



CLARIN AGREEMENT



CLARIN Agreement with the National Consortium of Iceland
(cf. Annex 1-A for the composition of the National Consortium)

This Agreement is made by and between:

- (1) CLARIN ERIC, a European Research Infrastructure Consortium established by Commission Implementing Decision (EU) 2012/136 of 29 February 2012 (hereinafter "CLARIN ERIC"), and,
- (2) the National Consortium of Iceland, represented by The Árni Magnússon Institute for Icelandic Studies (700806-0490) (hereinafter "National Consortium")

Hereinafter collectively referred to as the "Parties" and individually as "Party".

HAVE AGREED on the following specific terms and conditions for contribution of resources to CLARIN ERIC during the period February 1st, 2020 – December 31st, 2022.

Article 1 – Scope

The Parties have agreed to conclude an agreement in accordance with Article 6.5 of the Statutes, including the specific terms and conditions described in this Agreement which relate to Art. 2, 6.1 and 6.2 (d)-(m) of the Statutes.

Article 2 – Definitions

"National Coordinator" means the person appointed as a national coordinator by a CLARIN ERIC member in accordance with Article 6(2)(e) of the Statutes and which is responsible for the National Consortium.

"Statutes" means the Statutes of CLARIN ERIC including its Annexes.

Article 3 - Obligations

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

In addition to the specific contributions mentioned below in Article 3.2 below, the National Consortium shall (i) promote the adoption of relevant standards in national resource and tool creation projects, as well as the uptake of CLARIN services among researchers in their country, (ii) gather user feedback and requirements, and (iii) support CLARIN centres in the member country by facilitating integration into national and other relevant infrastructures.

3.2 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.

3.2.1. Contributions to be coordinated by CLARIN ERIC

(a) **Data and service centres:** In accordance with Art. 6.2 (f) of the Statutes, the National Consortium shall contribute at least one centre of type B (cf. the check list <https://office.clarin.eu/v/CE-2013-0095-B-centre-checklist-v6.pdf>).

Detailed descriptions of the centres and the services they offer are included in Annex 1-B. CLARIN centres are expected to obtain a deep level of integration with the infrastructure by connecting a substantial portion of their language resources and tools via:

- harvestable metadata for the VLO
- endpoints for the Federated Content Search
- web applications for the LR Switchboard

(b) **Access:** The National Consortium shall comply with statutes Article 19 of the Statutes on access policies for users. Any restrictions or exceptions are specified in Annex 1-F, to the extent it is applicable.

(c) **User Authentication and Authorization System:** The National Consortium shall comply with the specification given in Annex 1-C.

(d) **Knowledge sharing infrastructure (KSI):** The National Consortium shall contribute to the KSI by providing access to its knowledge and expertise to the CLARIN community at large in accordance with 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.

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

The National Consortium shall provide for:

(a) Enhancements and improvements to existing resources, tools and services in accordance with the detailed descriptions included in Annex 2-A.

(b) Creation of new resources, tools and services in accordance with the detailed descriptions 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 in accordance with the detailed descriptions 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 exchanging knowledge and expertise, and to advancing Humanities and Social Sciences research through the support by CLARIN. A brief description of those items that are not already covered by 3.2.1-(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).

3.3. National contributions made jointly with other parties

In accordance with Article 6.3 of the Statutes, 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 may be made between the CLARIN ERIC and those members / observers / third parties concerned.

Signatures

Date: March 6, 2020


Name: Eiríkur Rögnvaldsson
Function: National Coordinator

Date: 10 March 2020


Name: Franciska de Jong
Function: Executive Director

Date:

Annex 1 National contributions to be coordinated by CLARIN ERIC

Annex 1-A

The National Consortium of Iceland, CLARIN-IS, consists of the following partner(s):

1. The Árni Magnússon Institute for Icelandic Studies (coordinator) has competence in LRT, especially corpus linguistics and corpus development, lexicography, terminology, licensing, and Humanities archives.
2. University of Iceland, Institute of Humanities, has competence in linguistics, language resources, and LRT, especially treebanking.
3. Reykjavik University has competence in LRT, especially tagging and parsing, speech technology, artificial intelligence and neural networks.
4. The National and University Library has competence in digital archiving and cataloguing, and eLibrary functions.
5. National Archives of Iceland owns various types of archive materials and has competence in the handling of large digital resources archives and catalogues.
6. Icelandic Language Council has the role of giving advice to the Government in matters concerning the Icelandic language. It has expertise in language planning and language consultation.
7. Almennarómur is a self-owned foundation consisting of both academic institutions and private companies. It has been assigned the role of implementing the Government's Language Technology Program (2019-2023).
8. The Icelandic National Broadcasting Service (RÚV) is a company wholly owned by the State. It runs one TV station and two radio stations. RÚV has an enormous collection of spoken language resources which are in the progress of being digitized.

The National Coordinator is Prof. Eiríkur Rögnvaldsson, The Árni Magnússon Institute for Icelandic Studies, Þingholtsstræti 29, IS-101 Reykjavík - eirikur.rognvaldsson@arnastofnun.is, phone +354-525-4037.

The invoice for Annual fee is to be sent to Sigurborg Stefánsdóttir, Financial Manager, The Árni Magnússon Institute for Icelandic Studies, Árnagarði við Suðurgötu, IS-101 Reykjavík – sigurborg.stefansdottir@arnastofnun.is, phone +354-525-4025.

Annex 1-B

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

The Árni Magnússon Institute for Icelandic Studies has established a C-centre (<https://centres.clarin.eu/centre/64>). The institute will establish a service centre providing central national harvesting and long term preservation, and access to a repository of humanities resources.

The centre will aim at obtaining status as a CLARIN type B-centre, in accordance with the Technical and Scientific Description, Annex CLARIN Centre Types.

The Árni Magnússon Institute also aims at establishing a K-centre for the Icelandic language, in collaboration with the University of Iceland and the Icelandic Language Council. In addition to providing access to a number of Icelandic language resources, this centre will offer services such as question answering.

Annex 1-C

The Árni Magnússon Institute for Icelandic Studies and the University of Iceland are already connected to the CLARIN federation through the Danish national federation WAYF, as no national federation exists in Iceland. Until such a federation is established, users from other institutions in the National Consortium will have to get their own personal CLARIN account.

Annex 1-D

CLARIN-IS 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

CLARIN-IS 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:

CLARIN-IS participates in the Standing Committee for CLARIN Technical Centres and intends to participate in the CLARIN Legal Issues Committee.

Any other contributions are to be decided later.

Annex 2 National contributions to be coordinated at the national level

Annex 2-A

National Work Programme with timing:

Many of the consortium partners own a number of important language resources which they aim at giving access to through CLARIN. The coordinating partner, the Árni Magnússon Institute for Icelandic Studies, owns the greatest number of resources and many of them have already been made publicly available through META-SHARE and the local repository malfong.is.

The following resources have already been archived and registered in the CLARIN-IS repository and are publicly available:

The Database of Modern Icelandic Inflection	12-19
The Database of Modern Icelandic Inflection Core	12-19
ISLEX - Icelandic-Scandinavian dictionary	12-19
ISLEX recordings (pronunciation files)	12-19
IceNLP Natural Language Processing Toolkit	12-19
CombiTagger	12-19
Tokenizer for Icelandic Text	01-20
The Icelandic Confusion Set Corpus	01-20

The following resources will be enhanced, improved and made available through CLARIN in the course of the next few months:

The Saga Corpus	06-20
Icelandic Pronunciation Dictionary	06-20
Tagged Icelandic Corpus	06-20
The Icelandic Gigaword Corpus	06-20
The Icelandic Parliamentary Corpus	06-20
The Icelandic Spoken Language Corpus	06-20
The Icelandic Terminology Bank	06-20
MIM-Gold, standard for PoS tagging	06-20

From mid-2020, resources owned by other members of the consortium will gradually be added to the CLARIN infrastructure. It is expected that the preparation of these resources will be considerably more time-consuming than the preparation of those owned by the Árni Magnússon Institute. At present, it is not possible to give a timeline for these additions but the aim is to have added the most important resources from all consortium partners by the end of 2022.

Annex 2-B

National Work Programme with timing:

In the course of the next four years, a number of basic resources and tools for Icelandic will be created within the Government funded National Language Technology Programme. This includes resources and tools for speech recognition, speech synthesis, spell and grammar checking, and machine translation. These resources will be developed by SÍM (Samstarf um íslenska máltækni), a consortium consisting of three academic institutions, the National Broadcasting Service, The Association for the Visually Impaired, and four commercial companies, under the auspices of the self-owned institution Almannarómur, which is a

member of the Icelandic CLARIN consortium. Almannarómur will guarantee that all of these resources will be made freely available through CLARIN-IS. Since the detailed implementation plan for the National Language Technology Programme has not been finalized yet, it is not possible at present to make a timeline for the completion of these resources and tools but they are all supposed to have been submitted to CLARIN-IS by the end of 2023.

Annex 2-C

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

The coordinator will advise the partners to adopt CLARIN standards and will provide information to promote this adoption. The coordinator is a member of a special Standards group that the SÍM Consortium mentioned above has established. This group has been assigned the task of choosing standards and formats for the tools and resources that will be developed within the Language Technology Program and submitted to CLARIN-IS.

Annex 2-D

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

CLARIN-IS will cooperate with Almannarómur and with other European initiatives such as ELRC and ELG in promoting language technology by different means, such as national workshops and press releases.