:nbn:fi:lb-2014120211>

How to present metadata in XML files deposited in the Language Bank and instructions for properly encoding the corpora.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

#### **4. The data repository has an explicit mission in the area of digital archiving and promulgates it.**

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

#### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

In addition to the vision, mission and strategy of CSC – IT Center for Science that form the basis of the Language Bank of Finland's operation, the Language Bank adheres to the collaborating FIN-CLARIN consortium's strategy.

CSC – IT Center for Science's strategy:

<http://urn.fi/urn:nbn:fi:lb-2014120213>

CSC's strategy, consisting of mission, vision and values.

FIN-CLARIN's strategy:

<http://urn.fi/urn:nbn:fi:lb-2014120212>



Strategy, mission and vision of the FIN-CLARIN consortium.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

We would to see a more explicit mission statement for digital archiving under the Language Bank of Finland at the time of the next DSA renewal.

## **5. The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.**

### *Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

## **Applicant Entry**

### *Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### *Self-assessment statement:*

The Language Bank of Finland is not a legal entity; it is a service at CSC – IT Center for Science. The national FIN-CLARIN consortium coordinating the depositing of language resources is juridically represented by the University of Helsinki. Each deposited language resource is covered by a deposit agreement between FIN-CLARIN and the content provider. Each user is required to accept general terms and conditions as well as resource-specific licenses.

A user account can be terminated or suspended by CSC without notice in the event of any unauthorized use of the services or if CSC has a justified reason to suspect that the services are used contrary to the terms and conditions.

Information about agreements and licenses:

<http://urn.fi/urn:nbn:fi:lb-2014120215>

What kinds of license classes and settings are available and how to choose the correct one. The main categories are public (PUB), academic (ACA), and restricted (RES). These instructions assist the content providers in preparing their language resources for publication.

CLARIN legal information:

<http://urn.fi/urn:nbn:fi:lb-2014120216>

Definitions of licenses in the international CLARIN consortium. When possible, corpora in the Language Bank are given licenses that are compatible with CLARIN in order to guarantee maximal interoperability.

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 6. The data repository applies documented processes and procedures for managing data storage.

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

Each of the Language Bank of Finland's public services has a business continuity and disaster recovery plan. The plans are updated annually and approved by CSC's Head of Security. The documents comply with CSC's internal and external quality and security requirements. These services include:

Language Archive Tools (LAT)

- audiovisual corpus interface

- <https://lat.csc.fi>

- 

Korp

Data Seal of Approval Board

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

- concordance search service for text corpora

- <https://korp.csc.fi>

- 

#### META-SHARE

- metadata repository

- <https://metashare.csc.fi>

- 

#### metalb

- OAI-PMH service

- [metalb.csc.fi](http://metalb.csc.fi)

- 

#### Persistent identifiers (PID)

- persistent identifier system
- uses the Universal Resource Name (URN) technology
- pid.csc.fi

- 

#### KitWiki

- wiki platform
- <https://kitwiki.csc.fi>

- 

Each service has a Disaster Recovery and Business Continuity Plan. The systems are backed up daily, and backups are retained for three months. The backup process is internally documented.

The Language Bank has a policy for dealing with different versions and instances of the deposited language resources.

CSC is certified according to the standard ISO/IEC 27001:2013. The certification covers operations, development and management of CSC's ICT-platforms. The following functions related to the Language Bank are also covered:

- Datacenters
  
- Networks

- Virtualization platform
- Storage and backup
- Operations systems
- Information security and physical security
- Change, incident and capacity management

CSC's ISO/IEC 27001:2013 certification does not cover application level functions of the Language Bank.

The following OAIS functions are provided:



Responsibility of CSC:

- Access
- Archival Storage
- Administration
- Data Management

•

Responsibility of the user community:

- Ingest

•

Overview of the CSC data environment:

<http://urn.fi/urn:nbn:fi:lb-2014120217>

Information about the data systems at CSC. This includes the data environment of the Language Bank.

Overview of the CSC computing environment:

<http://urn.fi/urn:nbn:fi:lb-2014120218>

Information about the computing resources at CSC. This includes the computing environment of the Language Bank.

The Language Bank's software and tools:

<http://urn.fi/urn:nbn:fi:lb-2014120219>

A list of software deployed on the Language Bank's software servers and virtual machines at CSC.

Language resource versioning process:

<http://urn.fi/urn:nbn:fi:lb-2014120220>

Instructions how to manage versions of language resources in the Language Bank. Instructions for deciding whether a change in a file requires creating a new version or not.

## Reviewer Entry

*Accept or send back to applicant for modification:*

Send back to applicant for modification

*Comments:*

metalb.csc.fi and

- pid.csc.fi

resolve to an APACHE server test page.

Is this a problem?

## **7. The data repository has a plan for long-term preservation of its digital assets.**

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

### **Applicant Entry**

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

In order to mitigate the effect of file format obsolescence, depositing data in widely used, raw and open formats is encouraged. Keeping several formats in parallel may also be justified. The recommendations favor foreseeably longevous formats. Formats are monitored and requirements and recommendations regularly updated. Data in obsolete formats will be migrated into more current ones. Up-to-date metadata also contributes to retaining data usability.

CSC – IT Center for Science manages long-term preservation for the National Digital Library (NDL), a project that aims to ensure that electronic materials of Finnish culture and science are managed with a high standard and easily accessed and securely preserved well into the future. The infrastructure and knowledge accumulated in the project is planned to be extended to cover the Language Bank of Finland as well. The schedule for this project has not been set yet. NDL also has a detailed long-term preservation plan.

CSC participates in the Open Science and Research Initiative that promotes research information availability and open science. One of this project's goals is long-term preservation of research material.

National Digital Library:

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

<http://kdk.fi/en>

The National Digital Library, a project in which CSC is managing long-term preservation.

Digital preservation implementation plan of the National Digital Library:

<http://urn.fi/urn:nbn:fi:lb-2014120221>

This document and the knowledge and expertise accumulated in its creation will be used as an important guideline in the creation of the Language Bank's long-term preservation plan.

The Open Science and Research Initiative:

<http://openscience.fi>

Another project that will be looked closely into in the creation of the Language Bank's long-term preservation plan.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

This guidelines has been accepted. It would not be possible to renew without setting a clear timescale for long term digital preservation.

## 8. Archiving takes place according to explicit work flows across the data life cycle.

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

### Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

The data life cycle in the Language Bank of Finland is described by FIN-CLARIN's corpus production line procedure. Phases of the life cycle:

collecting linguistic material

- obtaining permissions from informants and IPR

- obtaining a research permit

- 

finding a suitable deployment platform for the resource

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)



- depends on data format, content and purpose

- 

depositing the resource

- signing an agreement between the content provider and the repository
- creating and publishing metadata
- assigning at least one persistent identifier for the resource
- choosing an accessibility level and a license, if applicable
- including the the new resource in the access rights application system

- 

publishing the resource

- granting users access, if applicable
- using the resource in research and education

- 

enriching the resource with

- additional depth of annotation

- publishing new versions

- 

Overview of the corpus production line:

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

<http://urn.fi/urn:nbn:fi:lb-201412022>

A visual representation of the life cycle of a corpus in the Language Bank of Finland. Produced by the University of Helsinki for instructing existing and future users and content providers. The corpora produced according to the production line are deposited in the Language Bank's repositories, including LAT and Korp.

Ministry of Culture and Education's guidelines for open science:

<http://urn.fi/urn:nbn:fi:lb-201412022>

The Language Bank follows the resource life cycle guidelines of the Ministry of Culture and Education.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

More detailed explanations for non-English language evidence would be preferable

## **9. The data repository assumes responsibility from the data producers for access and availability of the digital objects.**

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

Content providers are provided with instructions and guidance beginning with the process of compiling corpora. One of the aims of FIN-CLARIN's involvement in the process as early as possible is to ensure that the resulting language resources be maximally usable, for example to prevent agreements made with informants from excessively restricting the resource's availability.

For each resource, a deposit agreement is written between FIN-CLARIN and the content provider. FIN-CLARIN provides templates for the agreements.

Corpus management instructions:

<http://urn.fi/urn:nbn:fi:lb-2014120223>

How to write a corpus management plan and which questions the plan should answer. Having a corpus management plan is beneficial for both the content provider and the users. It also ensures the possibility of further development for the corpus. A plan is typically required by research funding bodies.

Instructions for creating corpus agreements for content providers:

<http://urn.fi/urn:nbn:fi:lb-2014120224>

How to create agreements concerning intellectual property and privacy protection. Proper agreements allow maximal openness for the material.

Informant agreement templates:

<http://urn.fi/urn:nbn:fi:lb-2014120225>

Templates for agreements between the content provider and the informant.

Deposit agreement templates:

<http://urn.fi/urn:nbn:fi:lb-2014120226>

Templates for deposit agreements for each of the main access categories of language resources (PUB, ACA and RES).

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## **10. The data repository enables the users to discover and use the data and refer to them in a persistent way.**

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

Each corpus in the Language Bank of Finland has a Universal Resource Name (URN). The URN system is provided by the National Library of Finland who has allocated the Language Bank its own namespace (urn:nbn:fi:lb). The URNs are predominantly created manually and stored on the server pid.csc.fi alongside their referent URLs. Based on the list, a Python script is used to generate an XML file that conforms to the National Library's specifications. The National Library harvests the contents of the XML file daily.

The Language Bank is also prepared for interoperability with the Handle PID system. A round trip conversion principle has been devised between the two systems, so that either form of PID can always be derived from the other.

CSC – IT Center for Science, the organization hosting the Language Bank, is a member of the European Persistent Identifier Consortium (EPIC).

The PIDs allocated by CSC:

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

<http://urn.fi/urn:nbn:fi:lb-2014120227>

The XML file containing each URN PID allocated by CSC for the Language Bank of Finland.

The National Library of Finland's URN information:

<http://urn.fi/urn:nbn:fi:lb-2014120228>

What is a uniform resource name (URN). Instructions provided by the National Library of Finland, the provider of the URN service used at the Language Bank.

The Handle system:

<http://www.handle.net>

The website of the commercial PID system used in the European Persistent Identifier Consortium (EPIC) with which the Language Bank's URN system has been made compatible.

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)



European Persistent Identifier Consortium:

<http://www.pidconsortium.eu>

The website of the PID consortium CSC is a member of.

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 11. The data repository ensures the integrity of the digital objects and the metadata.

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

### Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

The Language Bank's main services are the corpus interfaces Language Archive Tools (LAT, lat.csc.fi) and Korp (korp.csc.fi), and the application server Taito (taito.csc.fi). None of them have consistent integrity checking at the moment. Checking mechanisms based on md5checksums are being planned. LAT supports MD5 checksums natively; Korp and Taito both rely on directory structures where subdirectories can be "frozen" and their content regularly checked against the frozen state and changes reported.

Taito user guide:

<http://urn.fi/urn:nbn:fi:lb-201503131>

General instructions for using the Taito super cluster at CSC.

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 12. The data repository ensures the authenticity of the digital objects and the metadata.

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

FIN-CLARIN ensures that publicly available content is accompanied by the appropriate licenses and agreements. By signing the agreement, the content provider asserts the resource's authenticity. FIN-CLARIN's experts check the metadata provided by the content provider. Content providers are regularly requested to confirm that the metadata remains up to date and to update it when needed.

Instructions for creating language resources:

<http://urn.fi/urn:nbn:fi:lb-2014120229>

Comprehensive instructions and a checklist for content providers for creating language resources.

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

### **13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.**

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

#### **Applicant Entry**

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

The digital preservation system of the National Digital Library (NDL), which CSC – The IT Center for Science is implementing, is based on the OAIS reference model. The NDL principles are planned to be extended to cover the Language Bank of Finland as well.

National Digital Library:

<http://kdk.fi>

The National Digital Library, a project in which CSC is managing long-term preservation.

Digital preservation implementation plan of the National Digital Library:

<http://urn.fi/urn:nbn:fi:lb-2014120221>

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

This document and the knowledge and expertise accumulated in its creation will be used as an important guideline in the creation of the Language Bank's long-term preservation plan.

CSC security information:

<http://urn.fi/urn:nbn:fi:lb-2014120239>

Description of CSC's security policy and best practices.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

See previous comments on timescales

## 14. The data consumer complies with access regulations set by the data repository.

### *Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

## **Applicant Entry**

### *Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### *Self-assessment statement:*

Permissions to access the language resources in the Language Bank of Finland are applied for, processed, monitored and operated in the Language Bank Rights application (LBR). It is based on the Resource Entitlement Management System (REMS) technology developed at CSC – IT Center for Science.

A user of the Language Bank can be both a content provider and a user of other resources. The LBR system uses META-SHARE directly to provide information about the available resources. Other systems and platforms consult LBR for access information and act accordingly. Users are authenticated either by their Haka identity federation (or an international federation affiliated therewith) credentials or a local CSC user account.

The application process in LBR introduces the applicant to the licenses associated with access the selected resources and requires them to be accepted in order to proceed with the application. The licenses include CSC's general terms of service and resource-specific conditions, if applicable.

Language Bank Rights:

<https://lbr.csc.fi>

CSC's electronic user application system.

Resource Entitlement Management System:

<http://urn.fi/urn:nbn:fi:lb-2014120230>

Description of the Resource Entitlement Management System technology (REMS) which the Language Bank Rights application (LBR) is based on. Developed at CSC.

Instructions for applying rights to access language resources in different services:

<http://urn.fi/urn:nbn:fi:lb-2014120231>

How to apply for access to the Language Bank's language resources and what kind of services are included in the access rights.

## **Reviewer Entry**

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

*Accept or send back to applicant for modification:*

Accept

*Comments:*



**15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information.**

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

**Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

Prior to accessing any language resources in the Language Bank of Finland, each user is required to accept all terms and conditions and resource-specific licenses. The users agree:

- Not to share their credentials, leave them for others to see, or neglect any security responsibilities defined in the service description.
  
- Not to misuse or abuse any CSC or third party service or property, including intellectual property.
  
- Not to misuse or abuse user content, credentials or other confidential information.
  
- Not to send or transmit harassing, abusive, libellous, obscene or unsolicited communications.
  
- Not to tamper with or deliberately disrupt system resources or network traffic to the services.

- To notify CSC promptly if their account has been used without permission or if their credentials have been lost or stolen.

A user account can be terminated or suspended by CSC without notice in the event of any unauthorized use of the services or if CSC has a justified reason to suspect that the services are used contrary to the terms and conditions.

CSC's general terms of use:

<http://urn.fi/urn:nbn:fi:lb-2014120232>

The general terms of use of CSC's services.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

## **16. The data consumer respects the applicable licences of the data repository regarding the use of the data.**

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

In addition to the general terms and conditions, many resources are associated with specific licenses. Access rights are only granted to applicants who have accepted all applicable terms and licenses.

The language resources deposited in the Language Bank have a variety of licenses. They can be classified into three main groups: public (PUB), academic (ACA) and restricted (RES). Public resources can be accessed freely. Academic resources require the applicant to possess an academic status, such as researcher. Restricted resources require a personal permission granted by the content provider or their delegated contact person, who, in many cases, are the administrators of the Language Bank.

FIN-CLARIN experts aid content providers in choosing appropriate licenses for their language resources. There is also an online tool for assisting the process.

CLARIN license categories:

<http://urn.fi/urn:nbn:fi:lb-2014120233>

What kind of categories and traits are included in the different licenses that can be applied to CLARIN-compatible language resources in the Language Bank.

General public (PUB) license:

<http://urn.fi/urn:nbn:fi:lb-2014120234>

General academic (ACA) license:

<http://urn.fi/urn:nbn:fi:lb-2014120235>

General restricted (RES) license:

<http://urn.fi/urn:nbn:fi:lb-2014120236>

License selection helper tool:

<http://urn.fi/urn:nbn:fi:lb-2014120237>

An application for identifying what kind of a license is suitable for a given language resource.

Graphical representation of license types:

<http://urn.fi/urn:nbn:fi:lb-2014120238>

A chart for identifying what kind of a license is suitable for a given language resource.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*