



D5.1

Analysis of the situation in non-CLARIN countries

Document information

Title	Analysis of the situation in non-CLARIN countries
ID	CLARINPLUS-D5.1 (CE-2016-0846)
Author(s)	Bente Maegaard, Sussi Olsen, bmaegaard@hum.ku.dk , saolsen@hum.ku.dk
Responsible WP leader	Martin Wynne
Contractual Delivery Date	2016-08-31
Actual Delivery Date	2016-08-31
Distribution	Public
Document status in workplan	Deliverable

Project information

Project name	CLARIN-PLUS
Project number	676529
Call	H2020-INFRADEV-1-2015-1
Duration	2015-09-01 – 2017-08-31
Website	www.clarin.eu
Contact address	contact-clarinplus@clarin.eu

Table of contents

1	Executive Summary	3
2	Introduction	4
3	EU Member States	5
3.1	Belgium	5
3.1.1	Situation of roadmap	5
3.1.2	Core institutions	5
3.1.3	Comments	5
3.2	Croatia	5
3.2.1	Situation of roadmap	5
3.2.2	Core institutions	5
3.2.3	Comments	5
3.3	Cyprus.....	5
3.3.1	Situation of roadmap.....	5
3.3.2	Core institutions	5
3.3.3	Comments	6
3.4	France.....	6
3.4.1	Situation of the roadmap	6
3.4.2	Core institutions	6
3.4.3	Comments	6
3.5	Ireland	6
3.5.1	Situation of roadmap.....	6
3.5.2	Core institutions	6
3.5.3	Comments	7
3.6	Luxembourg.....	7
3.6.1	Situation of roadmap	7
3.6.2	Core institutions	7
3.6.3	Comments	7
3.7	Malta.....	7
3.7.1	Situation of roadmap	7
3.7.2	Core institutions	7
3.7.3	Comments	7
3.8	Romania.....	7
3.8.1	Situation of roadmap	7
3.8.2	Core institutions	7
3.8.3	Comments	8
3.9	Slovakia.....	8
3.9.1	Situation of roadmap	8
3.9.2	Core institutions	8
3.9.3	Comments	8
3.10	Spain	8
3.10.1	Situation of roadmap	8
3.10.2	Core institutions	8
3.10.3	Comments.....	8
3.11	United Kingdom.....	8
3.11.1	Situation of roadmap	8
3.11.2	Core institutions	8
3.11.3	Comments.....	9
4	Associated countries	10
4.1	Albania	10
4.1.1	Situation of roadmap	10
4.1.2	Core institutions	10
4.1.3	Comments	10

4.2	Bosnia and Herzegovina.....	10
4.2.1	Situation of roadmap.....	10
4.2.2	Core institutions.....	10
4.2.3	Comments.....	11
4.3	Faroese Islands.....	11
4.3.1	Situation of roadmap.....	11
4.3.2	Core institutions.....	11
4.3.3	Comments.....	11
4.4	Georgia.....	11
4.4.1	Situation of roadmap.....	11
4.4.2	Core institutions.....	11
4.4.3	Comments.....	11
4.5	Iceland.....	11
4.5.1	Situation of roadmap.....	11
4.5.2	Core institutions.....	11
4.5.3	Comments.....	11
4.6	Israel.....	11
4.6.1	Situation of roadmap.....	11
4.6.2	Core institutions.....	11
4.6.3	Comments.....	12
4.7	Serbia.....	12
4.7.1	Situation of roadmap.....	12
4.7.2	Core institutions.....	12
4.7.3	Comments.....	12
4.8	Switzerland.....	12
4.8.1	Situation of roadmap.....	12
4.8.2	Core institutions.....	12
4.8.3	Comments.....	12
4.9	The Former Yugoslav Republic of Macedonia.....	12
4.9.1	Situation of roadmap.....	12
4.9.2	Core institutions.....	12
4.9.3	Comments.....	12
4.10	Turkey.....	13
4.10.1	Situation of roadmap.....	13
4.10.2	Core institutions.....	13
4.10.3	Comments.....	13
5	Other countries.....	14
5.1	South Africa.....	14
5.1.1	Situation of roadmap.....	14
5.1.2	Core institutions.....	14
5.1.3	Comments.....	14
5.2	United States.....	14
5.2.1	Situation of roadmap.....	14
5.2.2	Core institutions.....	14
5.2.3	Comments.....	14
6	Conclusion.....	15
	References.....	16

1 Executive Summary

CLARIN ERIC was created in 2012 by 9 founding members. CLARIN ERIC now (August 2016) has 19 members. This is a very good progress, but there are still many EU and associated countries that have not yet joined.

This report describes the situation in these countries through the two ultimately most important pieces of information: 1) the situation of a national roadmap for research infrastructures, and 2) core institutions that would be relevant for a national CLARIN consortium, and finally any other comment. The report itself does not mention any person names or email addresses, but this information is available at the CLARIN Office.

The process of collecting the information for this survey has shown that there are many active and interested parties in non-member countries, and that several discussions and preparations for joining are going on. Contacts with those interested will be continued and will hopefully lead to more members joining. There are good contacts with relevant parties in all EU member states and with most associated countries. In a few associated countries no contact has been established yet, but CLARIN will keep its eyes open, and establish contacts when possible.

2 Introduction

CLARIN ERIC aims at covering as many languages as possible, and consequently CLARIN ERIC aims at having as complete a membership by European countries as possible.

CLARIN ERIC already has 19 members and one observer (August 2016). This document describes the situation in those EU member states that are not yet members of CLARIN ERIC and also includes a description of the situation in those of the associated countries for which we have been able to gather information.

We start in Section 3 with the EU member states, in alphabetical order, and then continue with associated countries in Section 4. Finally in Section 5 we add information about other third countries that have shown interest in joining CLARIN. In this section we only mention those countries that have or are preparing a Memorandum of Understanding with CLARIN.

3 EU Member States

Here we describe the situation in EU member states that are not yet CLARIN members. As can be seen from this section, there are 10 or 11 EU countries that are not yet members, depending on how the situation in Belgium is described. Here we have listed Belgium as a non-CLARIN country, but actually Flanders is participating.

Roadmaps on research infrastructures until October 2015 can be found here https://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri-national-roadmaps.

3.1 Belgium

Dutch Language Union, NTU, an intergovernmental institution on the Dutch/Flemish language, is a member of CLARIN ERIC.

3.1.1 Situation of roadmap

Roadmap is under preparation.

3.1.2 Core institutions

For Dutch Language Union: INT (Institute for the Dutch Language, formerly INL) in the NL.

For Flanders: CCL (University of Leuven, Centre for Computational Linguistics).

3.1.3 Comments

During the CLARIN Preparatory Phase (2008-2011) it was decided that the region Flanders would be the representing entity on behalf of Belgium. As the available funding in Flanders went to the NTU, it was decided that NTU would join CLARIN ERIC, and that the region of Flanders would participate under the NTU umbrella.

Discussions with Belgium, Flanders or NTU about the best future constellation have not yet taken place.

3.2 Croatia

3.2.1 Situation of roadmap

Roadmap published in April 2014.
CLARIN is mentioned.

3.2.2 Core institutions

University of Zagreb, Faculty of Humanities and Social Sciences

3.2.3 Comments

No further comments.

3.3 Cyprus

3.3.1 Situation of roadmap

Roadmap under preparation

3.3.2 Core institutions

Cyprus University of Technology, Department of Electrical Engineering,
Computer Engineering and Informatics, Digital Heritage Research Laboratory

3.3.3 Comments

No further comments.

3.4 France

3.4.1 Situation of the roadmap

The roadmap for France 2010-2020 contains the following paragraph:

“The strategy for this sector is marked by the desire to comfort the existing infrastructures and to confirm the French participation in the European projects DARIAH (digital infrastructure to study source materials in cultural heritage institutions), SHARE (data infrastructure for analysis of ongoing changes due to population ageing) and CESSDA (facilitate access of researchers to high quality data for social sciences) and to carry on the study for CLARIN (language resources) and ESS (European social survey).”

France has made a study of the situation wrt joining CLARIN, and is currently preparing to join as an observer.

3.4.2 Core institutions

France will be represented by CNRS via Huma-Num.

Huma-Num is a French Very Large Facility which aims to facilitate the digital turn in humanities and social sciences. Huma-Num has been created and is supported by the CNRS (National Centre for Scientific Research), the University of Aix-Marseille and the Campus Condorcet.

The French consortium is under construction and at present (June 2016) consists of the two French CLARIN C-Centres:

University Aix-Marseille, Speech and Language Data Repository

University Orléans-Tours, COCOON (Collections de COrpus Oraux Numériques)

The French Equipment Ortolang, a network infrastructure including a repository of language data wishes to become a CLARIN centre.

3.4.3 Comments

A broader initial consortium is under consideration.

3.5 Ireland

3.5.1 Situation of roadmap

Roadmap published in 2007

3.5.2 Core institutions

National University of Ireland, Galway

Dublin City University

National Library of Ireland

Royal Irish Academy (Digital Repository of Ireland)

These are the initial institutions – further national consortium is under development.

3.5.3 Comments

The Irish Research Council will be formally requested early in 2017 to consider again supporting CLARIN membership for Ireland. In the meantime, Ireland continues to be interested in attending CLARIN events and in maintaining contact with the ERIC.

3.6 Luxembourg

3.6.1 Situation of roadmap

No roadmap available

3.6.2 Core institutions

University of Luxembourg

3.6.3 Comments

The situation in Luxembourg with regard to language resources offers a rather scattered landscape, and a unified approach or platform on the national level is not yet available.

3.7 Malta

3.7.1 Situation of roadmap

Malta has made a report [The national Research and Innovation Strategy 2020](#) with a broader scope than just research infrastructures. This report describes clearly the need for infrastructures and the willingness to invest. CLARIN is not specifically mentioned, but ESFRI and the interest in connecting to large European infrastructures is clearly described, particularly in relation to ICT which has been identified as a key enabling area of specialisation.

3.7.2 Core institutions

University of Malta is the key institution. Several departments have expressed their interest.

3.7.3 Comments

No further comments.

3.8 Romania

3.8.1 Situation of roadmap

A roadmap for research infrastructures – the CRIC report – was published in 2007. This roadmap mentions CLARIN.

3.8.2 Core institutions

Romanian Academy, Research Institute for Artificial Intelligence
University "A.I. Cuza" of Iasi, ITI institute
"Politehnica" University of Bucharest,
University of Bucharest,
Linguistic Institute "Al. Philippide" of Iasi.
Technical University "Gheorghe Asachi" , Iași

There are several other universities with growing interests in these areas.

3.8.3 Comments

An important corpus – the Reference Corpus for the Contemporary Romanian, COROLA – is underway and will be finalized late 2017. This corpus has CMDI compatible metadata. The creation of this corpus is a priority project of the Romanian Academy and the management of the project is done by the two first institutes on the list above.

3.9 Slovakia

3.9.1 Situation of roadmap

No roadmap available

3.9.2 Core institutions

L. Štúr Institute of Linguistics, Slovak Academy of Sciences

3.9.3 Comments

Ongoing negotiations with the Ministry of Education.

3.10 Spain

3.10.1 Situation of roadmap

Roadmap: "National Scientific and Technological Infrastructure Map (ICTS)" for the period 2013-2016 and 2016-2020 where CLARIN is not mentioned.

3.10.2 Core institutions

IULA-UPF

Universidad Nacional de Educación a Distancia (UNED)

Universidad del País Vasco (IXA-UPV).

3.10.3 Comments

Spain has an acknowledged CLARIN Knowledge Centre with the three partners mentioned above. The knowledge centre offers tools and services for Spanish, Catalan and Basque texts.

3.11 United Kingdom

UK is currently observer in CLARIN.

3.11.1 Situation of roadmap

UK has published a report [INVESTING FOR GROWTH \(RCUK\), Strategic Framework for Capital Investment](#) from 2012, but the focus is different from the RI roadmaps and CLARIN is not specifically mentioned.

3.11.2 Core institutions

British Library

Centre for Research in Linguistics and Language Sciences, University of Newcastle (CRiLLS)

Centre for Corpus Research, University of Birmingham

Centre for Translation Studies, University of Leeds

Endangered Languages Archive, School of Oriental and African Studies, University of London

ESRC Centre for Corpus Approaches to the Social Sciences (CASS) and University Centre for Computer Corpus Research on Language (UCREL), Lancaster University

Institute for Language, Cognition and Computation (ILCC), University of Edinburgh

National Centre for Text Mining (NaCTeM), University of Manchester
Natural Language Processing Group, University of Sheffield
Oxford e-Research Centre, University of Oxford
Research Group in Computational Linguistics, University of Wolverhampton
School of Critical Studies, University of Glasgow
School of Humanities, Coventry University

3.11.3 Comments

A rearrangement of the funding councils of UK is ongoing including the creation of 'Research England' to oversee infrastructure funding for research.

4 Associated countries

All 14 associated countries that are not yet member of CLARIN have been contacted, either through a research contact or through their ESFRI representative, or both, but not all have responded. Below we list those countries in which we have established contacts. Information from additional countries (Moldova, Montenegro, Tunisia and Ukraine) will still be collected and used where possible.

4.1 Albania

4.1.1 Situation of roadmap

No roadmap available, but in the framework of government obligations, as a result of signing the EU Framework Programme Horizon 2020 agreement and its commitment towards ERA-European Research Area, three inter institutional Working Groups have been established by Decisions Nr 286,287,288 of Ministry of Education and Sport to draft a new National Strategy on Research, Technology and Innovation 2017-2022, a new Law on Research & Development and new Institutional Framework of National Research System. The Working Groups are undertaking a National Need Assessment on Research Human Resources and Infrastructures which will in the end lead to the creation of a Research Infrastructure Roadmap for Albania.

4.1.2 Core institutions

University of Tirana, Centre for Albanological Studies, Institute of Linguistics and Literature

4.1.3 Comments

No further comments.

4.2 Bosnia and Herzegovina

The official representing entity of the Bosnia and Herzegovina to the EU and other Countries is the Ministry of Civil Affairs - Bosnia and Herzegovina.

Bosnia and Herzegovina consists of two Entities – the Federation of Bosnia and Herzegovina and the Republic of Srpska, which are politically autonomous to an extent, as well as the district of Brcko, which is jointly administered by both. The Entities have their own constitutions and governments.

The Ministry of Civil Affairs - Bosnia and Herzegovina - is coordinator between two Ministries of Science and Technology from Entity level.

4.2.1 Situation of roadmap

No roadmap available.

4.2.2 Core institutions

National and University Library of the Republic of Srpska (Entity: Republic of Srpska)

Faculty of Philology Banja Luka, (Entity: Republic of Srpska)

Academy of Sciences and Arts of the Republika Srpska, (Entity: Republic of Srpska)

University Computer Center - URC Banja Luka (Entity: Republic of Srpska)

University Computer Center of East Sarajevo (URCIS) (Entity: Republic of Srpska)

National and University Library Dervis Susic - Tuzla (Entity: Federation of Bosnia and Herzegovina)

University Tele-informatics Center – UTIC - Sarajevo (Entity: Federation of Bosnia and Herzegovina)

University of Mostar - (Entity: Federation of Bosnia and Herzegovina)

4.2.3 Comments

In a future CLARIN-BH consortium both Entities will have to be represented with the joint objective to enhance a sustainable and coordinated network infrastructure and user community on top of partly existing and fragmented infrastructures.

4.3 Faroese Islands

4.3.1 Situation of roadmap

No roadmap available.

4.3.2 Core institutions

Faroese University, Department of Faroese language and literature
Faroese University, Department of history and social science

4.3.3 Comments

No further comments.

4.4 Georgia

4.4.1 Situation of roadmap

No roadmap available.

4.4.2 Core institutions

The Science Library in Tbilisi
Shota Rustaveli National Science Foundation <http://rustaveli.org.ge/en>
Ivane Javakhishvili Tbilisi State University, Arnold Chikobava Institute of Linguistics

4.4.3 Comments

No further comments.

4.5 Iceland

4.5.1 Situation of roadmap

There is no roadmap available yet. However, the Ministry of Science, Education and Culture and the Confederation of Icelandic Employers have recently joined forces with the aim of building infrastructure for Icelandic Language Technology. Each party has allotted 5 million ISK to hire a project manager who is currently collecting information, explaining the importance of Icelandic LT to IT companies, talking to academics, etc. He is supposed to come up with a plan in the fall, and the Minister has guaranteed that there will be money to implement that plan – hopefully with the support of industry.

4.5.2 Core institutions

University of Iceland
Árni Magnússon Institute for Icelandic Studies
Reykjavik University

4.5.3 Comments

No further comments.

4.6 Israel

4.6.1 Situation of roadmap

No roadmap available.

4.6.2 Core institutions

The departments of Computer Science at:

- The University of Haifa
- Technion, Israel Institute of Technology
- Bar Ilan University
- Ben Gurion University
- The Open University

4.6.3 Comments

No further comments.

4.7 Serbia

4.7.1 Situation of roadmap

RESEARCH FOR INNOVATION: Strategy on Scientific and Technological Development of the Republic of Serbia for the period 2016 – 2020. This document, which has a wider scope than just research infrastructures, was published in 2016. It does not mention any specific research infrastructures.

4.7.2 Core institutions

University of Belgrade (Faculty of Philology and Faculty of Philosophy)
University Library "Svetozar Markovic"

4.7.3 Comments

No further comments.

4.8 Switzerland

4.8.1 Situation of roadmap

Roadmap 2017-2020 proposal available at <http://www.sbfi.admin.ch/org/02839/index.html?lang=fr>. This has to be adopted before the end of 2016 by the Swiss parliament.

4.8.2 Core institutions

Universität Zürich, Institut für Computerlinguistik
Universität Zürich, Institut für Vergleichende Sprachwissenschaft
Università della Svizzera italiana, Faculty of Informatics, Lugano
Swiss national corpus (Schweizer Textkorpus). Zürich. <https://www.chtk.ch/>

4.8.3 Comments

No further comments.

4.9 The Former Yugoslav Republic of Macedonia

4.9.1 Situation of roadmap

No roadmap available

4.9.2 Core institutions

Faculty of Computer Science and Engineering
Institute of Macedonian Language "Krste Misirkov"
Research Centre for Areal Linguistics
Faculty of Philology " Blazhe Koneski"

4.9.3 Comments

No further comments.

4.10 Turkey

4.10.1 Situation of roadmap

Roadmap under preparation.

4.10.2 Core institutions

Istanbul Technical University

4.10.3 Comments

No further comments.

5 Other countries

While of course focussing on Europe and giving priority to European countries, the CLARIN interests go beyond Europe. Below we mention South Africa and the United States because these are the countries where we have more formalised collaboration, but there are active contacts in several other countries which may also develop into formal collaboration.

5.1 South Africa

5.1.1 Situation of roadmap

[A South African Research Infrastructure Roadmap](#) (SARIR report), November 2013 of the National Department of Science and Technology suggests the establishment of a National Centre for Digital Language Resources (NCDLR). Since then a new name has been decided: South African Centre for Digital Language Resources (SADiLaR). The SARIR report mentions CLARIN as a relevant RI to collaborate with. A decision on the funding of SADiLaR is expected late 2016 or early 2017.

5.1.2 Core institutions

SADiLaR will be hosted by the North-West University, and will also include the University of Pretoria, University of South Africa (UNISA), the Council for Scientific and Industrial Research (CSIR), and indirectly the University of the Free State and Stellenbosch University.

5.1.3 Comments

A concept Memorandum of Understanding (MoU) between SADiLaR and CLARIN ERIC is in the process of development. It will be formalised as soon as SADiLaR has commenced its work.

SADiLaR will integrate the current activities of the Language Resource Management Agency (RMA) which is also hosted by the North-West University
<http://www.rma.nwu.ac.za>

5.2 United States

A Memorandum of Understanding between Carnegie Mellon University and CLARIN ERIC was signed in 2014. The MoU aims at establishing a longer lasting collaboration agreement between CMU and CLARIN ERIC and also to investigate how other universities and institutions can be involved.

5.2.1 Situation of roadmap

Various proposals for infrastructural projects and activities in the area of language resources and tools have been made, but there is no national roadmap to refer to.

5.2.2 Core institutions

Carnegie Mellon University

5.2.3 Comments

Discussions are still ongoing on the exact way collaboration can be formalised.

6 Conclusion

The task of elaborating this survey has proven very fruitful, and has given highly useful input to CLARIN's efforts to enlarge the infrastructure in terms of members. CLARIN ERIC will carefully maintain the existing contacts. Contacts are being maintained through regular email or other contacts, by organising small meetings at conferences where possible, by inviting representatives of upcoming members to CLARIN workshops and to the annual conference. This approach has proven successful and we hope to be able to welcome more members in a not too far future.

Decision-making for this kind of matters takes time, however. There is a long way from a university or a group of institutions wishing to join, till a decision has been taken at the level of the government or research agency. Therefore CLARIN is continuously maintaining the relations to the researchers and to the research administrations.

References

National roadmaps for research infrastructures:

https://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri-national-roadmaps

<http://www.esfri.eu/national-roadmaps>

The ESFRI Roadmap 2016: <http://www.esfri.eu/roadmap-2016>

List of associated countries:

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/3cpart/h2020-hi-list-ac_en.pdf

National roadmaps

Below we list those national roadmaps we have been able to identify.

Note: Some of these documents have a broader scope than just research infrastructures (e.g. research and innovation strategy).

The documents are listed in the same order as the countries in sections 3, 4 and 5.

Croatia:

<http://public.mzos.hr/fgs.axd?id=21801>

France:

http://cache.media.enseignementsup-recherche.gouv.fr/file/Infrastructures_de_recherche/74/5/feuille_route_infrastructures_recherche_2016_555745.pdf

Ireland:

https://www.djei.ie/media/forfas061221_research_infrastructure.pdf

Malta:

[The national Research and Innovation Strategy 2020](#)

Romania:

http://www.research.ro/img/files_up/1212040805RAPORT%20CRIC%201_en.pdf

Spain:

http://www.idi.mineco.gob.es/stfls/MICINN/Innovacion/FICHEROS/ICTS_ing.pdf

UK:

[INVESTING FOR GROWTH \(RCUK\), Strategic Framework for Capital Investment](#)

Serbia:

<http://www.mpn.gov.rs/wp-content/uploads/2015/08/Strategija-ingleski-jezik.pdf>

Switzerland:

<http://www.sbf.admin.ch/org/02839/index.html?lang=fr>.

Ukraine:

http://www.bilat-ukraina.eu/_media/BILATUKRAINA_PolicyBrief5_Research_Infrastructuresv02_final.pdf

South Africa:

[*A South African Research Infrastructure Roadmap*](#) (SARIR report),