

# CLARIN AGREEMENT



2016-02-08



---

## CLARIN AGREEMENT

### CLARIN Agreement for Finland

#### *Period*

November 2015 – February 2017

CLARIN ERIC, represented by its Legal Representative Franciska de Jong, Executive Director,

of the one part

and University of Helsinki, representing the National FIN-CLARIN Consortium, cf. Annex 1-A,

of the other part

HAVE AGREED on the following specific terms and conditions for membership of CLARIN ERIC.

#### Article 1 – Scope

The two parties have agreed to conclude an agreement in accordance with the CLARIN ERIC statutes Art. 6.5, including the specific terms and conditions described in this CLARIN Agreement which relate to articles 2, 6.1, 6.2 and 8.2 of the CLARIN ERIC statutes.

#### Article 2 – Rights

The CLARIN ERIC statutes describe the rights of members, and they are not repeated here.

#### Article 3 - Obligations

##### 3.1 Contribution to CLARIN ERIC overall and specific goals and obligations listed in articles 2 and 6 of the statutes

The National Consortium will contribute to the construction and operation of CLARIN ERIC by means of its financial contribution, by means of acting in accordance with the statutes, and through the specific national contributions listed in Article 3.5 in this CLARIN Agreement. These include but are not limited to the variety of activities listed in article 2.2 and obligations listed in article 6.2 of the statutes.

In addition to the specific contributions mentioned below, the National Consortium will promote the adoption of relevant standards in national resource and tools creation projects. It will also promote the uptake of CLARIN services among researchers in their country, and gather user feedback and requirements, and will support CLARIN centres in the member country by facilitating integration into national and other relevant infrastructures.

##### 3.2 Financial Contribution

The annual membership fee is determined in Annex 2 to the statutes.

The annual membership fee must reach CLARIN ERIC no later than 28 February of the same year in order to secure the cash flow. An invoice for collecting the annual membership fee will be issued by CLARIN ERIC in the month of December of the preceding year. For entities joining CLARIN ERIC in the course of a year, the membership fee should reach CLARIN ERIC no later than two months after receipt of the invoice. The invoice will be sent to the address given in Annex 1-A.

### 3.3 National Consortium

A National Consortium or other organisational framework for coordinating and implementing the obligations following from this Agreement has been set up at the national level.

The National Consortium is described in Annex 1-A.

### 3.4 National Coordinator

A national coordinator responsible for the national consortium or entity has been appointed. Name, affiliation and contact details of the appointed National Coordinator are provided in Annex 1-A. All correspondence and contact between CLARIN ERIC and the National Consortium will be channelled through the National Coordinator.

### 3.5 Specific National Contributions

This section describes the national level in-kind contribution to CLARIN; part of this contribution will be coordinated at the transnational level through CLARIN ERIC, and part will be coordinated at the national level. This agreement lists the components of the offering at the time of signature. For each component, details are provided in technical annexes where relevant. Any changes (additions, deletions, changes of technical specifications) will be reflected by updates of the annexes.

#### 3.5.1. Contributions to be coordinated by CLARIN ERIC

(a) Data and service centres: In accordance with Art. 6.2 (f), at least one centre of type A or B should be contributed (cf. the Technical and Scientific Description, Annex CLARIN Centre Types; see also 3.6 below). Detailed descriptions of the centres and the services they offer are included in Annex 1-B.

(b) Access: The National Consortium will comply with statutes article 19 on Access Policies for Users. Any restrictions or exceptions are specified Annex 1-F, if applicable.

(c) User Authentication and Authorization System: Specification is given in Annex 1-C.

(d) Knowledge sharing infrastructure (KSI): The National Consortium will contribute to the KSI by providing access to its knowledge and expertise to the CLARIN community at large. See Annex 1-D.

(e) Membership of CLARIN ERIC committees and any other contribution being offered to CLARIN and lending itself for coordination by CLARIN ERIC at the European level. Detailed descriptions are included in Annex 1-E.

#### 3.5.2. Contributions coordinated at the national level, but compliant with CLARIN

(a) Enhancements and improvements to existing resources, tools and services. Detailed descriptions are included in Annex 2-A.

(b) Creation of new resources, tools and services. Detailed descriptions are included in Annex 2-B.

(c) Measures to promote the adoption of and adherence to CLARIN standards in national resources and tools creation programmes and projects. Detailed descriptions are included in Annex 2-C.

(d) Measures to create a knowledge-sharing infrastructure at the national level, with a view to promoting uptake of CLARIN services by the target audience, to exchange knowledge and expertise, and to advance Humanities and Social Sciences research by using CLARIN. A brief description of those items that are not already covered by 3.5.1-(d) is provided in Annex 2-D.

(e) Any other contribution relevant to the goals of CLARIN, including collaboration with third parties that CLARIN as a whole could contribute to or from which CLARIN could benefit. Detailed descriptions are included in Annex 2-E (if applicable).

### 3.6. National contributions made jointly with other members

According to the statutes, Article 6.3, members may decide to provide some of the national contributions in cooperation with other members, observers or third parties. In such cases, a separate multi-party agreement is made between the CLARIN ERIC and those members / observers / third parties concerned.

Signatures

Date:

Date:

Jukka Kola  
Rector, University of Helsinki

Franciska de Jong  
Executive Director, CLARIN ERIC

Date:

Esa Hämäläinen  
Director of Administration,  
University of Helsinki

Date:

Krister Lindén  
National Coordinator, FIN-CLARIN

## Annex 1 National contributions to be coordinated by CLARIN ERIC

### Annex 1-A

The National Consortium consists of the following partner(s):

1. The University of Helsinki (UHEL) offers expertise, e.g., in the following areas: European and world languages, corpus linguistics, Variation of English (VARIENG), language universals, language technology including morphological, surface-syntactic parsing and finite-state transducer technology (HFST), speech synthesis as well as language data including synonym and terminology databases, and annotated corpora of Finnish.
2. CSC–IT Center for Science Ltd. offers services for language research including an extensive number of text and speech data collections as well as software suitable for research in an efficient computing environment. CSC offers its expertise, e.g., in the following areas:
  - Evaluation, installation, tailoring, development and maintenance of linguistic software and tools.
  - Web-based collaboration environments for research communities as well as web interfaces for linguistic tools, e.g. as part of the Scientist’s User Interface, [korp.csc.fi](http://korp.csc.fi) and [lat.csc.fi](http://lat.csc.fi).
  - For the European CLARIN infrastructure and as needed for implementing functionalities for international use, e.g., technical definitions, PID compatibility for resources, authorization service extensions, installation and operation of new resources and their required computing and archiving servers, agreements concerning international resource use, research tool compatibility and integration on an international level.
3. The Institute for the Languages of Finland (KOTUS) offers expertise, e.g., in the following areas: internet access to scientific dictionaries, other dictionaries, lexicons and wordlists, resources for linguistic and onomasilological maintenance, text, speech and video corpora as well as other audio-visual digital data collections. Resources are available in particular for Finnish and Swedish but also for languages related to Finnish.
4. The Aalto University offers expertise, e.g., in the following areas: automatic speech recognition, speaker adaptation for speech recognition and synthesis, alignment of text and speech, indexing, retrieval and browsing of audio and audio-visual material, methods and tools for the analysis and parameterization of the glottal source, vocoders in HMM-based speech synthesis, methods for robust parameterization of speech, as well as resources with simultaneous recordings of speech and EGG data.
5. The University of Eastern Finland offers expertise, e.g., in the following areas: planning, recording and annotation of speech corpora and creation of speech and text corpora for pathological language; multi-lingual resources, in particular for translation and contrastive language studies.
6. The University of Jyväskylä offers expertise, e.g., in the following areas: language skills assessment corpora, multi-lingual learners’ corpora, corpora of English as a teaching language and as a foreign language, as well as corpora of child and student language use.
7. The University of Oulu offers expertise, e.g., in the following areas: Finnic minority and regional languages, Northern varieties of Finnish, learners’ corpora, corpus linguistics, corpus methodology, Saami linguistics, Saami language technology,

training researchers and other professionals who have a profound knowledge of Saami language and culture.

8. The University of Tampere offers expertise, e.g., in the following areas: search tools, text corpora and parallel text corpora.
9. The University of Turku offers expertise, e.g., in the following areas: syntactic and semantic annotation of English bio-medical text as well as a general Finnish Treebank, text retrieval and analysis tools, the dialect corpus of the syntax archive, and the morpho-syntactic database of the works of Michael Agricola, as well as a Finnish parser, health data corpora, bio-molecular interaction resources, search and analysis services, and collections of Satakunta regional data, Finnish language learner data, and old literary Finnish.
10. The University of Vaasa offers expertise, e.g., in the following areas: terminology, especially conceptual analysis, research in language for special purposes as well as discourse analysis.

The CLARIN centres of Finland offer an R-type service (*Recognized Centre*) according to the *CLARIN ERIC Technical and Scientific Description, July 2011*, unless otherwise specified.

The National Coordinator is:

Krister Lindén  
Research Director  
Faculty of Arts  
Department of Modern Languages  
P.O.Box 24, Unioninkatu 40B  
FI-00014 University of Helsinki, Finland  
Phone: +358 2941 29303  
Email: [krister.linden@helsinki.fi](mailto:krister.linden@helsinki.fi)

The invoice for the Annual fee is to be sent to:

*Note! The Academy of Finland only accepts online or e-invoices.*

Academy of Finland EDI identifier: 003702458939  
Opus Capital Group Oy (operator): 003710948874  
Academy of Finland Business ID: FI0245893-9

Contact information:

Merja Särkioja  
Academy of Finland  
Hakaniemenranta 6  
POB 131  
FI-00531 Helsinki  
Phone: +358 295 335 000  
Registrar's Office: +358 295 335 049  
Emails: [merja.sarikoja@aka.fi](mailto:merja.sarikoja@aka.fi), [kirjaamo@aka.fi](mailto:kirjaamo@aka.fi)

## Annex 1-B

The following data and computing centres will be providing services to CLARIN:

1. The University of Helsinki (<http://www.helsinki.fi/fin-clarin>) is responsible for the contents of the FIN-CLARIN Language Bank (<http://www.kielipankki.fi>).
2. CSC - IT centre for science Ltd, (<http://www.csc.fi/>) offers data processing services for the FIN-CLARIN Language Bank, including storage capacity and computing facilities for storing and analyzing language resources, provision of tools that can access language resources, implementing and maintaining a web-based interfaces to language bank tools, user management and authorization infrastructure for the language resources, support of persistent identifiers for digital objects, participation in further development of IT infrastructure for CLARIN in accordance with European CLARIN partners and developments, and maintenance of language bank services.

The following FIN-CLARIN partners are nominated for CLARIN Type A + B centres:

- CSC - IT centre for science Ltd

B centres and certification status:

- CSC - IT centre for science Ltd has been certified as a CLARIN B centre in 2015.

## Annex 1-C

Specification of the implementation of user authentication and authorisation system: All national centres (or the universities they belong to) have joined the AAI federation of the Finnish National Research and Education Network (HAKA).

## Annex 1-D

FIN-CLARIN will (i) publish the relevant Knowledge-Sharing Infrastructure activities and facilities it offers on the CLARIN website according to guidelines provided by CLARIN ERIC, (ii) will keep this list up to date, and (iii) will participate in the coordination of these activities with other National Consortia in order to maximize their efficiency and effectiveness.

## Annex 1-E

FIN-CLARIN will contribute to the permanent CLARIN ERIC committees where applicable, and will participate in ad hoc committees as required and within the limits of the national budget:

- National Coordinators' Forum: Krister Lindén (UHEL)
- Standing Committee for CLARIN Technical Centres: Martin Matthiesen (CSC)
- Legal Issues Committee: Krister Lindén (UHEL, chair)
- Standards Committee: Martin Matthiesen (CSC)

The cost of travel to CLARIN ERIC meetings by consortium members will be covered from the National Consortium budget.

Any other contribution: None.

## Annex 1-F

Resources, tools, and services provided by FIN-CLARIN will comply with statutes article 19 on Access Policies for Users. Any restrictions or exceptions are specified in the Language Bank Rights authorization tool or similar means of getting access to the resources.

## Annex 2 National contributions to be coordinated at the national level

### Annex 2-A

National Work Programme with timing:

The FIN-CLARIN partners provide the following resources, tools, and services through the Language Bank of Finland. Further LRTs may be contributed in addition to this list.

#### SERVICES AND TOOLS

CSC – IT Center for Science Ltd. and University of Helsinki:

- URL: <https://www.csc.fi/-/i-4>  
Name: IDA Storage Service
- URL: <http://lat.csc.fi/>  
Name: LAT (Language Archiving Technology)  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2015050902>
- URL: <http://korp.csc.fi>  
Name: Korp  
Resource type: web-based concordance search service  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2015050901>
- URL: <https://github.com/aalto-speech/AaltoASR>  
Name: Aalto University Automatic Speech Recognition System  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2014091904>
- URN: <http://urn.fi/urn:nbn:fi:lb-2014052713>  
Name: Finnish WordNet  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2014052714>
- Name: WordNet, Kielipankki Version  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2015082502>
- Name: finnish.analyze.sh  
Resource type: script for analyzing Finnish texts  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2015050903>
- Name: omorfi-analyze.sh  
Resource type: script for analyzing Finnish texts  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2015050904>

- Name: Connexor Machine Syntax for Finnish  
Resource type: script for analyzing Finnish texts  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2015022001>
- Name: TextMorfo  
Resource type: tool for parsing and disambiguating Finnish texts  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2014073089>
- Name: Connexor Machine Syntax for Swedish  
Resource type: script for analyzing Swedish texts  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2014073089>
- Name: Connexor Machine Syntax for English  
Resource type: script for analyzing English texts  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2015022003>
- Name: Helsinki Finite-State Transducer Technology  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-20140730183>

## CORPORA

### Aalto University:

- Name: Speech and EGG (Electroglottography) Simultaneous Recordings  
Resource type: audio corpus  
Contact: [paavo.alku@aalto.fi](mailto:paavo.alku@aalto.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-20140730182>

### University of Helsinki:

- Name: Corpus of Finnish Magazines and Newspapers from the 1990s and 2000s  
Resource type: text corpus  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2016011101>
- Name: amph-Corpus  
Resource type: text corpus  
Contact: [antti.arppe@ualberta.ca](mailto:antti.arppe@ualberta.ca)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2015021301>
- Name: The Newspaper and Periodical OCR Corpus of the National Library of Finland (1771-1874)  
Resource type: text corpus  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2015051201>



- Name: Digital Morphology Archives  
 Resource type: text corpus  
 Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
 Metadata: <http://urn.fi/urn:nbn:fi:lb-201403261>
- Name: ELFA Corpus  
 Resource type: audio and text corpus  
 Contact: [anna.mauranen@helsinki.fi](mailto:anna.mauranen@helsinki.fi)  
 Metadata: <http://urn.fi/urn:nbn:fi:lb-201403262>
- Name: Finnish Broadcast Corpus  
 Resource type: text and video corpus  
 Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
 Metadata: <http://urn.fi/urn:nbn:fi:lb-201403265>
- Name: Fenno-ugrica, Kielipankki Version  
 Resource type: text corpus  
 Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
 Metadata: <http://urn.fi/urn:nbn:fi:lb-2015101902>
- Name: Finnish TreeBank 1  
 Resource type: text corpus  
 Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
 Metadata: <http://urn.fi/urn:nbn:fi:lb-20140730138>
- Name: Finnish TreeBank 2  
 Resource type: text corpus  
 Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
 Metadata: <http://urn.fi/urn:nbn:fi:lb-201407163>
- Name: Finnish TreeBank 3  
 Resource type: text corpus  
 Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
 Metadata: <http://urn.fi/urn:nbn:fi:lb-201406022>
- Name: The Finnish N-grams 1820-2000 of the Newspaper and Periodical Corpus of the National Library of Finland  
 Resource type: textngram  
 Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
 Metadata: <http://urn.fi/urn:nbn:fi:lb-2014073038>
- Name: Finnish Text Collection  
 Resource type: text corpus  
 Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
 Metadata: <http://urn.fi/urn:nbn:fi:lb-201403268>
- Name: A Multimodal Corpus of Tourist Brochures Produced by the City of Helsinki, Finland (1967-2008)

- Resource type: text corpus  
 Contact: [tuomo.hiippala@iki.fi](mailto:tuomo.hiippala@iki.fi)  
 Metadata: <http://urn.fi/urn:nbn:fi:lb-201411281>
- Name: The Finnish Gutenberg Corpus  
 Resource type: text corpus  
 Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
 Metadata: <http://urn.fi/urn:nbn:fi:lb-2014100301>
  - Name: Helsinki Corpus of Swahili  
 Resource type: text corpus  
 Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
 Metadata: <http://urn.fi/urn:nbn:fi:lb-2014032624>
  - Name: The Helsinki Korp Europarl Bilingual Corpora  
 Resource type: text corpora  
 Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
 Metadata: <http://urn.fi/urn:nbn:fi:lb-2015043011>
  - Name: The Helsinki Korp JRC-Acquis Bilingual Parallel Corpora  
 Resource type: text corpora  
 Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
 Metadata: <http://urn.fi/urn:nbn:fi:lb-2015061210>
  - Name: The Longitudinal Corpus of Finnish Spoken in Helsinki (1970s, 1990s and 2010s)  
 Resource type: audio and text corpora  
 Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
 Metadata: <http://urn.fi/urn:nbn:fi:lb-2014073041>
  - Name: The HS.fi News and Comments Corpus  
 Resource type: text corpus  
 Contact: [jutta.salminen@helsinki.fi](mailto:jutta.salminen@helsinki.fi)  
 Metadata: <http://urn.fi/urn:nbn:fi:lb-2014052718>
  - Name: Aleksis Kivi Corpus (SKS)  
 Resource type: text corpus  
 Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
 Metadata: <http://urn.fi/urn:nbn:fi:lb-201405274>
  - Name: The Newspaper and Periodical Corpus of the National Library of Finland, Kielipankki Version  
 Resource type: text corpus  
 Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
 Metadata: <http://urn.fi/urn:nbn:fi:lb-201405276>
  - Name: Multilingual Resource Collection of the University of Helsinki Language Corpus Server  
 Resource type: text corpora

- Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-201403269>
- Name: Collection of OTA Texts in Public Use  
Resource type: text corpus  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2015082501>
  - Name: Route to A wing  
Resource type: text, audio and video corpus  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2014101401>
  - Name: SFNET Corpus  
Resource type: text corpus  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-20150126>
  - Name: Finnish Folk Poetry  
Resource type: text corpus  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2014052712>
  - Name: The Swedish N-grams 1770-1940 of the Newspaper and Periodical Corpus of the National Library of Finland  
Resource type: textngram  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2014091902>
  - The Suomi 24 Corpus  
Resource type: text corpus  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-201412171>
  - Name: The Suomi 24 2001-2014 (Sample) Corpus  
Resource type: text corpus  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2015120101>
  - Name: The Suomi 24 Sentences Corpus  
Resource type: text corpus  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2015091701>
  - Name: Frequency Lexicon of the Finnish Newspaper Language  
Resource type: lexical conceptual resource  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-201405272>

- Name: Ylilauta Corpus  
Resource type: text corpus  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2015031802>

University of Jyväskylä:

- Name: Jyväskylä Corpus of Middle French  
Resource type: text corpus  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-201403264>

University of Oulu:

- Name: International Corpus of Learner Finnish (ICLFI)  
Resource type: text corpus  
Contact: [sisko.brunni@oulu.fi](mailto:sisko.brunni@oulu.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-20140730163>
- Name: Oulu Corpus  
Resource type: text corpus  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-201403266>

University of Tampere:

- Name: FiRuLex: Finnish-Russian Comparable Corpus of Legal Texts  
Resource type: text corpus  
Contact: [mikhail.mikhailov@uta.fi](mailto:mikhail.mikhailov@uta.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-201407161>
- Name: MULCOLD - Multilingual Corpus of Legal Documents  
Resource type: text corpus  
Contact: [mikhail.mikhailov@uta.fi](mailto:mikhail.mikhailov@uta.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-201405278>
- Name: ParFin: Finnish-Russian Parallel Corpus of Literary Texts  
Resource type: text corpus  
Contact: [mikhail.mikhailov@uta.fi](mailto:mikhail.mikhailov@uta.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2014052710>

University of Turku:

- Name: The Advanced Finnish Learners' Corpus  
Resource type: text corpus  
Contact: [kirsti.siitonen@uta.fi](mailto:kirsti.siitonen@uta.fi); [ilmari.ivaska@utu.fi](mailto:ilmari.ivaska@utu.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-201407167>

#### The Institute for the Languages of Finland:

- Name: Karelian Dictionary  
Resource type: lexical conceptual resource  
Contact: [leena.joki@kotus.fi](mailto:leena.joki@kotus.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-20140730157>
- KOTUS Finnish-Swedish Parallel Corpus  
Resource type: text corpus  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-201406036>
- Name: The Finnish Dialect Syntax Archive  
Resource type: audio and text corpus  
Contact: [maria.vilkuna@helsinki.fi](mailto:maria.vilkuna@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2014052716>
- Name: Follow-up Study of Dialects of Finnish  
Resource type: audio and text corpus  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-2014073043>
- Name: Samples of Spoken Finnish  
Resource type: audio and text corpus  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-201407141>
- Name: New Year's Speeches of the Presidents of the Republic of Finland  
Resource type: text corpus  
Contact: [finclarin@helsinki.fi](mailto:finclarin@helsinki.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-20140730150>
- Name: Corpus of Old Literary Finnish  
Resource type: text corpus  
Contact: [pirkko.kuutti@kotus.fi](mailto:pirkko.kuutti@kotus.fi)  
Metadata: <http://urn.fi/urn:nbn:fi:lb-201407165>

#### ON-GOING ENHANCEMENT WORK

1. The University of Helsinki will improve and extend Finnish language corpora, such as the National Library Newspaper collection, the social media corpus Suomi24, Finnish bilingual corpora as well as the national scientific term bank.
2. The Institute for the Languages of Finland (KOTUS) will improve their existing resources and services.
3. The Aalto University will improve and develop the speech recognition software as well as the glottis analysis and speech parameterization tools.
4. The University of Eastern Finland will produce a baseline corpus of typical Finnish speech data for use in acoustic assessment protocols for speech pathology.

5. The University of Jyväskylä will improve their corpora of language proficiency assessments.
6. The University of Oulu will improve and maintain the Oulu Tape Archive (ONA), the Saami Culture Archive (SKA), and the International Corpus of Learner Finnish (ICLFI).
7. The University of Tampere will improve and maintain the Russian-Finnish parallel corpus of literary texts (ParRus), Multilingual corpus of juridical texts (MulJur), Comparable Russian-Finnish corpus of juridical texts (FiRuLex), a Resource Collection of human-computer dialogues, and the `mustikka.uta.fi` web server.
8. The University of Turku will improve and maintain English biomedical syntactically and semantically annotated text corpora, a Treebank of general Finnish, text retrieval and analysis tools, the syntax archive dialect corpus and related user guidance, a morphosyntactically annotated corpus of the works of Michael Agricola and its documentation
9. The University of Vaasa will maintain WasaTerm - Terminologia sanastoja, Terminology Forum

## Annex 2-B

National Work Programme with timing:

Creation of new resources, tools and services for CLARIN.

FIN-CLARIN is one of the more than 30 infrastructures to be built according to the Finnish national roadmap of research infrastructures. The FIN-CLARIN plan is to create new language resources and tools for language processing, collect and upgrade existing language resources to make them widely usable, and include new language resources created in other projects as they become available:

1. The University of Helsinki will create a Finnish Treebank and a parse bank as well as parsers for the annotation of the Treebank as well as annotating new data of various genres. Tools for named entity recognition using HFST and tools for semantic annotation will also be created.
2. The Institute for the Languages of Finland (KOTUS) will create general language resource, in particular electronic dictionaries for Finnish and Swedish dialects as well as for old Finnish. Text and audio-visual corpora will also be developed for Finnish and Swedish dialect material as well as corpora for guiding language use and maintenance.
3. The Aalto University will create tools for indexing, searching and browsing audio-visual corpora based on speech recognition and metadata, multimodal corpora for studying movement and language, as well as a vocoder for HMM synthesis.
4. The University of Eastern Finland will conduct Cross-linguistic Studies of Phrasal Idioms: Digitalized Text and Word-list Corpora of English, German, Swedish, Russian, Finnish – a joint UEF project, as well as Finnish control data for *Speech Examination* test suite.
5. The University of Jyväskylä will continue development of the TOPLING test text corpus for second language learners of Swedish and English, and the DIALUKI Finnish-English-Russian data set with language tasks, questionnaires, psycholinguistic tasks, interviews and observation data, as well as the procurement planning and the coordination of the syntax and error annotation of the ICLFI corpus.

6. The University of Oulu will create the Giellagas Corpus of Spoken Saami Language and an error annotated version of the International Corpus of Learner Finnish, ICLFI.
7. The University of Tampere will create a Finnish-Russian corpus of literary texts (ParFin), and the English-Finnish-English parallel corpus (TamBic)
8. The University of Turku will create a parser for Finnish, a corpus of medical data, services for retrieving and analysing biomedical resources, Sapu – a corpus with additional samples of speech data from Satakunta, a corpus with additional samples and annotations of advanced learners of Finnish, as well as an annotated corpus of old Finnish.

### Annex 2-C

Measures to promote the adoption of and adherence to CLARIN standards in national resources and tools creation programmes and projects.

FIN-CLARIN cooperates with the Academy of Finland in including appropriate funding requirements in research project plans that intend to produce language resources. The goal is to have research proposals already include funding for making resources available in standard formats with proper metadata so that the resources can be seamlessly integrated in the FIN-CLARIN archives, preserved and used by other scholars. In addition, instructions and documentation will be developed to help create metadata in CLARIN compatible format:

1. The University of Helsinki will promote CLARIN standards in seminars for the Language Bank of Finland on IPR issues, procuring metadata and annotating data sets, as well as providing support and help desk for users of FIN-CLARIN web applications.
2. CSC–IT Center for Science Ltd. will promote CLARIN standards as follows: CSC assumes an active role in implementing inter-federated AAI. CSC offers expertise in Resource Access Management, PIDs (persistent identifiers), compatible and integrated research tools, as well as service desk support for users of FIN-CLARIN's applications.
3. The Institute for the Languages of Finland (KOTUS) will further streamline and document the process of creating and procuring language resources.
4. The University of Eastern Finland will integrate the CLARIN standards and policies in all its language resource-related courses on BA, MA and PhD levels. Research projects will be required to make their language resources openly available via FIN-CLARIN.
5. The University of Oulu will go over its resources to make sure they are CLARIN conformant. The Giellagas Corpus of Spoken Saami Language and the Oulu Tape Archive ONA will be included in the FIN-CLARIN Language Bank.
6. The University of Tampere will go over its resources to make sure they are CLARIN conformant.
7. The University of Turku will port its language archive dialect corpus to the Max-Planck Institute Language Archive Technology.

In addition, for all partners, we foresee:

- Close cooperation with associations of the humanities and social sciences
- Specific workshops, trainings, summer schools for researchers from the humanities and social sciences

- Hosting and training of junior researchers in cooperation with Finnish graduate schools
- Development of teaching modules for student curricula both on bachelor's and master's degree levels

#### Annex 2-D

Contributions to the Knowledge Sharing Infrastructure mainly targeting national audiences and not already covered by Annex 1-D:

Members of the national consortium have their areas of expertise as mentioned in the description of each member. Each member creates and maintains resources, tools and services related to its area of expertise in FIN-CLARIN and will advise students and researchers of other FIN-CLARIN members on the creation and usage of such resources, tools and services. In particular, the following areas of activity are foreseen:

- Creation of a national web application hosting or linking to web services of all Finnish CLARIN centres
- Close cooperation with associations of the humanities and social sciences
- Specific workshops, trainings, summer schools for researchers from the humanities and social sciences
- Hosting and training of junior researchers
- Development of teaching modules for student curricula
- Support and help desk for users of FIN-CLARIN web applications

#### Annex 2-E

Any other contribution:

None.