EUROPEAN LANGUAGE EQUALITY 2

D1.2

Report on consultations with funding agencies, policy makers etc.

Authors	Sabine Kirchmeier, Georg Rehm, Marie Mattson, Davyth Hicks, Jane Dunne
Dissemination level	Public
Date	12-05-2023

About this document

Project	European Language Equality 2 (ELE2)
Grant agreement no.	LC-01884166 – 101075356 ELE2
Coordinator	Prof. Dr. Andy Way (DCU)
Co-coordinator	Prof. Dr. Georg Rehm (DFKI)
Start date, duration	01-07-2022, 12 months
Deliverable number	D1.2
Deliverable title	Report on consultations with funding agencies, policy makers etc.
Type	Report
Number of pages	40
Status and version	Final
Dissemination level	Public
Date of delivery	12-05-2023
Work package	WP1: SRIA: Further Consultations and Documentation of Stakeholder Commitment
Task	Task 1.3 Liaise with national and international funding agencies and policy makers as well as national and regional language institutes
Authors	Sabine Kirchmeier, Georg Rehm, Marie Mattson, Davyth Hicks, Jane Dunne
Reviewers	Federico Gaspari, Maria Giagkou
EC project officer	Susan Fraser
Contact	European Language Equality 2 (ELE2) ADAPT Centre, Dublin City University Glasnevin, Dublin 9, Ireland
	Prof. Dr. Andy Way – andy.way@adaptcentre.ie
	European Language Equality 2 (ELE2) DFKI GmbH Alt-Moabit 91c, 10559 Berlin, Germany
	Prof. Dr. Georg Rehm – georg.rehm@dfki.de
	http://www.european-language-equality.eu
	© 2023 ELE2 Consortium

Consortium

1	Dublin City University (Coordinator)	DCU	IE
2	Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (Co-coordinator)	DFKI	DE
3	Univerzita Karlova (Charles University)	CUNI	CZ
4	Universidad Del Pais Vasco/ Euskal Herriko Unibertsitatea (University of the Basque Country)	UPV/EHU	ES
5	Athina-Erevnitiko Kentro Kainotomias Stis Technologies Tis Pliroforias, Ton Epikoinonion Kai Tis Gnosis	ILSP	GR
6	European Federation of National Institutes for Language	EFNIL	LU
7	Réseau européen pour l'égalité des langues (European Language Equality Network)	ELEN	FR

Contents

List of Figures

List of Tables

List of Acronyms

AI	Artificial Intelligence
CLARIN	Common Language Resources and Technology Infrastructure
CRACKER	Cracking the Language Barrier (EU project, 2015–2017)
DLE	Digital Language Equality
EC	European Commission
EFNIL	European Federation of National Institutes for Language
ELE	European Language Equality
ELE1	European Language Equality (preceding project)
ELE2	European Language Equality *(this project)*
ELE Programme	European Language Equality Programme *(the long-term, large-scale funding programme specified by the ELE project)*
ELEN	European Language Equality Network
ELG	European Language Grid (EU project, 2019-2022)

EP	European Parliament
EU	European Union
GDPR	General Data Protection Regulation
LT	Language Technology/Technologies
LLM	Large Language Models
META	Multilingual Europe Technology Alliance
META-NET	EU Network of Excellence to foster META
MEP	Member of the European Parliment
SRIA	Strategic Research and Innovation Agenda
STOA	Science and Technology Options Assessment

Abstract

In order to reach digital language equality (DLE) in Europe by 2030, support and commitment from different stakeholder groups are necessary. This report presents the efforts that have been taken to reach policy-makers, national language institutes, regional language institutes and funding agencies. In particular, policy-makers were reached through workshops and virtual meetings between Commissioner Mariya Gabriel (European Commission), MEP Jordi Solé (European Parliament) and representatives from the ELE 2 project. Consultations with national and regional language institutes were done through an online questionnaire which was then followed up by one-to-one consultations. In addition, a list of 103 relevant funding agencies covering 48 European countries and regions was created for future consultation.

1 Introduction

It is envisaged that achieving digital language equality (DLE) in Europe by 2030 can only be accomplished through the support and commitment of multiple stakeholder groups working towards the goal of establishing a joint large-scale programme. ELE 2 partners engaged with the relevant policy-making bodies and funding agencies, research, industry and user stakeholders on European, national and regional levels to systematically expand the list of stakeholders identified in ELE 1, and document their commitment. This deliverable outlines the activities undertaken in order to reach relevant stakeholders, including the organisation of questionnaires, direct consultations, etc as presented in the Specification of approach for consultations and for documentation of stakeholder commitment (Hegele et al., 2022).

2 Overview of funding agencies, policy makers, national and regional language institutes in Europe

As part of ELE 2, national and international funding agencies and policy makers were singled out due to their relevance and importance for the goal of achieving DLE for all languages of Europe by 2030. National and regional language institutes were also targeted due to their central roles in monitoring the official language(s) of their countries, advising on language use and learning, or developing language policies. They are additionally particularly relevant to the objectives of ELE due to the fact they usually maintain the national or regional corpora as well as other language datasets and digital resources. These resources are, in many cases, only accessible for research purposes. This deliverable documents the interactions with these stakeholders to "unlock" these national or regional corpora and to make them available for computational purposes, especially for the development of large language models (LLMs) for the languages of Europe.

EFNIL and ELEN together with all ELE 2 partners concentrated on liaising with these two stakeholder groups. Through various activities like joint meetings, exploiting synergies and collaborations with ongoing related projects (e. g. OpenGPT-X in Germany) it was envisaged that ELE 2 would align objectives and drive home the message that well curated language data from the national language institutes can make a difference. The effort made by EFNIL and ELEN was targeted to convince the national language institutes that (a) their data will not be sold or published if they are submitted to the training of LLMs, that (b) it will not be possible to reconstruct their data if they are used to train LLMs, and (c) that their submission of data for training purposes is essential to bring about DLE for all European languages.

3 Consultations with policy-makers

For ELE 2, the project consortium set out to use the project results from ELE 1 further to communicate the potential of Language Technologies (LTs) made *in* Europe *for* Europe, especially to policy-makers and to key political and industrial decision-makers. This was a continuation of the highly important strategic roadmapping activity that was originally started in META-NET and then continued in CRACKER, among other EU-funded projects. Given that ELE 2 was a 12-month project, targeted meetings with specific stakeholders were organised, acknowledging the critical role that the funding agencies and policy-makers (European, national and regional) have in the realisation and sustainability of the Strategic Research and Implementation Agenda (SRIA) and Roadmap drawn up by the ELE (ELE Consortium, 2022; Rehm and Way, 2023) to achieve DLE in Europe by 2030.

3.1 European Parliament (Science and Technology Options Assessment Panel, STOA)

On 8 Nov. 2022 the Science and Technology Options Assessment (STOA) Workshop "Towards full digital language equality in a multilingual European Union" took place, organised by the ELE 2 consortium together with the European Parliament.[1] It was the third STOA workshop to feature the topic of LT after events in 2013 and 2017. STOA Panel member Jordi Solé chaired and opened the workshop by underlining the importance of protecting multilingualism in the EU, which comprises more than 90 national, regional and minority languages across Europe. MEP Solé was followed by Prof. Andy Way (DCU) who introduced the ELE project as co-coordinator and briefly presented its strategic agenda to achieve full DLE in Europe by 2030. "If the EU does not address the matter properly, the worst-case scenario is that some or maybe even most of these languages will eventually suffer from digital [...] extinction," concluded Prof. Andy Way. After the introduction, Dr. Maria Giagkou (ILSP) provided a snapshot of the results from the first 18 months of the ELE project highlighting the considerable inequality in technological support amongst European languages, unsurprisingly showing English, French, German and Spanish as particularly dominant. Examples of how three languages – Irish, Bulgarian and Basque – are currently digitally supported were presented by researchers and speakers from their respective countries (Dr. Teresa Lynn, DCU; Prof. Svetla Koeva, BAS; Prof. German Rigau, EHU), with ELE co-coordinator Prof. Georg Rehm (DFKI) presenting the overarching recommendations from the SRIA (the ELE ELE Consortium (2022)). The session concluded with a panel discussion involving Jordi Solé (Greens/EFA, Spain) and Yana Toom (Renew, Estonia), policy analysts from the European Parliament (Magdalena Pasikowska-Schnass) and the European Commission (June Lowery-Kingston) and European language technology experts (Prof. Barbara Plank, LMU).

In addition to the ELE-centric workshop described above, another relevant exchange took place with the STOA panel. In late 2022, early 2023, DFKI produced the technical feasibility study "European Streaming Platform for National News Accessible in all EU Languages" for the STOA. The final study was presented by the study coordinator Prof. Georg Rehm to the members of STOA on 20 April.[2] The study explores the technical feasibility of the development of a European streaming platform, which processes the media programmes of European broadcasters in such a way that they are translated into all European languages and then made available for search and discovery through a central platform. This platform would have to make use of various sophisticated LTs that support *all* European languages,

---

[1]   https://www.europarl.europa.eu/stoa/en/events/details/towards-full-digital-language-equality-i/20220711WKS04301

[2]   https://multimedia.europarl.europa.eu/en/webstreaming/panel-for-future-of-science-and-technology_20230420-0930-SPECIAL-STOA

which is why there is a direct connection to the vital topic of DLE. The study recommends, among others, to foresee a tight collaboration between such a potential future European streaming platform project and European Language Equality (ELE), (ELG), European Language Data Space (LDS) and European Media Data Space. If set up in the right way, such a joint collaboration could produce very positive effects with regard to the production of new datasets for all European languages and for arriving at full DLE across all European languages eventually.[3]

3.2 European Commission

On 14 March 2023, a virtual meeting between Commissioner Mariya Gabriel (European Commission), MEP Jordi Solé (European Parliament) and ELE co-coordinators Prof. Andy Way (DCU) and Prof. Georg Rehm (DFKI) took place. The goal of the meeting was to present the key results and main recommendations of the ELE project up to that point and to discuss potential next steps towards the realisation of the ELE Programme. Commissioner Gabriel expressed her support for the ELE recommendations and suggested to organise, as a next step, a meeting with Commissioners Hahn, Breton and herself to discuss how a complex and ambitious endeavour such as the ELE Programme can be structured, organised and financed. This follow-up meeting with the three Commissioners is currently being planned, considering the respective commitments. ELE collaborates with colleagues in the European Parliament including Jordi Solé's office, in order to organise this meeting.

3.3 Language European Digital Infrastructure Consortium (EDIC)

The European Digital Infrastructure Consortium (EDIC) is a new mechanism which became available to Member States in January 2023 as a way to implement Multi-Country Projects. EDICs provide a specific set of features, including swift set-up by several Member States (while keeping open access for others to join throughout the duration of the EDIC), legal personality or the possibility to combine different sources of funding. It is envisaged that such features will facilitate infrastructure deployment and ensure its long-term sustainability.

While an EDIC would implement a concrete Multi-Country Project it is not considered a funding programme. In early 2023, multiple EU Member States have expressed their interest in developing a *Language EDIC*, i. e., a Multi-Country Project that addresses LT-related topics and goals. At the time of writing (early May 2023), a working group with representatives of those countries had expressed their interest in developing the main components of the Language EDIC in collaboration with the Language Data Space procurement project, which started in January 2023 and which is coordinated by DFKI. The emerging Language EDIC represents an important landmark insofar as it is the first time that an official, pan-European coordination body has the opportunity to discuss, decide upon, work on and influence high-level topics. DFKI will attempt to ensure that the ELE recommendations and DLE are among the high-level topics to be addressed by the Language EDIC if/when it is established later in the second half of 2023.

4 Consultations with national language institutes

The European Federation of National Institutions for Language (EFNIL) is a network of the central or national institutions for research, documentation and policy relating to the officially recognised standard languages in Europe. EFNIL mainly represents the official lan-

---

[3] While the study is finalised, it has not been published yet on the European Parliament website.

guages of the European countries including Georgian and Ukrainian. Some EFNIL institutions, however, cater also for minority languages.

EFNIL institutions are mainly research institutions and specializing on the national languages and not on LT as such, although many institutions are rather advanced users of specialized LT solutions for the purpose of creating language descriptions such as structured text collections (corpora), dictionaries and grammars. Some of EFNIL's member institutions are strongly involved in language policies of their countries and may even be in charge of implementing and monitoring them, and some also have the responsibility of developing LT services for the public user.

EFNIL institutions maintain close contacts with other private or public language institutions and research institutes in their countries and publish a list of these institutions as part of the European Language Monitor which is available on the EFNIL website[4]. These contacts have also been included in the consultations.

The consultations with language institutions were primarily based on:

- a questionnaire addressing missing language resources (LRs)

- Direct consultations with selected respondents to the questionnaire

4.1 Methodology and background of the questionnaire

The target group of the questionnaire were European language institutions that are involved in documenting, developing and safeguarding national, regional and minority languages and the cultural heritage that they represent. EFNIL was responsible for sending out the questionnaire and perhaps therefore most of the respondents are members of EFNIL. For a list of recipients of the questionnaire, see Appendix A and B.

4.1.1 Background of the questionnaire

There were three main goals of the questionnaire:

- **Missing language resources:** Find out if there are LRs that are missing from the ELG platform and, if possible, help upload them to ELG;

- **Difficulties:** Find out what issues language institutions have when sharing digital language data and what kind of changes they want to see;

- **Contact persons:** Find contact persons from language institutions to discuss missing LRs and DLE further and guide through the ELG platform if nessercary.

The questionnaire was sent out to language institutions via e-mail. It consisted of a total of 23 questions of which 9 were closed (single or multiple choice) and 14 open. Four questions were mandatory and 19 were optional (see Table 1). The full questionnaire, as published online, is presented in Appendix D.

The survey was structured in four main parts.

- **Part A. Respondents' profiling:** the first part of the survey included three questions for the demographic profiling of respondents with emphasis on characteristics relevant to the task at hand, i. e.

  - Name of the organisation/representative body the respondents work for
  - Country the respondents are based in

---

[4]   http://efnil.org/projects/elm

	Mandatory	Optional	Totals
Closed	1	8	9
Open-ended	3	11	14
Totals	4	19	23

Table 1: Type of survey questions

- Languages the organisation works with

- **Part B. Creation of digital LRs: looked into whether the institutions work with digital LRs, i. e.,:**
  - If they are responsible for developing or maintaining digital LRs
  - The future plans regarding digital LRs

- **Part C. Difficulties in sharing of digital LRs, i. e.,**
  - What challenges they face when working with sharing of LRs;
  - What measures should be taken to make it easier to share digital LRs;

- **Part D. Information about the respondent, i. e.,**
  - Their name and title
  - Their contact information
  - Any optional additional comments

Follow-up: When respondents gave their contact information, they accepted that they could be contacted to discuss these questions further.

The survey was distributed by EFNIL through emails to members of the EFNIL network and other language institutions in Europe that are not members of EFNIL. The survey was sent out to 21 EFNIL member institutions and 63 other European language institutions. The survey was opened on 24 February 2023 and closed on 26 April 2023. In total, 25 responses were collected from national language institutes.

4.2 Analysis of questionnaire responses

4.2.1 Respondent's profiling and language coverage

In total, 20 countries were covered by the questionnaire. Out of all countries represented, there were more than two respondents from four countries (Belgium, Denmark, Luxembourg and Slovenia). There were two countries with three respondents (Iceland and Norway). The following countries were covered via the survey: Belgium, Croatia, Czech Republic, Denmark, Estonia, Greece, Iceland, Latvia, Lithuania, Luxembourg, Malta, Norway, Serbia, Slovakia, Slovenia, Switzerland, The Netherlands, United Kingdom and Wales. A total of 41 languages were covered by the institutions who responded (see Figure 1).

4.2.2 Creation of digital language resources

Most respondents, 18 (72%), replied that they are responsible for developing, updating or maintaining digital LRs, while seven (28%) stated that they are not (Figure 2). Digital LRs include dictionaries, corpora, grammars and language descriptions among others.
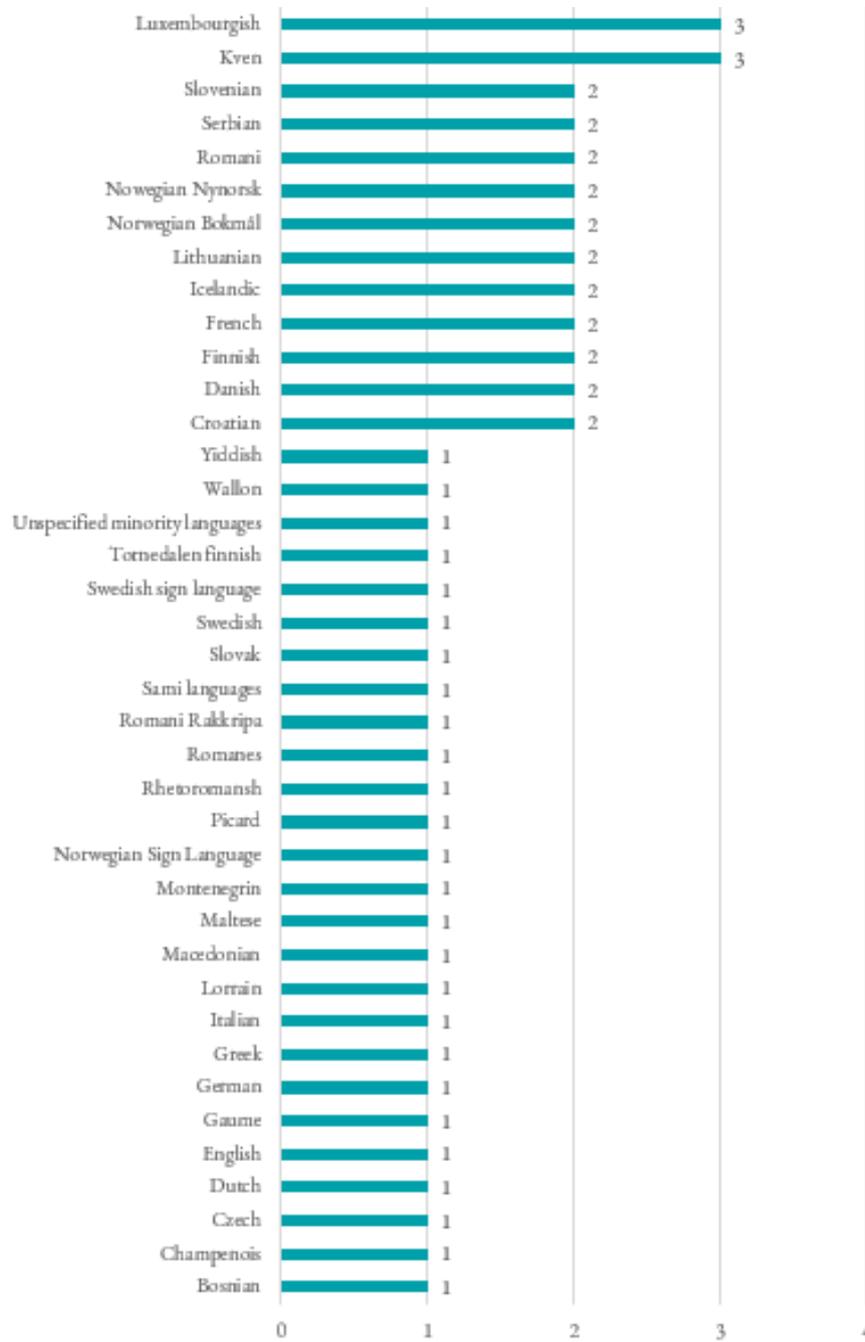
Figure 1: Which language(s) is your institution concerned with, according to its mandate?
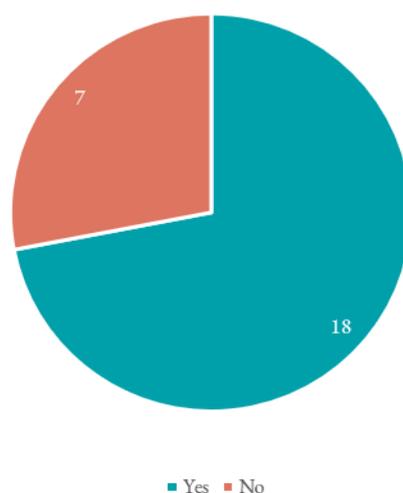
Figure 2: Is your institution responsible for developing, updating or maintaining LRs?

When asked to provide more information about what type of resources they are responsible for, 18 (72%) of them did. Most responses consisted of lists with examples of several openly available digital LRs, as well as some links to platforms or repositories such as CLARIN, ELG or Github. The most commonly mentioned type of digital LR was electronic dictionaries, which was mentioned by 15 language institutions, followed by digital corpora, which was mentioned by 10 language institutions, as can be seen in Table **??**. It is, however, important to note that this is not a complete list of what digital LRs the responding institutions are responsible for. It is only a list of the types of resources mentioned in the open-ended responses.

Type of resource	Quantity
Online dictionary	15
Corpora	10
Linguistic Software/LT tool	8
Terminological resources	7
Other kinds of lexicon	5
Specialized linguistic database	4
Unspecified/other	5

Table 2: Please provide more information about the resources your institute is responsible for. (Summary)

Sixteen language institutions responded that they plan to create more digital LRs in the future as seen in Figure 3. Six respondents stated that they do not have any plans for creating future resources and three responded "other". They also commented that their plans for future development of LRs depend on external factors such as funding.

Of all 25 respondents, 15 institutions provided more information about what LRs they are planning to create in the future. Several of the responses mainly focused on maintaining and expanding the resources they currently have. Some examples of the responses include:

- digital and multilingual dictionaries
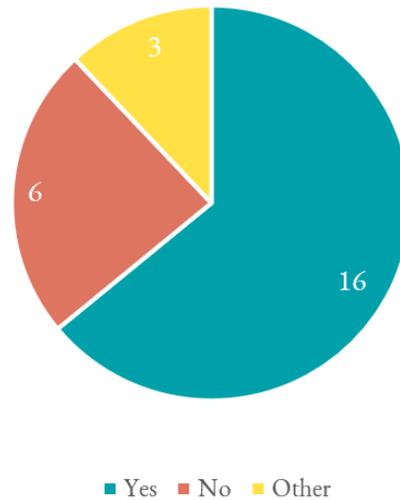
Figure 3: Is your institution planning to create more resources in the future?

- corpora, especially multimodal and audio corpora
- language models
- LT tools such as catcher of neologisms
- expansion of current resources

A little less than half of the respondents reported that there are LRs that they do not have access to, but would like to (see Figure 4). When asked to provide more information about which resources they would like to access, larger and better corpora, literary texts, audio data, historical linguistic resources and better LT tools are a few examples that respondents mentioned. Some also mentioned the need for more resources for minority languages.
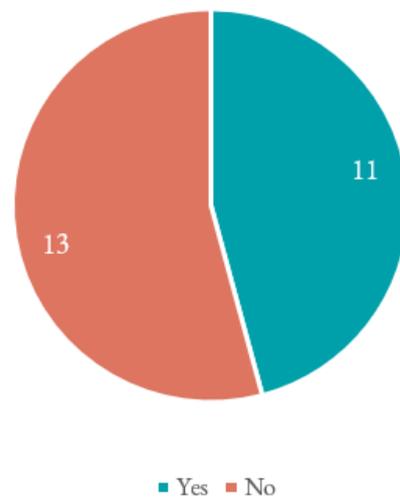
Figure 4: Are there LRs that you would like to use but do not have access to (either because they are possessed by someone else or do not exist yet)?

When it comes to the respondents' willingness to share their LRs, 14 stated that they are willing to share all their resources and 7 replied that they are willing to share some of their resources (see Figure 5. None of the respondents picked the option that they are not willing to share any resources. Three respondents picked the "other". Most of them stated that they are willing to share as many of their resources as possible, but there could possibly be some restrictions such as copyright or collaboration partners that could be unwilling to share resources widely.

- ■ We are willing to share all our resources　　■ We are willing to share some of our resources
- ■ We are not willing to share any of our resources　■ Other
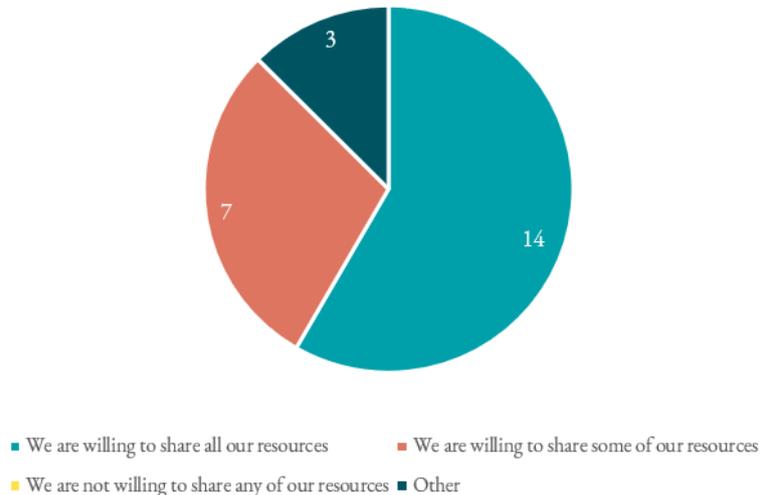
Figure 5: What is your policy regarding the sharing of your LRs? Optional question.

When asked what kind of access can be granted to LRs, the organizations could choose among multiple options. 12 responded that they can be viewed online, 13 responded that they can be downloaded, eight said they can only be used for academic or non-commercial purposes and two responded that they can be acquired through a commercial license agreement. None of the respondents stated that they do not give access at all and 8 individuals responded with "other" (see Figure 6. Most of the respondents commented that they share as much as possible, but it sometimes depends on copyright or other external factors.

Figure 6: What kind of access can be given to your LRs? Optional question, multiple-choice.

Most respondents, 14, stated that they have resources published on an online data platform, such as CLARIN or ELG, and 10 respondents (42%), stated they have not shared resources on any platform (see Figure 7). The following platforms were mentioned in the optional comment:

- ELRC-SHARE

- CLARIN

- ELG

- META-SHARE

- HuggingFace

- Github

- National platform/other

One respondent mentioned that they would like to give access to an online data platform.
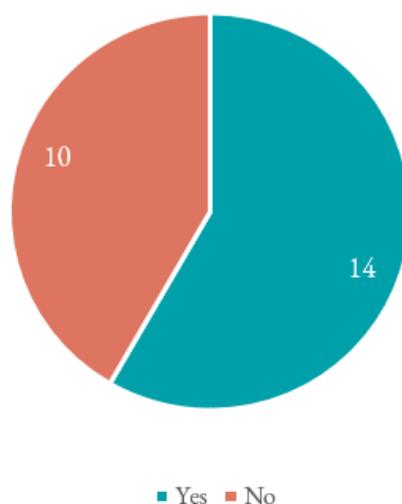
Figure 7: Do you have resources that have already been uploaded to CLARIN, ELG Platform, META-SHARE, Hugging Face, Github or other similar online data platforms?

4.2.3 Issues that prevent sharing of digital languages resources

The survey section on issues that prevent sharing of digital LRs started with an open-ended question on the kind of resources that the respondents have not made publicly available. There were ten responses to this open-ended question. The responses to this question can be summarised into two groups:

- All language data is already shared.

- There are some types of data, such as audio or literature, where the copyright is not clear.

When asked about the reasons for not making resources publicly accessible, copyright issues was the most common one, with nine responses. Five stated that it was because of privacy concerns, four responded that it was due to lack of funding or personnel (see Figure 8). Most of the five respondents who selected "other" stated that they have not yet considered these issues or that the specific questionnaire item was not relevant.
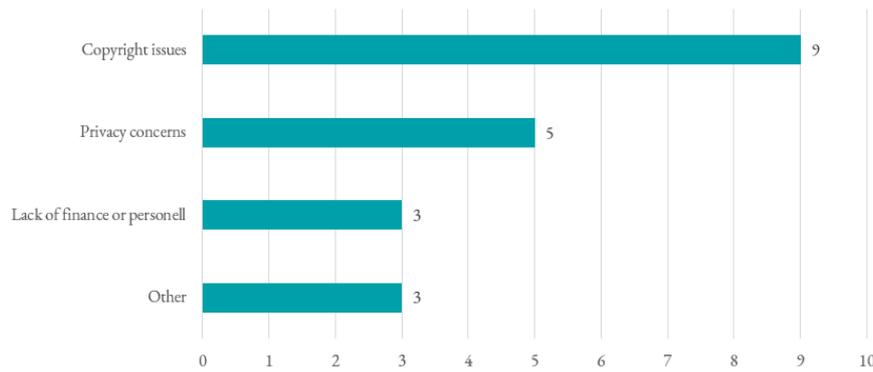
Figure 8: What are the reasons for not making these resources accessible?

When asked to provide more information about the issues, most responded that some authors do not want to give full access to their resources. Some respondents also mentioned that their resources might be made available with some restrictions, such as for research purposes only or through a corpus investigation interface. One respondent noted that most of their audio data cannot be shared due to privacy concerns. One respondent also mentioned ethical considerations, such as old recordings of people who are not able to consent to the publication of their recordings.

Among the 12 institutions that suggested specific measures to facilitate the full or partial availability of resources, almost all (nine respondents), argued that legal measures should be taken. The second most frequent response was financial support, with five institutions indicating this as a desirable measure and three institutions expressed that technical measures, such as anonymisation and/or scrambling, would be helpful. Additionally, three institutions mentioned other measures (see Figure 9).
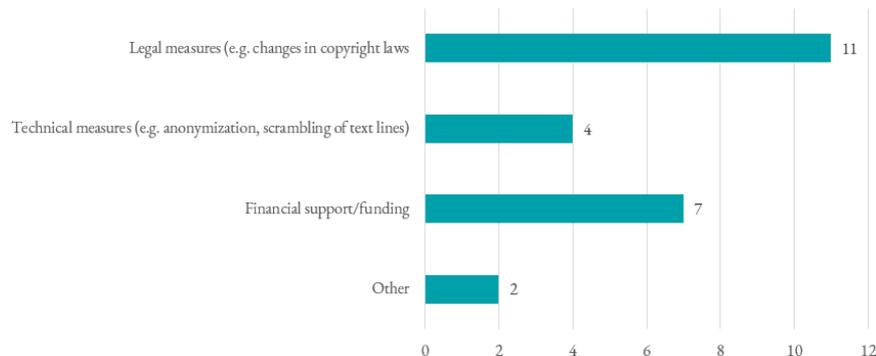
Figure 9: What measures should be taken so that you could make the non-accessible resources fully or partially accessible?

When asked to describe other possible measures that would be helpful, only two institutions responded. They stated that scrambling of data might not be enough and that legal measures are the most important.

4.3 Direct consultations

Some of the respondents of the questionnaire were contacted to discuss the questionnaire in more detail. The goal of these direct consultations was to get a deeper understanding of how

language institutions work with digital LRs and what issues they face when sharing digital LRs. Three national language institutes and one regional language institute participated in these targeted follow-up online meetings. A summary of the direct consultations is presented in this section.

4.3.1 Structure of consultations

The consultations consisted of four main parts:

- Background of the ELE 2 project.

- Background of the language institute.

- Discussion about digital LRs.

- Discussion about issues that prevent from sharing LRs.

The meetings lasted for about 20 - 40 minutes. All the follow-up meetings were one to one meetings, except for one case that the language institute was represented with several participants.

4.3.2 Result of the direct consultations

In the first part of the meeting the main goals of the ELE 2 project were presented and the participant was requested to endorse the SRIA.[5]

The next section of the meeting was about the language institute: what languages they work with and in what way. Two of the institutes (the Norwegian Language Council and the Võro institute) work with minority languages and three institutes (the Norwegian Language Council, the Centre for the Greek Language and Institut Grand-Ducal) work with national languages. The following are the covered languages:

- Greek

- Kven Finnish

- Luxembourgish

- Norwegian

- Norwegian sign language

- Romani

- Seto

- Võro

Not all language institutes were responsible for developing and maintaining LRs, but those who did had already uploaded most or all of them to platforms such as ELG. All participants expressed that there is a need for more LRs in their respective languages, namely large corpora, syntactically tagged corpora. The Centre for the Greek language also mentioned the need for better and publicly available LT tools, such as OCR and tokenisation tools. Most organizations planned to publish more resources in the future, such as spoken corpora, annotated sign language corpora and general expansions and improvements of already existing

---

[5] https://european-language-equality.eu/endorse-the-ele-sria/

LRs. After discussing issues that might prevent the sharing of digital LRs, it was clear that copyright issues were the main problem. One solution to these issues has been to only make resources available for academic or non-commercial purposes, as this might make authors and copyright holders more willing to give access to resources. This has also been a solution for making spoken data, such as interview recordings, available. When discussing funding, one issue that was brought up is that most of the financial support comes from funding for isolated projects that are limited in time, and it is therefore difficult to plan long term work with LRs. Personnel issues were also mentioned by some language institutes. It is especially difficult to find personnel with interdisciplinary competences. This is especially an issue for smaller languages, e. g., Võro. Technical problems were not brought up as a main issue by the institutes. However, the Centre for the Greek language mentioned that anonymisation might be useful for making future resources publicly accessible. An example was students' exams which would have to be anonymised before publishing.

5 Consultations with regional language institutes

5.1 ELEN

ELEN is the leading international organization for the protection, promotion and well-being of European lesser-used languages. ELEN comprises 174 member organisations representing 50 different European languages across 25 States. The majority of ELEN members are civil society organisations working to maintain and develop their languages, as well as universities and research institutes. For ELEN the aim of achieving DLE for all European languages is a vital part of the revitalization process and the overarching aim of maintaining European language diversity. To foster work in this direction, the ELE project brought together LT experts from across Europe and language organisations such as ELEN and EFNIL with the support of the EU.

5.2 Methodology and background of the questionnaire

This section presents the results of a survey of government supported territorial (hereafter RML) language institutes and uses the survey compiled by EFNIL (see the methodology discussion in Section 4.1)[6]. In using the same survey it adds some uniformity to the responses and acts so as to not differentiate between EU official and non-official European languages. ELEN sent the questionnaires by e-mail to most of the territorial language institutions in Europe. These language institutes are not ELEN members, but ELEN regularly collaborates with them in terms of language policy development and implementation. Most of these institutions are supported or are a department of an autonomous or regional government with the responsibility and competences for the development of their respective language, as well as having a budget to support language projects. In total nine responses were received from RML Institutes. These are: Welsh Government (Welsh); Ofis ar Brezhoneg (Breton); Bòrd na Gàidhlig (Gaelic); Võro Institute (Võro) ; Province of Fryslân (Frisian); Centro Ramón Piñeiro (Galician); Cornwall Council (Cornish); Catalan Government (Catalan and Aranes); and the Welsh Language Commissioner (Welsh).

---

[6]   One regional language institute, Võro institute, was also a part of a direct consultation, which is presented in Section 4.3 along with the other direct consultations with language institutes.

5.2.1 Creation of digital language resources

Question: Is your institution responsible for developing, updating and/or maintaining digital LRs?

The vast majority (88.9%, all except one) replied that they are responsible for developing, updating and/or maintaining digital language resources (see Figure 10). Digital LRs include digital dictionaries, corpora, grammars and language descriptions, amongst others.
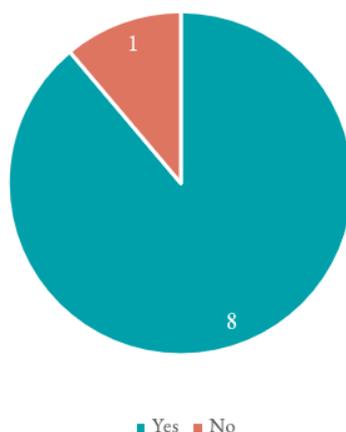
Figure 10: Is your institution responsible for developing, updating or maintaining LRs?

Question: Please provide more information about the resources your institute is responsible for.

The institutes were asked to provide more information about their digital LRs. Most responses consisted of lists with examples of digital LRs. A clear trend is that the "bigger" the language and the more power the government has, the more resources it had. For example, Catalan (c. 10 million speakers and with an autonomous government with competence for language technology development) has far more resources and a broader range of resources than Cornish (c. 1000 speakers with no autonomy and no budget or competences for LT development). The most commonly mentioned type of digital LR was digital dictionaries, which was mentioned by 9 language institutions, and digital corpora, also mentioned by 9 language institutions, as well as maintaining a place-names database (9). Most organisations (8) also worked on developing machine translation and speech synthesis, as can be seen in Table 3. It is, however, important to note that this is not an exhaustive list of what digital LRs the responding institutions have, just a list of the ones mentioned in the responses. Note also how the larger more well-resourced administrations (e. g., Catalonia) mentioned that they are able to out-source LT development work, while the less well-resourced administrations are conducting most of their LT work in-house (e. g., Brittany).

5.2.2 Language Resources in the Future

Question: Is your institution planning to create more resources in the future?

All institutes replied that they are planning to create more resources in the future.

Question: Please, describe the resources that you are planning to create.

Type of resource	Quantity
Online dictionaries	9
Online corpora	9
Place-names database	9
Machine translation	8
Speech technology	8

Table 3: Types of resources mentioned

All institutes replied and described which resources they are planning to create, giving a wide range of responses. Nearly all respondents stated that they would expand existing resources such as online dictionaries and several said they would start to develop new resources such as speech-to-text and aim in the long term for machine translation. The Catalan Government described its AINA project which creates corpora and language models for Catalan to enable companies to create their own apps such as voice assistants.

Question: Are there LRs that you would like to use but do not have access to (either because they are possessed by someone else or do not exist yet)?

All except one respondent said that there are LRs that they would like to use but don't have access to.

Question: describe the resources that you would like to be able to use.

There was a whole range of responses on the question of resources that institutes would like to use with each institute giving different answers. These included book texts (Voru), speech technology and conversational AI (Welsh), large language corpora (Breton), screen readers, Sat Nav, Smart TV apps (Welsh), live AI subtitling (Gaelic), broad corpus and labelling (Galician), parallel corpora of translation (from Galician), machine translation, voice recognition (Cornish), more models trained for machine translation and conversation management, text-to-speech, speech-to-text (Catalan and Aranès Occitan).

Question: What is your policy regarding the sharing of your LRs?

All respondents were unanimous in agreeing to share LRs. One respondent added that a policy of sharing "is the philosophy with which we currently approach the creation of new resources. However, previously created resources are in different legal situations. We cannot be sure of full openness in the resources that will be created in the future, but the intention is that they can be fully shared."

Question: What kind of access can be given to your LRs?

When asked what kind of access can be given to LRs, the organizations had a range choices.

- Seven responded they can be freely downloaded from the Web.

- Three said they can be used only for academic or non-commercial purposes.

- Two said they can only be viewed online.

- One said they can be acquired via a commercial license agreement.

One institute added that training data is available for download under open licence whenever possible and another pointed out that while offering free downloads this would also depend on the constraints of the format. One institute gave no answer.

Question: Do you have resources that have already been uploaded to CLARIN, ELG, META-SHARE, Hugging Face, Github or other similar online data platforms?

50% of respondents (4) stated that they do not have resources published to the online data platforms listed with one institute not replying. The figure indicates a marked difference on this question to that of the EU official language institutes where the majority have uploaded resources to one of the listed data platforms. Of those that had uploaded resources the following platforms were mentioned in the optional comment: ELRC, ELG, GitHub, HuggingFace, Zenodo. One institute stated that it had uploaded resources to the websites of partners.

5.2.3 Issues that prevent sharing of digital languages resources

Question: If your institution has digital LRss that are not publicly accessible, please, provide a list and if possible links.

The first question regarding issues that prevent sharing of digital LRs was for the respondents to write an open-ended response to what kind of resources they have that are not publicly available. The majority of respondents (5) did not answer this question. The 4 responses received were varied. One institute stated that only some books and texts were not publicly available, another that they shared all resources but this was not always reciprocated, another that all resources were publicly available, and one which stated that important running resources are publicly available.

Question: What are the reasons for not making these resources accessible?

When asked about the reasons for not making resources publicly accessible, the majority of respondents (6) gave no answer. The three respondents who did answer listed copyright issues, privacy issues and the lack of finance and personnel as the main reasons.

Question: What measures should be taken so that you could make the non-accessible resources fully or partially accessible?

When asked what measures should be taken so that an institute could make the non-accessible resources fully or partially accessible, most respondents (6) gave no answer. Of those that did one called for legal measures such as a change in copyright laws and two called for financial support.

Question: Describe possible other measures.

When asked to describe other possible measures that would be helpful, only two institutions responded. One institute proposed having: "Case studies of instances where opinion shifts have happened. E. g., YouTubers don't always seek permission to include copyrighted images within their videos. Sometimes they attribute source instead of seeking consent. Have there been positive effects of individual owners' contributions being used as training data where consent wasn't explicitly gained?" Another stated that they lack "Funding for language technology capacity to complete resources. We have Cornish language speakers with the expertise but they are of working age and need to be paid for their time."

Question: Would your organization be willing to give access to the non-accessible text data it holds for the purpose of creating a resource where the original text would not be recognizable or reconstructable?

The majority of respondents replied to this question with four saying "Yes", one saying that they were not aware if they had this data, and one saying that they would be open to discussion on the topic.

6 Consultations with funding agencies

In late 2022, a list of more than 100 national funding agencies was compiled and contact persons were identified as points of departure for communication about the SRIA (the ELE ELE Consortium (2022)) and consultations about the support for the ELE recommendations. However, since in parallel there have also been several initiatives by the EC, EU policy-makers, government agencies and ELE members regarding the Language EDIC and the European Language Data Space, it was decided not to go ahead with further contacts to funding agencies in order not to create confusion about overlapping initiatives. The list of funding agencies in the different countries is presented in Appendix C.

7 Conclusion

This deliverable explores the commitment of multiple stakeholder groups, policy makers, funding agencies and language institutes with the goal of establishing a joint large-scale programme to ensure DLE for all languages in Europe. From the consultation with policy-makers such as the European Parliament and the European Commission we have seen positive feedback with regard to the ELE initiative's findings and strategic recommendations but no concrete commitment concerning the financing and implementation of the ELE Programme. However, at the same time other developments have taken place in parallel, which include, among others, the funding and establishment of the European Language Data Space as well as the emerging Language European Digital Infrastructure Consortium (EDIC). It remains to be discussed how these various newly established initiatives can eventually work together and what the concrete synergies between them can actually be in the future.

The consultation with institutes for national languages (EFNIL and others) and minority languages (ELEN) had the purpose of promoting the ELE initiative and of identifying missing resources and issues that prevent the institutes from developing and sharing more resources. The consultation showed that the institutes primarily compile mono- and bilingual dictionaries, corpora and terminology resources, but also develop the tools that are needed to search and process these resources. Especially institutes for minority languages also produce technological resources such as machine translation and speech synthesis or voice recognition systems by themselves as these are generally not produced by tech companies.

All institutes that maintain resources are basically willing to share them for the benefit of their languages, and many have already shared whatever possible via ELG, CLARIN or other platforms. The main obstacle against sharing even more resources are legal issues such as copyright and GDPR. This is particularly the case for older resources that were compiled in times where the awareness of the usefulness of LRs for the development of LT and AI was not as high as it is today. Some of these restricted resources are available for research purposes, but not for public or industrial use. Another obstacle is the lack of experts with knowledge of both the language and the technology.

Many institutes would like to have access to more text data, audio data and historical texts and are asking for access to conversational AI and spoken dialogue systems to enable the speakers of their languages to participate in the digital age. For minority languages, we see a similar situation as for national languages where those which have a large number of speakers and a regional governmental structure that supports them, such as Catalan (10 million speakers), have more resources and more advanced technology than those with few speakers, few resources, and access to only basic technology, such as Cornish (1000 speakers).

There is a consensus that the most beneficial initiatives to improve the situation would be funding for especially less resourced languages, a workaround for the legal issues, and education of more researchers and developers with solid knowledge of these languages.

References

ELE Consortium. Deliverable 3.4 Digital Language Equality in Europe by 2030: Strategic Agenda and Roadmap, November 2022. URL https://european-language-equality.eu/wp-content/uploads/2022/11/ELE___Deliverable_D3_4__SRIIA_and_Roadmap__final_version_-1.pdf. ELE Deliverable D3.4.

Stefanie Hegele, Annika Grützner-Zahn, Katrin Marheinecke, Georg Rehm, Maria Giagkou, and Stelios Piperidis. Deliverable 1.1 Specification of approach for consultations and for documentation of stakeholder commitment, September 2022. URL https://european-language-equality.eu/wp-content/uploads/2022/10/ELE2___Deliverable_D1_1.pdf. ELE 2 Deliverable D1.1.

Georg Rehm and Andy Way, editors. *European Language Equality: A Strategic Agenda for Digital Language Equality*. Cognitive Technologies. Springer, June 2023. In print.

Appendix

A Recipients of the questionnaire - EFNIL member institutions

Table 4: Recipients of the questionnaire – EFNIL member institutions

Country	Institution	Contact
Austria	Austrian Centre for Digital Humanities, Austrian Academy of Sciences	Jutta Ransmayr
Austria	Österreichisches Sprachen-Kompetenz-Zentrum	Gunther Abuja
Belgium	Union for the Dutch Language	Kris Van de Poel
Belgium	Wallonia Brussels Federation	Aurore DUMONT
Bulgaria	Institute for Bulgarian Language, Bulgarian Academy of Sciences	Svetla Koeva
Croatia	Institute of Croatian Language and Linguistics	Kristina Despot
Croatia	Institute of Croatian Language and Linguistics	Zeljko Jozić
Croatia	Institute of Croatian Language and Linguistics	*
Czech Republic	Czech Language Institute of the CAS	Martin Prošek
Denmark	Danish Language Council	Ida Elisabeth Mørch
Denmark	Danish Language Council	Thomas Hestbæk Andersen
Denmark	Danish Language Council	*
Estonia	Estonian Language Council	Birute Klaas
Estonia	Institute of the Estonian Language	Avri Tavast
Estonia	Institute of the Estonian Language	*
Finland	Institute for the Languages of Finland	Aino Piehl
Finland	Institute for the Languages of Finland	Anna Maria Gustafsson
Finland	Institute for the Languages of Finland	Leena Nissilä
Finland	Institute for the Languages of Finland	Salli Kankaanpää
France	General Delegation for the French language and the languages of France	Claire-Lyse Chambron
France	General Delegation for the French language and the languages of France	Jean-François Baldi
France	General Delegation for the French language and the languages of France	Paul de Sinety
France	General Delegation for the French language and the languages of France	*
Georgia	Tbilisi State University, State Language Department	Maka Tetradze
Georgia	Tbilisi State University, State Language Department	*
Germany	German Academy for Language and Poetry	Jürgen Schiewe
Germany	Leibniz-Institute for the German Language	Andreas Witt
Germany	Leibniz-Institute for the German Language	Gerd Piroth

Continued on next page

Table 4 – continued from previous page

Country	Institution	Contact
Germany	Leibniz-Institute for the German Language	Gerhard Stickel
Germany	Leibniz-Institute for the German Language	Henning Lobin
Greece	Faculty of Philosophy of Aristotle University	Maria Arapopoulou
Greece	Centre for the Greek Language	Prof. Dimitri Koutsogiannis
Greece	Centre for the Greek Language	Vassiliki Dendrinou
Greece	Centre for the Greek Language	*
Hungary	Hungarian Research Centre for Linguistics	Gábor Prószéky
Hungary	Hungarian Research Centre for Linguistics	István Kenesei
Hungary	Hungarian Research Centre for Linguistics	Tamás Váradi
Iceland	Icelandic Language Council	Dr. Ármann Jakobsson
Iceland	Icelandic Language Council	Professor Angantýsson
Iceland	Icelandic Language Council	Steinunn Stefánsdóttir
Iceland	The Árni Magnússon Institute for Icelandic Studies	Ari Páll Kristinsson
Ireland	Foras na Gaeilge	Anna Davitt
Ireland	Foras na Gaeilge	Sean Ó Cearnaigh
Italy	Accademia della Crusca	Cecilia Robustelli
Italy	Accademia della Crusca	*
Italy	CNR The Italian Dictionary	Giulio Vaccaro
Italy	CNR The Italian Dictionary	Lino Leonardi
Italy	CNR The Italian Dictionary	Luca Barbieri
Italy	CNR The Italian Dictionary	*
Latvia	Latvian Language Agency	Jānis Valdmanis
Latvia	Latvian Language Institute	Ina Druviete
Lithuania	Institute of the Lithuanian Language	Albina Auksoriūtė
Lithuania	Institute of the Lithuanian Language	*
Lithuania	The State Commission of the Lithuanian Language	Audrys Antanaitis
Lithuania	The State Commission of the Lithuanian Language	Violeta Meiliūnaite
Lithuania	The State Commission of the Lithuanian Language	*
Luxembourg	Institut Grand-Ducal	Guy Berg
Luxembourg	Institut Grand-Ducal section de linguistique, d'ethnologie et d'onomastique	Amira-Louise Ouardalitou
Luxembourg	Institut Grand-Ducal section de linguistique, d'ethnologie et d'onomastique	Amira-Louise Ouardalitou
Luxembourg	Zenter fir d'Lëtzebuerger Sprooch (ZLS)	Luc Marteling

Table 4 – continued from previous page

Country	Institution	Contact
Luxembourg	Zenter fir d'Lëtzebuerger Sprooch (ZLS)	*
Malta	National Council for the Maltese Language	Ray Fabri
Malta	National Council for the Maltese Language	Thomas Pace
Malta	National Council for the Maltese Language	*
Netherlands	Dutch Language Institute	Frieda Steurs
Netherlands	Dutch Language Institute	*
Netherlands	Dutch Language Union	Annemieke Hoorntje
Netherlands	Dutch Language Union	Johan Van Hoorde
Netherlands	Dutch Language Union	Karlijn Waterman
Northern Macedonia	Krste Misirkov Institute of Macedonian Language	Jovanova-Grujovska
Norway	The Language Council of Norway	Nina Teigland
Norway	The Language Council of Norway	Åse Wetås
Poland	Council for the Polish Language, Polish Academy	Anna Dabrowska
Poland	Council for the Polish Language, Polish Academy	prof. Władysław Miodunka
Romania	Romanian Academy	Elena Tamba
Serbia	Serbian Language Institute	Jasna Vlajic Popovic
Slovakia	Slovak Academy of Science	Jana Levická
Slovakia	Slovak Academy of Science	Júlia Choleva
Slovakia	Slovak Academy of Science	Slavomir Ondrejovic
Slovenia	Fran Ramovš Institute of the Slovenian Language	Kozma Ahačič
Slovenia	Fran Ramovš Institute of the Slovenian Language	Nataša Gliha Komac
Slovenia	Slovenian Language Service	Irena Grahek
Slovenia	Slovenian Language Service	Magda Stražišar
Sweden	Language Council of Sweden	Harriet Kowalski
Sweden	Language Council of Sweden	Jennie Spetz
Sweden	Svenska Akademien	*
Switzerland	Institut de plurilinguisme – Institut für Mehrsprachigkeit	Susanne Obermayer
Ukraine	Office of the Ukrainian State Language Preservation Commissioner	Andrii Vitrenko
Ukraine	Office of the Ukrainian State Language Preservation Commissioner	Taras Kremin
United Kingdom	British Council	Ann Veitch
United Kingdom	British Council	John Simpson
United Kingdom	British Council	Maria Graczyk
	European Commission	Filip Majcen

Continued on next page

Country	Institution	Contact
	Table 4 – continued from previous page	
	European Commission	Javier Hernandez-Saseta

* In case of blanks, the email was sent to the main address of the institution

B Recipients of the questionnaire – other institutions

Table 5: Recipients of the questionnaire – other institutions

Country	Language	Institution	Type
Austria	Austrian German	Society for Austrian German	Language institute
Austria	Slovene	Rat der Kärntner Slowenen	Policy org.
Belgium	Dutch	Dutch royal academy for linguistics and litterature	Language institute
Belgium	Dutch	Variaties vzw Koepelorganisatie voor dialecten en oraal erfgoed in Vlaanderen	Policy org.
Belgium	Dutch	Ons Erfdeel (our heritage)	Policy org.
Czech Republic	Czech	The Institute of the Czech National Corpus	Language institute
Denmark	Faroese	The Faroese Language Committee	Language institute
Denmark	Greenlandic	Greenlandic language council	Language institute
Denmark	Danish	Society for danish language and literature	Language institute
Denmark	Danish	The Danish Language Society	Policy org.
Denmark	Faroese	University of Faroe Islands	Research institute
Estonia	Voro	The Võro institute	Language institute
Estonia	Estonian	NGO Estonian Association of the Deaf	Policy org.
Estonia	Estonian	Emakeele Selts (Mother Tongue Society)	Research institute
Estonia	Estonian	MTÜ Fenno-Ugria Asutus (NPO Fenno-Ugria)	Policy org.
Europe	All	European Centre for modern languages	Language institute
Finland	Norht, Inari, Skolt saami	The finnish sami parliment	Language/culture institute
Finland	Swedish	The Swedish Assembly of Finland, or Folktinget	Policy org.

Continued on next page

Table 5 – **continued from previous page**

Country	Language	Institution	Type
Germany	German	Goethe-Institut	Language/culture institute
Germany	German	Verein Deutsche Sprache (VDS)	Policy org.
Germany	German	Society for the German Language	Policy org.
Hungary	Hungarian	Int. Society of the Hungarian Language and Culture (Conf. on the Mother Tongue)	Language institute
Hungary	Hungarian	Institute for Hungarian Language Strategy	Language institute
Hungary	Hungarian	Service Office for the Hungarian Language	Language institute
Hungary	Hungarian	Association of Cultivators of the Mother Tongue	Policy org.
Hungary	Croatian	Institute of Science for Croats in Hungary	Research institute
Hungary	Slovak	Research Institute for the Slovaks in Hungary	Research institute
Hungary	Romanian	Institutul de Cercetări al Românilor din Ungaria = Research Institute for Romanians in Hungary	Research institute
Hungary	Hungarian	Termini Research Network for the Hungarian Language	Research institute
Iceland	Icelandic sign language	Language Council of Icelandic Sign Language	Language institute
Iceland	Icelandic sign language	The Communication Centre for The Deaf and Hard of Hearing	Language institute
Iceland	Nordic minority languages	The Vigdís International Centre for Multilingualism and Intercultural Understanding	Research institute
Latvia	Latvian	State Language Centre	Language institute
Lithuania	Lithuanian	Fellowship of the Lithuanian Language	Language institute
Lithuania	Lithuanian	Gathering of specialists of Lithuanian philology	Language institute
Netherlands	Frisian	Foundation Frisian Academy	Language institute
Netherlands	Limburgish	Council for the Limburgish language	Language institute
Netherlands	Dutch	Association Our Language	Language institute
Netherlands	Yiddish	Yiddish Foundation	Language institute
Netherlands	Dutch	International Association for Dutch Studies	Language institute
Norway	Kven	Kvensk institutt	Language institute
Norway	Sami	Sami parliment	Language institute
Norway	Norwegian	The National Team for Language Unity	Language institute
Norway	Norwegian nynorsk	Noregs Mållag	Language institute
Norway	Norwegian sign language	Norwegian Association of the Deaf	Language institute
Norway	Norwegian sign language	Foreningen for norsk tegnspråk (FONTS)	Language institute

Table 5 – continued from previous page

Country	Language	Institution	Type
Norway	Norwegian bokmål	Riksmåls society	Language institute
Portugal	Portugese	Comunidade dos Países de Língua Portuguesa	Language institute
Portugal	Portugese	Academia das Ciências de Lisboa	Language institute
Portugal	Portugese	Camões, Instituto da Cooperação e da Língua, IP	Language institute
Slovakia	Slovak	Slovenský národný korpus	Language institute
Slovakia	Hungarian	Gramma	Research institute
Slovakia	German	Karpatendeutscher Verein – Karpatskonemecký spolok na Slovensku	Research institute
Slovenia	Slovene	Centre for Slovene as a Second/Foreign Language	Language institute
Slovenia	Slovene	Office of the Government of the Republic of Slovenia for Slovenians Abroad	Policy org.
Slovenia	Slovene	Cultural Diversity and Human Rights Service	Language/culture institute
Slovenia	Slovene	Office for National Minorities	Policy org.
Slovenia	Slovene	Trojina, Institute for Applied Slovene Studies	Research institute
Slovenia	Slovene	Centre for language resources and technologies at the University of Ljubljana	Research institute
Slovenia	Slovene	Association Societies of Society for Slavic Studies of Slovenia	Research institute
Sweden	Sami	Sami parliment	Language institute
United Kingdom	Gaelic	Bòrd na Gàidhlig (Gaelic Board)	Language institute
United Kingdom	Welsh	Welsh language commissioner	Language institute

C List of funding agencies

WP1: SRIA: Further Consultations and Documentation of Stakeholder Commitment

D1.2: Report on consultations with funding agencies, policy makers etc.

Table 6: List of funding agencies

Country/Region	Organisation
Albania	Ministry of Education and sport
Armenia	Armenian National Science and Education Fund (ANSEF)
Armenia	Foundation for Armenian Science and Technology (FAST)
Austria	Austrian Research Promotion Agency (FFG)
Austria	Austrian Ministry for Climate Action, Environment, Energy, Mobility and Technology (BMK)
Austria	Austrain Science Fund (FWF)
Austria	Austrian Centre for Social Innovation (ZSI)
Azerbaijan	Science Development Foundation under the President of the Republic of Azerbajan
Basque Country	Basque Govt DG Language Policy
Belarus	National Academy of Sciences of Belarus
Belgium	Innoviris
Belgium	Research Foundation – Flanders (FWO)
Belgium	Flanders Innovation & Entrepreneurship (VLAIO)
Belgium	Fund for Scientific Research (FNRS)
Belgium	Research Wallonia Ministry for Public Service (SPW)
Bosnia	Federal Ministry of Education and Science
Brittany	Ofis ar Brezhoneg
Bulgaria	Bulgarian national Science Fund (BNSF)
Catalonia	Catalan Govt DG Language Policy
Croatia	Croatian Science Foundation (HRZZ)
Czechia	Czech Science Foundation (GACR)
Czechia	Technology Agency of the Czech Republic (TA CR)
Cyprus	Research and Innovation Foundation (RIF)
Cyprus	Research Promotion Foundation (RPF)
Denmark	Innovation Fund Denmark (IFD)
Denmark	Independent Research Fund Denmark (DFF)
Denmark	Danish National Research Foundation (DG)
Denmark	Ministry of Higher education and Science
Estonia	Estonian Research Council (ETAG)
Finland	Finnish Funding Agency for Technology and Innovation (TEKES)
Finland	Academy of Finland (AKA)

ELE

Table 6 – **continued from previous page**

Country	Institution
France	French National Research Agency (ANR)
France	EuroScience - European Association for the Advancement of Science and Technology
France	Core Technologies for Life Sciences
France	Institut Français de Recherche pour l'Exploitation de la Mer (Ifremer) - French Research Institute for Exploitation of the Sea
France	National Center for Scientific Research (CNRS)
Galicia	Galician Govt DG Language Policy
Georgia	Georgian Research and Development Foundation (GRDF)
Georgia	Shota Rustavelly National Science Foundation of Georgia
Germany	Project Management Jülich (PTJ)
Germany	German Aerospace Centre – Project Management Agency (DLR)
Germany	German research foundation (DFG)
Germany	Federal Ministry of Education and Research (BMBF)
Germany	Federal Ministry for Economic Affairs and Climate Action (BMWK)
Greece	Foundation for Research and Technology-Hellas (FORTH)
Greece	Ministry of Development and Investments
Greece	General Secretariat for Research and Development, Ministry of Development and Investments
Greece	Ministry of Digital Governance
Greece	Hellenic Foundation for Research and Innovation
Hungary	Hungarian Academy of Sciences (MTA)
Hungary	Hungarian Scientific Research Fund (OTKA)levente.emody@aok.pte.hu
Iceland	Rannis - Iceland Centre for Research
Ireland	Health Research Board (HRB)
Ireland	Irish Research Council (IRC)
Ireland	Science Foundation Ireland (SFI)
Italy	Ministry for Education, Universities and Research (MIUR)
Italy	National Research Council (CNR)
Lithuania	Research Council of Lithuania (LMT)
Latvia	State Education Development Agency Republic of Latvia (VIAA)
Latvia	Latvian Council of Science (LZP)
Luxembourg	Luxembourg National Research Fund (FNR)
Malta	Malta Council for Science and Technology (MCST)
Moldova	Moldovan Agency for Innovation and Technology Transfer (AITT)

Table 6 – **continued from previous page**

Country	Institution
Moldova	Moldovan Academy of Sciences
Montenegro	Ministry of Science and Technological Development
The Netherlands	The Dutch Research Council (NWO)
North Macedonia	Ministry of Education and Science
Norway	The Research Council of Norway (RCN)
Poland	National Science Centre (NCN)
Poland	Foundation for Polish Science (FNP)
Poland	Poland Ministry of Science & Higher Education (NAUKA)
Portugal	Portuguese national funding Agency for Science, Research and Technology (FCT)
Portugal	Regional Fund for Science and Technology oft he Azores Regional Government (FRCT)
Romania	Executive Agency for Higher Education, Research, Development and Innovation Funding (UEFISCDI)
Romania	The National Authority for Scientific Research (ANCS)
Romania	National Council for Scientific Research (CNCS)
Russia	Russian Foundation for Basic Research (RFBR)
Russia	Russian Science Foundation (RSF)
Serbia	Science Fund of the Republic of Serbia (SFRS)
Serbia	Serbian Academy of Sciences and Arts
Serbia	Ministry of Education, Science and Technological Development
Slovakia	Slovak Research and Development Agency (SRDA)
Slovenia	Slovenian Research Agency (ARRS)
Slovenia	Slovenian Science Foundation (SZF)
Spain	State Research Agency (AEI)
Spain	Ministry of Economy and Competitiveness (MINECO)
Spain	Carlos III Health Institute
Spain	Secretary of State for Research, Development and Innovation
Spain	Spanish National Research Council (CSIC)
Sweden	Vinnova
Sweden	”Forte - Swedish Research Council for Health, Working Life and Welfare
Sweden	Formas – The Swedish Research Council for Sustainable developement
Sweden	Swedish Energy Agency (SWEA)
Sweden	Vetenskapsrådet - The Swedish Research Council
Sweden	Swedish Foundation for Strategic Research (SSF)
Sweden	Swedish Foundation for International Cooperation in Research and Higher Education (STINT)

Table 6 – continued from previous page

Country	Institution
Switzerland	Swiss Naitonal Science Foundation (SNSF)
Turkey	The Scientific and Technological Research Council of Turkey (TÜBITAK)
Ukraine	National Research Foundation of Ukraine (NRFU)
United Kingdom	Economic and Social Research Council (ESRC)
United Kingdom	Arts and Humanities Research Council (AHRC)
United Kingdom	Innovate UK
United Kingdom	UK Research and Innovation (UKRI)

ELE

D1.2: Report on consultations with funding agencies, policy makers etc.

ELE

D Questionnaire sent to national and regional language institutes

Figure 11: Full questionnaire as published (page 1/7)

Questionnaire Digital Language Resources

4. Is your institution responsible for developing, updating and/or maintaining digital *
   language resources such as digital dictionaries, corpora of modern or historical
   language, grammars, semantic nets or other language descriptions?
   Dictionaries that are not online, but printed from digital sources are also
   relevant.

   Check one answer.

   ⃝ No *Continue to question 6*
   ⃝ Yes

5. Please, provide more information about which digital language resources your
   institution is responsible for.

   _____

   _____

   _____

   _____

   _____

   Language resources in the future

6. Is your institution planning to create more resources in the future?

   Check one answer.

   ⃝ Yes
   ⃝ No
   ⃝ Other: _____

7. Please, describe the resources that you are planning to create.

   _____

   _____

   _____

   _____

   _____

   2/8

Figure 12: Full questionnaire as published (page 2/7)

D1.2: Report on consultations with funding agencies, policy makers etc.

ELE

Questionnaire Digital Language Resources

8. Are there language resources that you would like to use but do not have access to (either because they are possessed by someone else or do not exist yet)?

Check one answer.

◯ Yes

◯ No

◯ Other _____

9. Please, describe the resources that you would like to be able to use.

_____

_____

_____

_____

_____

Policies for language resources

10. What is your policy regarding the sharing of your language resources?

Check one answer.

◯ We are willing to share all our resources

◯ We are willing to share some of our resources

◯ We are not willing to share any of our resources

◯ Other: _____

3/8

Figure 13: Full questionnaire as published (page 3/7)

Questionnaire Digital Language Resources

11. What kind of access can be given to your language resources?

Multiple choice question

☐ They can only be viewed online
☐ They can be freely downloaded from our webpage
☐ They can be used only for academic or non-commercial purposes
☐ They can be acquired via a commercial license agreement
☐ We do not give access
☐ Other: _____

12. Do you have resources that have already been uploaded to CLARIN, ELG Platform (European Language Grid), META-SHARE, Hugging Face, Github or other similar online data platforms?

Check one answer.

◯ Yes
◯ No
◯ Other: _____

13. Please specify on which platforms your digital language resources are available.

_____
_____
_____
_____
_____

Issues that prevent sharing of language resources

4/8

Figure 14: Full questionnaire as published (page 4/7)

Questionnaire Digital Language Resources

14. If your institution has digital language resources that are not publicly accessible, please, provide a list and if possible links (please note, that we will *not* share them with others).

If possible, add information about size, e.g. number of words, entries, or pages etc.

_____
_____
_____
_____
_____
_____

15. What are the reasons for not making these resources accessible?

Multiple choice question.

☐ Copyright issues
☐ Privacy concerns
☐ Lack of finance or personnel
☐ Other: _____

16. Please, describe what the issues are in more detail.

_____
_____
_____
_____
_____

5/8

Figure 15: Full questionnaire as published (page 5/7)

D1.2: Report on consultations with funding agencies, policy makers etc.

ELE

Questionnaire Digital Language Resources

17. What measures should be taken so that you could make the non-accessible resources fully or partially accessible?

Multiple choice question.

- ☐ Legal measures (e.g. changes in copyright laws)
- ☐ Technical measures (e.g. anonymization, scrambling of text lines)
- ☐ Financial support/funding
- ☐ Other: _____

18. Please, describe possible other measures.

_____

_____

_____

_____

_____

19. Would your organization be willing to give access to the non-accessible text data it holds for the purpose of creating a resource where the original text would not be recognizable or reconstructable, like a frequency list or a language model which consists of probability distribution over sequences of words in the text, but does not allow the reconstruction of the text itself.

Check one answer.

- ◯ Yes
- ◯ No
- ◯ Other: _____

Contact

Please, provide your contact information if you are willing to discuss these questions further or would like more information about how to share your language resources. We will then contact you.

Otherwise, please submit your answers at the bottom of this page.

6/8

Figure 16: Full questionnaire as published (page 6/7)

Questionnaire Digital Language Resources

20. Name:

_____

21. Role/Position:

_____

22. E-mail:

_____

23. Please write any comments or questions you may have here.

_____
_____
_____
_____
_____

Google Forms

7/8

Figure 17: Full questionnaire as published (page 7/7)