



the Digital Language Diversity Project

The Roadmap

Executive Summary

- » Many of the 6,000 to 7,000 languages spoken in the world today are in danger of becoming extinct.
- » In order to establish a sustainable policy for safeguarding and promoting European and global linguistic diversity, policies for digital development have to be embraced. With so much of our lives happening on the internet and through digital devices, the digital domain represents a context that cannot be ignored.
- » If we are to ensure European digital language diversity for future generations we need substantive European level policies and legislation that acts to protect that diversity. If not speakers of these languages are facing a digital time-bomb as with current trajectories of digital development they will be excluded from using their own language in an increasingly digitized society.
- » The European Union has 24 official languages and 60 lesser-used languages. While they should all enjoy the same status only the official languages are fully recognised. The potential of digital development presents us with the opportunity to bridge the digital divide, and to progress towards greater digital language equality.
- » DLDP research shows that digital development for lesser-used languages must be community driven and not driven by the interests of the LT industry and corporations.
- » Language equality and diversity are at the heart of the European project and the EU now has the opportunity to empower all citizens to be able to use their own languages digitally. With these factors in mind, and in order to ensure language equality for the digital future, the DLDP consortium makes a wide-ranging set of recommendations.
- » Key recommendations are for legislation to ensure that States develop digital services in lesser-used languages so as to deal with the impending digital timebomb; and the extension of the Digital Language Diversity Project to conduct a Europe-wide survey for the EU on the digital needs of all European lesser-used languages.

Imprint

The DLDP Roadmap

Authors:

Davyth Hicks, Paola Baroni, Klara Ceberio Berger, Antton Gurrutxaga Hernaiz, Eleonore Kruse, Valeria Quochi, Irene Russo, Tuomo Salonen, Anneli Sarhima, Claudia Soria

This work has been carried out in the framework of The Digital Language Diversity Project (www.dldp.eu), funded by the European Union under the Erasmus+ Programme (Grant Agreement no. 2015-1-IT02-KA204-015090)

© 2018

This work is licensed under a Creative Commons Attribution 4.0 International License.

Cover design: Eleonore Kruse

Disclaimer

This publication reflects only the authors' view and the Erasmus+ National Agency and the Commission are not responsible for any use that may be made of the information it contains.



www.dldp.eu



www.facebook.com/digitallanguagediversity



dldp@dldp.eu



www.twitter.com/dldproject



the
Digital Language
Diversity Project

Contents

Introduction 4

1.1 Why is there a need for digital development for lesser-used languages? 5

1.2 Promoting Digital Language Diversity 6

1.3 Existing digital language provision, the ‘low digital representation’ problem. 7

1.4 Bridging the technology gap between European languages 8

1.5 The provision of public and private sector services in RMLs. 9

1.5 The provision of public and private sector services in RMLs. 10

2.1 Language community priorities for digital development: the DLDP Survey 10

2.3 The DLDP Training Programme 13

2.4 The ‘Digital Language Survival Kit’. 14

2.5 Implementing the DLDP Survival Kit and Training Programme 14

Policy Recommendations 15

Current Initiatives 16

Road Map Timeline 17

Review and Reporting 19

Glossary

- DSM Digital Single Market
- RMLs Regional or minority languages
- STOA Science and Technology Options Assessment (European Parliament)
- LT Language Technology
- ELEN European Language Equality Network

List of Figures

Fig. 1: Use of Basque on the Internet. 11

Fig. 2: Use of Sardinian for electronic communication. 12



Introduction

The Digital Language Diversity Project, a three-year project funded under the Erasmus+ Programme, aims to address the problem of low digital representation of EU 'regional' and minoritised languages by giving their speakers the skills to create, share, and re-use online digital content. The availability of digital content and technical support to use it are essential pre-requisites for the development of language-based digital applications, which in turn can boost digital usage of these languages.

The mission of DLDP is to advance the sustainability of Europe's regional and minority languages in the digital world by empowering their speakers with the knowledge and abilities to create and share content on digital devices.

The DLDP consortium is led by the Consiglio Nazionale delle Ricerche ILC (Claudia Soria, with Irene Russo, Valeria Quochi, and Paola Baroni), Elhuyar (Klara Ceberio Berger, Antton Gurrutxaga Hernaiz), Johannes Gutenberg University, Mainz (Anneli Sarhimaa, Eleonore Kruse), European Language Equality Network – ELEN (Davyth Hicks), Karelian Language Society (Tuomo Salonen).

1.1 Why is there a need for digital development for lesser-used languages?

According to linguists, there are between 6,000 and 7,000 spoken languages and perhaps as many sign languages. Some countries are particularly diverse, for instance, Papua New Guinea (830 languages), Indonesia (722 languages), Nigeria (over 500 languages), India (22 official languages, 400 languages, and more than 4,000 dialects). These countries not only have the highest levels of linguistic diversity, but also have the highest levels of language endangerment.

About 97% of the world's people speak 4% of the world's languages; and conversely, about 96% of the world's languages are spoken by about 3% of the world's people (Bernard, 1996).

Europe in comparison has far fewer living languages with 255 languages being noted. Europe's linguistic diversity, with 3% of the world's living languages, suggests that managing European linguistic diversity should be a much simpler issue than in any other continent in the world.

Many of the 6,000 to 7,000 languages spoken in the world today are in danger of becoming extinct. These are languages "where there is a significant probability of dying within the lifetime of the current generations" (Baker, 2011)¹. Those with fewer than 10,000 speakers are especially vulnerable. About half or even more of the world's languages are in the situation that they do not regenerate themselves. These languages are usually not transmitted within the family and also there is little or no backup within the education system to ensure that new generations of children speak the languages. It is very possible that within the next hundred years that there will be fewer than 1000 languages spoken in the world.

With the development of today's European nation-states, an increasing emphasis was placed on developing a more uniform and often centralised system for public administration along with a 'national culture' through state policy. This model of the nation-state also promoted the concept that a state should have a common language and history. Where this was not present it was often created through a new policy for language and education which promote the state language to the exclusion of all other languages present within the territory of the state. This meant that non-state languages were either suppressed or ignored and led to the monolingual mindset prevalent in many states today. This suppression led to high levels of assimilation over a period of time and many regional or minority languages were undermined to a point of endangerment. This process gained pace as governments played an increasingly prevalent role in the daily life of nation states and the cultural and linguistic monopoly of the state became even stronger. However, many countries have seen a revival of interest in regional languages and cultures over the past century.

With increasing globalization the world's largest economies have opened their borders in favour of free trade and this trade has developed to include cultural as well as economic goods. Lifestyle, music, fashion and means of communication have become global and are often closely linked to global languages such as English and Spanish. State languages have also played an important role in this development. However, without state backing, regional and minority languages have found it difficult to find a foothold in this ever evolving scenario.

¹ Baker, Colin. 2011. Endangered languages: Planning and revitalization. In Foundations of bilingual education and bilingualism. 5th ed. Edited by Colin Baker: 40–63. Bristol, UK: Multilingual Matters.

Language preservation is not easy, both at a practical and emotional level. Some languages fall into a downward spiral through a lack of usage, its speakers become demoralised and begin to feel that their language has no status within their region or nation and that it has little social or economic value. Many regional languages continue to be discriminated against by States which implement language policies designed deliberately to eradicate them and to stigmatise their speakers. These policies have left many of Europe's regional languages in a position of endangerment and led to the breakdown in intergenerational transmission. While there have been some moderate successes in revitalization, with some languages holding their ground, the overall picture continues to be one of constant battles for basic provisions and rights.

Considering the scenario outlined above, digital development for lesser-used languages could not only bridge the digital divide and allow these languages to 'catch-up' with well-resourced languages, but may also act to help revitalization, reverse endangerment, and allow these languages to thrive.

1.2 Promoting Digital Language Diversity

In order to establish a sustainable policy for safeguarding and promoting linguistic diversity, policies for digital development have to be embraced. With so much of our lives happening on the internet and through digital devices, the digital domain represents a context that cannot be ignored. Speakers of major languages can access almost unlimited amounts of Web content, easily perform searches, interact, communicate through social media and voice-based applications. They can enjoy interactive e-books, have fun with word games for mobiles, engage in multi-player video games, or take advantage from innovative language learning facilities for other widely-spoken languages.

So-called "smaller" languages, however, do not enjoy the same range of opportunities. Welsh speakers were denied the publication of e-books in Welsh over Amazon's Kindle platform, because of lack of available Welsh electronic dictionaries. There is no Wikipedia for Mansi; speakers of Saami or Tongva have no localized interface for Facebook, and there is no Google translation for Sardinian, Igbo, or Breton. This inequality of digital opportunities further discriminates against minority languages, by relegating them once more to the realm of family communication and restricted topics. For Kornai, the vast majority (over 95%) of languages have already lost the capacity to ascend digitally².

Minority languages, instead, need to gain access to all contexts of life to be perceived as vibrant and fully usable languages. The presence of a language on the internet is of paramount importance for the impact it has on its speakers, especially the young generation. We must ensure, therefore, that the range of usage opportunities for all languages is increased and enlarged. Multilingualism cannot be truly and effectively enforced if all languages are not put in the conditions to act digitally. Empowering all languages, regional and minority ones in particular, with instruments that put them on a par with more widely-spoken languages, is a matter of equal digital opportunities for the speakers of those languages.

Digital language diversity not only needs to be recognized but needs to be upheld in law if we are to make any progress in protecting European and ultimately global language diversity. One of the key European values, forming primary EU Law, is the 'respect for linguistic and cultural diversity, while in 2,000 the EU Council

2 Kornai, A. (2013). Digital Language Death. PLoS ONE, 8(10).

and the European Parliament stated that: “All the European languages, in their spoken and written forms, are equal in value and dignity [...] and form an integral part of European cultures and civilisation”³

If we are to ensure European digital language diversity for future generations we need substantive European level policies and legislation that acts to protect that diversity. If not speakers of these languages are facing a digital time-bomb as with current trajectories of digital development they will be excluded from using their own language in an increasingly digitized society.

1.3 Existing digital language provision, the ‘low digital representation’ problem

Europe’s regional and minority languages (RMLs) are poorly represented digitally for a number of reasons. Firstly, many regional and minority are not officially recognised and rarely fully supported (e.g. ‘regional’ languages in France). Low prestige and a weak socio-political profile result in speakers turning to other languages when accessing the digital world.

The presence of RMLs in digital media and their usability through digital devices is usually limited to instances of digital activism and/or by means of cultural initiatives focused on the preservation of cultural heritage.

Another reason is the fact that virtually no European citizen is monolingual in a regional or minority language: everyone can always make use of an official EU language instead of a minority one, thus making EU regional and minority languages non-essential for communication purposes. European RMLs are of little economic interest for companies developing language-based digital applications, since virtually no prospective customer would be unable to communicate if these languages were not supported. As a consequence, the provision of state-of-the-art language-based applications, which would enable and foster their use over digital media and devices, is severely limited.

In addition, for a language to be used digitally, it has to be “digitally ready”, i.e. it must enjoy the range of tools and technical support available for other major languages. This is not always the case, see for instance the recent battle for the adoption of a keyboard that supports French regional languages.

The majority of EU RMLs are affected by the problem of weak technological support, with the notable exceptions of Basque, Catalan, Galician, Welsh and to a lesser extent, Frisian. The digital readiness of a language is inextricably linked to its digital presence: whenever a language is technologically supported and thus widely digitally usable, its digital representation expands. Digital data has become easily and readily available to be exploited to develop new and better applications, which in turn fosters even wider use.

This relationship between digital readiness and digital usability has become a vicious circle for RMLs in that the development of language-based applications crucially depends on the availability of large quantities of good quality open data (see Soria et al., 2014)⁴, but this data can only become available if RMLs are widely used digitally, and this requires the support of technology.

3 Decision No 1934/2000/EC of the European Parliament and of the Council <http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32000D1934&from=EN>

4 Soria, C., Calzolari, N., Monachini, M., Quochi, V., Bel, N., Choukri, K., Mariani, J., Odijk, J., and Piperidis, S. (2014). The language resource strategic agenda: the FlaReNet synthesis of community recommendations. *Language Resources and Evaluation*, 48(4):753–775.

The consequences are not only technological and not only short-term. It has been argued that the lack of digital usability represents a severe threat for languages. The research carried out by the Meta-Net network, which culminated in the publications of 30 Language White Papers⁵, illustrated how 29 European languages are at risk of “digital extinction” because of lack of sufficient support in terms of language technologies. The situation for other regional and minority languages is worse given their almost complete lack of technological support.

Since everyday life makes an increasingly extensive use of digital devices that involve language use, the usability of a language on digital devices indicates how a language is modern and relevant to current lifestyles and capable of dealing with all the demands of the 21st century.

1.4 Bridging the technology gap between European languages

In order to bridge the technology gap, policies should focus on fostering basic research, applied research and development and innovation for language technologies for all European languages. Research and technology transfer among languages, along with increased collaboration across linguistic communities must receive more attention. Funding schemes, such as Horizon 2020 and its successor funding programme, could boost knowledge and technology transfer between countries and languages that already have excellent research and innovation hubs in LT and those that do not; the goal would be to enable the less-resourced languages, including sign languages, to benefit from technologies already developed for other languages.

Europe is a multilingual society. All official languages of the European Union enjoy, at least according to the founding documents of the EU, equal status. Nevertheless, more than 20 Treaty languages and many RMLs are in danger of digital language endangerment. Language technologies may serve as a vehicle for the protection and promotion of smaller and non-official languages. At present, LT is primarily used in relation to national and large regional languages, partly due to the investment required. However, to ensure the historical cultural and linguistic diversity of Europe and to facilitate an active participation of all European citizens in our democratic processes, it is also important for the smaller languages in Europe to make use of language technologies.

Relevant surveys reveal striking differences in LT support among the different European languages. While there are good-quality software and resources available for a few languages and certain application areas, other, usually smaller, languages show substantial gaps. Digital support for 21 of the 30 languages investigated by Meta-Net, for example, was “non-existent” or “weak” at best.

Certainly, the amount and quality of technologies available to a certain language tends to correlate with the number of speakers of that language. Among other reasons, companies refrain from investing in the development of sophisticated language technologies for languages spoken by a small number of speakers, thereby deepening the gap between well-supported languages and under-resourced ones.

⁵ Rehm, G. et al., editors. (2012). META-NET White Paper Series: Europe's Languages in the Digital Age. Springer. <http://www.meta-net.eu/whitepapers/overview>

Moreover, not all countries have the required expertise or human resources necessary to support their languages. European policies must address this deficit.

Although much is said by the European institutions about the importance of linguistic diversity⁶, very few policy initiatives are undertaken and even less funding is provided to support European linguistic diversity. They should aim to highlight this deficiency and promote the need for more support for all indigenous languages of Europe to ensure that our rich landscape of languages, many of them highly endangered, thrive well into the future.

1.5 The provision of public and private sector services in RMLs

As more and more public service provision is digitized, speakers of European lesser-used languages will become increasingly excluded from using their language as the process of digitalisation continues. RML language rights are upheld by domestic legislation in most EU Member States, with the exception of France and Greece, and upheld by international Treaties such as the Council of Europe's European Charter for Regional or Minority Languages and the Framework Convention for National Minorities. However, none of these domestic or international laws are future-proofed so as to ensure that all European lesser-used languages are included in digital development that the more widely spoken languages will benefit from. Furthermore, an ongoing problem with the ECRML and FCNM, for example, is that many signatory states are failing to implement Treaty provisions coupled with long overdue reporting cycles, while, in some cases, domestic legislation is proving to be ineffective in combatting anti-regional language discrimination overall.

The lack of digital provision for RML speakers and their effective exclusion raises several serious human rights issues which the EU and its Member-States will be obliged to address. Areas of special concern are that of access to health service and local government services.

6 See, for example, the 2013 European Parliament Resolution on Endangered Languages in the EU, passed with a 93% majority but not implemented by the EU.
<http://www.europarl.europa.eu/sides/getDoc.do?type=TA&reference=P7-TA-2013-0350&language=EN>

1.5 The provision of public and private sector services in RMLs

The impact of digital exclusion on RML speakers with public and private sector services will be profound and will undermine citizen's rights. Studies from Canada for example (see Bowen, S., 2001)⁷ have indicated the negative impact on health service provision:

- » Barriers to participation in health promotion and prevention activities;
- » Barriers to initial access for most health services;
- » Increased risks of misdiagnosis;
- » Poorer patient understanding and adherence to prescribed treatment;
- » Lower patient satisfaction;
- » Poorer management of chronic disease;
- » Less effective pain management.

The negative impact of digital exclusion on how elderly RML speakers deal with increasingly digitized government services will be immeasurable. In addition, for many RML languages that do not have supporting language legislation there will be no obligations on governments to provide services in the RML. This will have severe impacts on how RML speakers access justice, taxation, and employment services, for example. The private sector may well be a worse scenario with no obligation or incentive to provide services in RMLs. From the EU perspective digital exclusion wholly undermines any notion of future protection of linguistic diversity and a multilingual digital single market.

The European Union has 24 official languages and 60 lesser used languages. While they should all enjoy the same status only the official languages are fully recognised. The potential of digital development presents us with the opportunity to achieve language equality, at least digitally.

2.1 Language community priorities for digital development: the DLDP Survey

The DLDP surveyed four languages, Breton, Sardinian, Basque and Karelian⁸. Each language survey gave different results in terms of digital requirements. Importantly the survey gives an indication of the actual language community requirements and underlines why a full survey of the digital needs of all European RMLs is needed.

Breton speakers have worked to get the digital basics in place, but there continues to be a clear need and demand for popular apps to be provided in Breton. If these are not provided it will leave the language less able to compete with the huge number of French apps and will therefore make the language appear less attractive to young speakers and learners. Regarding social media, there is some provision in social media in Breton, for example, a Breton interface version of Facebook. However, users complain of poor translation for this. Furthermore, Breton will need to have machine translation and inclusion on apps such as Google Translate. Machine translation is almost completely lacking except for the online translation of Breton to French offered by Ofis ar Brezhoneg. By far the most worrying feature is the lack of provision in healthcare and customer care despite Breton demand for it. Breton has an ageing population with most speakers being over the age of

7 Bowen S. (2001) 'Language Barriers in the Access to Health Care' www.canada.ca/en/health-canada/services/health-care-system/reports-publications/health-care-accessibility/language-barriers.html

8 See: <http://www.dldp.eu/en/content/reports-digital-language-diversity-europe>

65. Increasingly state health service provision is being digitalized but Bretons will not be able to access these services in their own language therefore compelling them to use French.

One example of this being a life threatening problem is that patients who have strokes often lose the ability to speak in their second language and can only speak in their mother tongue. Furthermore, in terms of rights, Breton-speaking taxpayers who have paid into the French health system all of their lives should have access to services in Breton. To a great extent Breton is a fairly well developed digitally, compared to other minoritised languages, and able, with care, to be further developed so as to prepare itself for increasing digitalization during the 21st century. However, there are various structural and legal issues that need to be addressed. So far, much of the development has been conducted voluntarily by a hugely under-resourced speech community relying on activists to make the provision that most small, medium and large state languages would be able to take for granted, and that are sponsored by the state itself. In addition, there is no legal base for the protection of Breton meaning that any future state-supported digital development is dependent on the goodwill of the government and that the government would be unlikely to provide public sector services in Breton as French remains the only official language in Brittany. It is this kind of unfair playing field that could lead to Breton being further marginalized digitally in the future leading to speakers being compelled to use French simply to access basic services. This scenario would further undermine Breton revitalisation and its use as an everyday language, and especially undermine the language amongst young people.

The DLDP also surveyed Basque. The results of the survey showed that Basque is a digitally 'fit' language and that it is actively used. The respondents have a high linguistic competence and a good knowledge of the existing digital tools and resources. There are a high amount of respondents who state that they use Basque regularly on the Internet (websites, e-communication, blogs, Wikipedia, etc.). This applies also to social media, especially to Facebook and Twitter, which shows significant activity in Basque. Being that these types of networks are mostly associated with

informal registers, this supports the vision of Basque as a language both used and useful for everyday spoken and written online communication, an undeniable sign of vitality. This is an important factor for the digital survival of a language. In contrast, Basque is less used in LinkedIn, the business and employment-oriented social networking service. This may reflect the situation of the Basque language, where use has grown in the familiar and informal settings, but where more work needs to be done within professional and formal contexts. Concerning localised digital services and interfaces for Basque, it is of note that the majority of the respondents know about the existence of Basque versions of the mentioned services. Nevertheless, despite knowing about their existence, some of the respondents are not using Basque in their devices, applications or software. There may be many reasons for that, but this survey gives us an insight into Basque speaker's digital habits considering that: nearly a third of respondents feel that using Spanish tools is easier; the way of finding and installing software in Basque is not as easy as it is in other languages, and, as a result of that, the user has to make an extra effort. Some of the respondents are calling for a site where all available resources in Basque are listed and ready to download for the users. In addition, it has been mentioned that there is

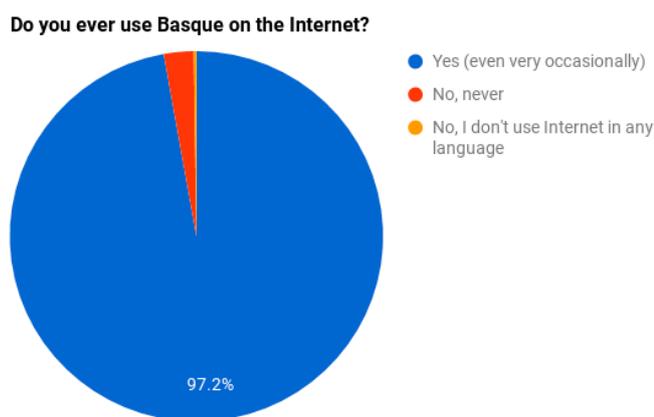


Fig. 1: Use of Basque on the Internet

the need to disseminate information about the existing services and interfaces in Basque, especially among young people. Finally, the survey indicated a demand for more entertainment products in Basque, especially for young people. Most people are using computer or mobile games in other languages because finding them in Basque is a bit harder for them.

Sardinian is extensively used online by the large majority of its speakers. 73% of survey participants declare the use of Sardinian everyday in electronic communication, in particular texting and chatting. The percentage reaches 87% for surfing, reading or writing on the Net. Sardinian appears to be particularly vital on social media, with Facebook being by far the most used network. The survey reports 143 different groups, pages or accounts that are accessed on a large scale. For Facebook there is also localised interface available. The use of Sardinian on social media reflects the fact that, despite official recognition for the language being granted by the Italian State in 1998, the language is relatively little used on public sites and that it continues to be a language mostly for private use. The vitality of Sardinian on social media does not

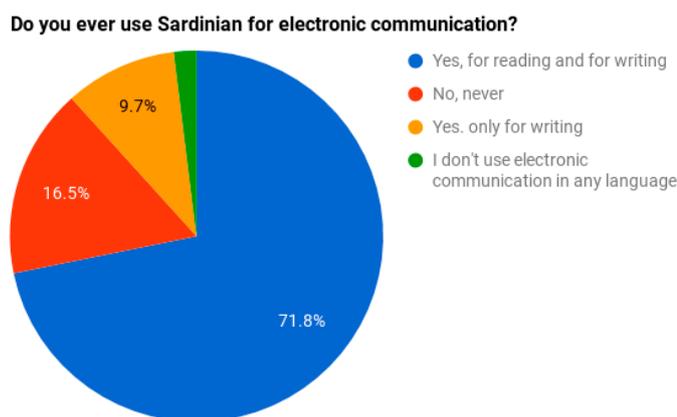


Fig. 2: Use of Sardinian for electronic communication

correspond, however, to a strong and diffused availability of Internet media. The existence and availability of digital services in Sardinian is not widely known. Online newspapers and news are widely available, as is entertainment and, thanks to a previous investment by the Region, some public administration services. More advanced media such as smartphone apps, Internet TV, streaming audio and video are lacking. People express a strong desire to be able to use Sardinian on the internet as part of their everyday life. It is worth noting that, in contrast to other regional languages, there are no technical obstacles for using Sardinian online. Typing is possible using the standard keyboard. Many respondents, however, declare a lack of competence in written Sardinian that prevents them writing from the language with the necessary confidence. The results clearly show the importance of encouraging the speaker community to use their language online as much as possible. The existence of a considerable number of language resources such as dictionaries, spell-checkers, and even an automatic translation system, is a good sign of the potential for this language to become a fully digital language, provided that the speakers are supported and encouraged to overcome the psychological barriers that continue to hold them back from considering Sardinian as a language to be used in all sociolinguistic domains.

The survey illustrates how Karelian's digital presence is developing, but still needs much work done. Speakers have a high linguistic self-esteem and this translates to a will to use the language online. Online presence is also particularly important for Karelian revitalization, since as a non-territorial language minority the speakers live far from each other in all parts of Finland. However, there are a lot of necessary online and digital resources missing, and speakers often have misleading or incorrect information about those that do exist. In terms of social networks the use of Karelian is very heavily restricted to Facebook. This could be because Facebook has users from all demographic groups and it is very well suited to extensive written communication, making it easy to connect with Karelians from different parts of the country as well as across the border. The situation is not likely to change considerably anytime soon, although the amount of people using Karelian on

Facebook may increase and there may be an increased Karelian presence on Twitter and Instagram, if more young people get interested in the language and start using it and if the Karelian revitalization efforts can get a better visibility in Finnish political discussion. The results clearly highlight the importance of informing the speaker community of the existing resources. The existence of online dictionaries and a Karelian Wikipedia could benefit many Karelian speakers and the community if only they were aware of the existence of such resources. The results also show that many people are unaware of the fact that they could easily change their keyboard settings so that they could write Karelian letters. This would be an important point to communicate to the community, since if writing Karelian is perceived as something that is difficult, it can easily demotivate people from using it.

2.2 Language community digital priorities

As expected the priorities of each different language community vary. To summarise, overall each language is digitally active and has made progress. Breton speakers emphasise more provision of public sector services, Basque speakers want more apps, Sardinians emphasised building confidence in writing the language, and Karelians, while using the language on Facebook, lack digital provision in all areas. Furthermore, the survey underlines the need for a complete survey of all European lesser-used languages so that we can build a complete picture of the digital provision required for our languages so as to enable them to be used fully in the digital field.

The next step is to develop the needs identified in the survey into substantive language policy and planning proposals that will form part of regular language planning for these languages. Furthermore, the survey needs to be extended to all European lesser used languages to assess their needs and to guide future digital development.

In addition, DLDP research shows that digital development for lesser-used languages must be community driven, that RML communities must be fully consulted, and not driven by the interests of the LT industry and corporations.

2.3 The DLDP Training Programme

As part of the DLD Project, the Training Programme has been created to support the long-term maintenance of Europe's regional and minority languages through fostering their digital presence. This innovative training programme is available to speakers of RMLs who want to learn why and how to increase the presence of his/her language online, and how to do this practically: it looks at which tools and techniques are available, which media are more suitable, and which aspects to address first. The programme is modular with each module addressing one or more particular level of Digital Capacity. The training programme, therefore, works to increase the digital survival capacity of a language.

2.4 The 'Digital Language Survival Kit'

The 'Digital Language Survival Kit' developed by the DLDP is an instrument allowing regional and minority languages speakers and communities to:

- a) Self-assess the degree of digital fitness of their language, by pinpointing current gaps and areas where action can and needs to be taken.

- b) Learn about what concrete actions and initiatives can be put in place depending on the particular digital fitness level identified. For instance, a minimal degree of digital fitness will require a level of "digital survival capacity" (in increasing order of necessity): ensuring connectivity; develop and adopt a standardized encoding; develop a standardized orthography, and some basic language resources, at least a corpus, a spell-checker, and a lexicon. Higher levels of digital fitness will require other types of measures, such as, for instance, creating or enriching a Wikipedia in the language, push for having localized version of important sites, main operating systems and social media interfaces. The 'kit' will serve as a companion to the training programme and will have a similar modular structure.

2.5 Implementing the DLDP Survival Kit and Training Programme

The DLDP Survival Kit and Training Programme will be disseminated to all European RML communities by the following means:

- » The DLDP will access the ELEN network and disseminate the Kit and Training Programme to all member organisations.
- » The DLDP will disseminate the Kit and Training Programme to all language policy departments of all governments that are host to RML communities.
- » ELEN will hold training courses with its members on how to use the Kit and Training Programme.

THE DLDP ROADMAP

Policy Recommendations

Language equality and diversity are at the heart of the European project and the EU now has the opportunity to empower all citizens to be able to use their own languages digitally. RML speakers are facing a time-bomb as with current trajectories of digital development they will be excluded from using their own language in an increasingly digitized society. With these factors in mind, and in order to ensure language equality for the digital future, the DLDP consortium makes the following Europe-wide recommendations:

- 1 Recommends European legislation that acts to ensure that lesser-used languages, including smaller state languages, have the digital provision required to function in the digital era. For example, if the Digital Single Market is to succeed citizens must have access to it in their own languages including in RMLs and smaller state languages. Not to do so undermines the project as well as raising numerous rights questions.
- 2 Recommends that current language legislation and international treaties protecting regional or minority languages have additional protocols added so as to ensure the digital provision of public sector services on RMLs and for digital RML development to keep pace with that of more widely used languages.
- 3 Recommends the extension of the Digital Language Diversity Project (DLDP) to assess all European RMLs. In order to collect data on RML digital needs so as to build the evidence base required to inform policy and planning, the DLDP needs to be extended to cover all European RMLs and smaller state languages that are lacking in digital provision. In order to facilitate this the EU could either establish a closed call for a project proposal or set up a Partnership with the DLDP consortium.
- 4 Recommends, further to 3 above, a Europe wide survey on the status of language technologies and resources for all European languages.
- 5 Recommends a large-scale, long-term research and development and innovation funding programme at European and State level, in order to develop innovative technologies and services, as well as contributing to the reduction of the technology gap between European languages.
- 6 Recommends the creation of a European language technology platform with representatives from all European languages that enables the sharing of language technology-related resources. Furthermore, promote technology transfer and sharing of resources and methodologies through the creation of cooperation programmes, including transnational programmes, between related languages.
- 7 Recommends that EU institutions raise awareness of the benefits for companies, public bodies, and citizens of the availability of online services, contents and products in lesser-used languages, for overcoming language barriers, as well as for the preservation of the cultural heritage of language communities.
- 8 Calls upon administrations at all levels to improve access to online services and information in lesser-used languages, and to use pre-existing language technology such as machine translation, speech recognition and text-to-speech, in order to improve the accessibility of those services.
- 9 Recommends that Social media and LT /ICT companies work with DLDP to ensure digital development for language diversity.
- 10 Recommends ensuring the production of open data, in the sense of data that is easily obtainable and can be used with few, if any, restrictions. Availability of open data is of paramount importance for the rapid development of software and technologies for languages for which data are relatively scarce.



Current Initiatives

There are two current (January 2018) initiatives in the European Parliament that will, if enacted, help to push the DLDP agenda for languages. Firstly, the STOA document Language Equality in the Digital Age¹, how this fits in with the DSM proposals, and mainstreaming the DLDP into that.

Secondly, the European Parliament INI Report Language Equality in the Digital Age, rapporteur Jill Evans MEP. ELEN and the DLDP have provided extensive input into the Report which calls for the DLDP to be extended to survey all European lesser used languages and to mainstreamed into the 'Human Language Project' proposals.

The two initiatives also fit in with the EU's overarching Digital Single Market² (DSM) initiative where ELEN has been arguing for the place of lesser-used languages in those proposals and underlining how developing RMLs and considering the language dimension will actually help the development of the DSM.

1 [http://www.europarl.europa.eu/RegData/etudes/STUD/2017/598621/EPRS_STU\(2017\)598621_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2017/598621/EPRS_STU(2017)598621_EN.pdf)

2 https://ec.europa.eu/commission/priorities/digital-single-market_en

THE DLDP ROADMAP

Road Map Timeline

Short term (2018-2020) measures with European and international institutions.



1

European Parliament (EP) initiative to support digital language diversity

Own initiative (INI) report with Jill Evans MEP as the CULTURE Committee rapporteur. This will commence in 2018. DLDP advisors have already been consulted on the Report and ELEN will ensure that the DLDP and ELEN will input into the early drafts of the Report. While the Report is only an own initiative Report and not a legislative report (i.e. it does not oblige the Commission to make it into EU law) it will, if successful, reflect the opinion and support of the EP. Given that the area is relatively uncontentious the EU may opt to implement it.

In addition, another wide-reaching initiative is the *Human Language Project* proposals as outlined in the EP's STOA document. If this succeeds it will be vital that the DLDP findings are taken on board and that RML digital development is embedded and prioritized within the project. (2018-2020)

2

DLDP results to MEPs/ Commission/ CoE. 2018

The DLDP project will be finishing in 2018. Before that the DLDP will have meetings with the relevant Commissioners (Ansip, Navracsics, Gabriel) in order to showcase the results of the project and to call for the project to be extended to all European RMLs. (2018-2019)

3

DLDP results to cross-cut into Digital Single Market debate, Geolocation, STOA Report

The DLDP outputs and results will be used by DLDP members to intervene into several ongoing debates in the EU including DSM, geolocation, national minority report, in order to mainstream the DLDP findings and proposals into the ongoing debate and proposals. (2018-2019)

4

Build on existing support from Member-States

Several member states with under-resourced languages have shown strong support for digital language developments such as the Baltic states and Slovenia. This support can be used to help gain traction for the DLDP proposals to be mainstreamed into the broader debate on digital development and provision.

5

Council of Europe

DLDP results and proposals will be presented to the COMEX of the ECRML and the Advisory Committee of the Framework Convention for National Minorities, and the Parliamentary Assembly of the Council of Europe. It will be highlighted how digital development for RMLs will aid better implementation of the Charter and FCNM. As with the EU, the DLDP will call for support for the project to be extended to cover all European RMLs. Call for monitoring and development of digital language provision for RMLs is added to the ECRML and FCNM as additional Protocols. (2018-2019).

6

United Nations

DLDP/ELEN will meet with the UN representatives from the Permanent Forum for Indigenous Peoples and Forum for Minorities to discuss how digital development benefits lesser used and indigenous languages, and to recommend adoption of the Training Programme and Survival Kit. (2018/2019)

7

UNESCO

DLDP/ELEN will meet with UNESCO to present the DLDP outputs and findings. In 2019 UNESCO will be launching the Year of Indigenous Languages with ELEN participating. Meetings will be arranged with the Knowledge Society Division to promote the DLDP with UNESCO. (2018/2019)



THE DLDP ROADMAP

Road Map Timeline

Short term (2018-2020) measures with European and international institutions.



1

Build awareness of DLDP results with each European RML community.

Via the ELEN network of RML organisations, the DLDP will work to disseminate project findings to all RML stakeholders in Europe including NGOs, universities, research centres, and governments.

2

Work to ensure that RML communities take action and adopt DLDP findings.

Once RML communities are made aware of the DLDP findings and proposals it will be important for each of them to apply the findings, using the Digital Language Survival Kit and training course, to their own language. Each community may wish to conduct its own survey to ascertain digital language priorities and develop grass-roots initiatives to aid development.

3

Translate RML communities support into national and European level initiatives.

Once RML communities have acted to implement the DLDP proposals it will be important for them to develop local and national level initiatives, such as RML digital language projects that will focus on developing their languages' digital capacity. They may also introduce legislation to help this process.

4

Encourage RML community to develop working links with major digital language developers.

DLDP will help RML communities develop their own links with the major digital providers eg Google, Microsoft, so as to ensure that the major companies work to ensure provision in their language. While ELEN will work for this at the political level it is more effective to have the actual language community engage with the major developers themselves to press their case.



Co-ordinate with Meta-Net and LT Innovate

Meta-Net issued its own detailed roadmap setting out the path for LT development. The road map is LT industry based whereas the DLDP roadmap is RML community based. Therefore, it is important that some coordination is planned in order for the LT industry to be made aware not just of the DLDP findings but of RML communities' digital needs, over and above the research Meta-Net has already conducted. Therefore, meetings will be held with Meta-Net and LT Innovate to establish a framework for regular dialogue between DLDP and ELEN on RML digital development. (2018/2019)

Co-ordinate with the LT industry on the Human Language Project

As mentioned above if the HLP is successful it will be vital for RML digital development to be prioritized in that project. It will require partnership with the LT industry and ongoing surveys of RML community digital requirements. (2020-)

1

2

THE DLDP ROADMAP

Road Map Timeline

Medium to Long Term measures with European and international institutions.



EU Directive to ensure digital language diversity.

A Directive will help to ensure that States and the EU support the development of digital provision for under-resourced languages. The proposal will act to complement the existing Digital Single Market initiative. A draft document will outline the case for a Directive in 2018 which will be presented to the Commission and MEPs.



Embed DLD recommendations with ICT/LT industry

Following discussions with DLDP and ELEN, the ICT/LT industry will work to ensure that DLDP recommendations are adopted.



Embed DLD recommendations with public sector organisations.

DLDP and ELEN will work to ensure that Member States, autonomous governments and local governments implement DLDP recommendations with regard to RML digital provision.



Ensure synergy between DLDP proposals and the new Donostia Protocol to Ensure Language Rights.

The Protocol¹ is the new standard-setting tool for the promotion and protection of RMLs with recommendations for ensuring LT development for RMLs.

Review and Reporting

The aim of the road map is for the normalisation of RML digital provision in the ICT/ LT industry where all European languages are included as well as in public sector digital provision.

In order to monitor progress the EU will need to establish a reporting mechanism, established in the Directive, whereby RML digital provision is monitored and supported in each Member State.

¹ See <http://protokoloa.eus/en/>