



**European Language
Resource Coordination**
Connecting Europe Facility

Deliverable D3.2.18

Task 8

ELRC Workshop Report for Cyprus



Author(s): Natassa Avraamides, Press and Information Office

Dissemination Level: Public

Version No.: <V1.0>

Date: 2019-01-08



Contents

1	Executive Summary	3
2	Workshop Agenda	4
3	Summary of Content of Sessions	6
3.1	Welcome by the EC	6
3.2	Welcome by the local organizer	6
3.3	Breaking down Language Barriers	6
3.4	The CEF eTranslation platform @ work	6
3.5	The European Language Resource Coordination (ELRC) action	7
3.6	National initiatives for digital public services	7
3.7	ELRC in Cyprus	7
3.8	Multilingualism and digital services in Cyprus	8
3.9	Legal Aspects in language data contribution	8
3.10	Tracking Greek Terminology	8
3.11	Identifying and managing your data: Questions & Answers	9
3.12	Conclusions	9
4	Synthesis of Workshop Discussions	10
4.1	ELRC and Open language Data in	10
4.2	Success stories and lessons learnt	10
5	Workshop Presentations	11

1 Executive Summary

The present document is a report regarding the 2nd ERLC Workshop which was held on 4th December 2018 at the EU-House in Nicosia (30, Byron Street, 1096). It includes the final agenda of the event and brief summaries of the sessions which included speeches, presentations and discussion. The event was attended by representatives of public services handling digital data and offering digital services to the public.

The workshop's highlight was Cyprus' position in a digitally connected multilingual Europe and its prospects through the lens of national strategies and actions for eGovernment, multilingualism and open data. Furthermore, there was discussion on how Cyprus can maximize the impact of EC initiatives. The contributions from the experts from the European Commission were particularly constructive, informative and gave an optimistic view of developments in the area of language digitization within the EU, whereas the statements made by senior representatives from the Cypriot public administration and government were promising towards identifying and collecting relevant sources of multilingual language resources in order to enhance the development of CEF eTranslation to the needs of national public services. The work of collecting language resources in Cyprus was taken one step ahead by creating a very positive environment and convincing the representatives about the significance of the work of ELRC in general, and more particularly with regard to the strengthening of the linguistic resources in the Greek language.

The workshop participants were from different governmental departments but also from non-governmental agencies in an effort to not only disseminate the message of the workshop but also to create a network of collaboration for a common cause. One of the main outcomes of the event is that the Press and Information Office of the Republic of Cyprus, has undertaken the role of acting as a national coordinator for open language data collection in Cyprus in cooperation with the Cyprus Open Data Portal and its network of public sector representatives from the national administration bodies.

The event was hosted in a dedicated webpage, http://www.lr-coordination.eu/l2cyprus_agenda, where the presentations are also available.

2 Workshop Agenda

- 08:00 – 09:00 **Registration**
- 09:00 – 09:10 **Welcome by the EC**
Thea Pieridou, Deputy Head of the European Commission Representation in Cyprus
- 09:10 – 09:15 **Welcome by the local organizer**
Sophie Michaelides, Director of Press and Information Office
- Session 1**
- 09:15 – 09:35 **Breaking down Language Barriers**
Christos Ellinides, DGT Deputy Director – General
- 09:35 – 09:50 **The CEF eTranslation platform @ work**
Micha Jellinghaus, eTranslation Engines Project Manager, DGT, EC
- 09:50 – 10:35 **The European Language Resource Coordination (ELRC) action: why and how you can engage and benefit**
Stelios Piperidis, Institute for Language and Speech Processing / Athena R.C., ELRC
- 10:35 – 10:45 **National initiatives for digital public services**
Sophie Michaelides, Director of Press and Information Office
- 10:45 – 11:00 **Coffee Break**
- Session 2**
- 11:00 – 11:20 **ELRC in Cyprus**
Fryni Kakoyianni-Doa – University of Cyprus - ELRC National Anchor Point (Technology)
Natassa Avraamides – PIO - ELRC National Anchor Point (Public Services)
- 11:20 – 12:20 **Multilingualism and digital services in Cyprus: an outlook into current and future challenges – Panel session**
Moderator: Natassa Avraamides, PIO, ELRC Public Services National Anchor Point
Panelists:
- *Katerina Pericleous, Director of Department of Information Technology Services – eGovernment – Ministry of Finance (represented by Charikleia Olympiou)*
 - *Lenia Orphanidou, Deputy Director of Department of Public Administration and Personnel, Open Data, Ministry of Finance (represented by Dimitris Michail)*
 - *Constantinos Karageorgis – Senior Officer at Service of Industry and Technology, Ministry of Energy, Commerce, Industry and Tourism*
 - *Andronikos Kakkouras, Senior Officer, Digital Communications, Ministry of Transport, Communications and Works*
- 12:20 – 13:20 **Lunch break**
- Session 3**
- 13:20 – 13:50 **Legal aspects in language data contribution**
Tatiana Synodinou, Associate Professor, University of Cyprus

ELRC Workshop Report for Cyprus

- 13:50 – 14:20 **Tracking Greek Terminology**
Ioannis Bersos, Terminology Coordination Unit, European Parliament
- 14:20 – 14:40 **Identifying and managing your data: Questions & Answers**
Moderator: Fryni Kakogianni-Doa
- 14:40 – 15:00 **Conclusions**
Moderator: Natassa Avraamides
- 15:00 – 15:30 *Coffee and networking*

3 Summary of Content of Sessions

3.1 Welcome by the EC

Deputy Head of the EC Representation in Cyprus, Ms Thea Pieridou, first gave a brief retrospect on the development of translation technology and digital commerce since 1998 and stressed the importance of digital commerce and the investment by gigantic companies such as Facebook, Microsoft in machine translation today. While stressing the importance of multilingualism in the EU, Ms Pieridou also underlined the need to prevent digital extinction of smaller languages, such as Greek, not only through learning, but also through eTranslation. Ms Pieridou also said that contributing language resources to the ELRC and to eTranslation will enable national public services and citizens in their digital transactions in their native language.

3.2 Welcome by the local organizer

Director of the Press and Information Office, Mrs. Sophie Michaelides addressed her welcome speech as the local organizer, stressing the need to find ways to promote multilingualism through cooperation between the public services of Cyprus. Ms Michaelides, in her welcome speech also read a message from the Minister of Interior, Mr. Constantinos Petrides. In his message, Mr. Petrides conveyed his support in the work of the EC for the digitization of languages and public services and stressed the importance of the Press and Information Office as the national coordinator in the process of language resources collection.

3.3 Breaking down Language Barriers

Deputy Director of the DGT, Mr. Christos Ellinides, stressed the importance of breaking down language barriers within the EU and the EC's commitment in supporting the cooperation with the public services in order to promote the Greek language in the digital age. Mr. Ellinides referred to the significance of good communication between the EU institutions at all levels of society in the member states, especially at a time when a strengthening of "euroscepticism" is sensed, and that this communication can only be achieved through languages. Mr. Ellinides also mentioned that the gap between widely used languages and languages that are less spoken must be bridged especially in matters of technology. Mr. Ellinides said that when public services across Europe can communicate without any language barriers, this would be useful for business and everyday life and this is the meaning of the single market. Until today, he said, the language resources in Greek are second in line of success among 36 data collections exchanged within the ELRC, with only 3 of them originating from Cyprus (sent by the Press and Information Office). Mr. Ellinides also mentioned the upgraded eTranslation platform and encouraged public services to develop multilingual information services allowing citizens to benefit from eTranslation with the use of a safe and high quality system. Finally, Mr. Ellinides referred to the EC's determination to render Europe multilingual and more competitive and expressed the wish to have the support of the public services in safeguarding multilingualism and in ensuring the digital presence of the Greek language and Cypriot terminology in the digital world.

3.4 The CEF eTranslation platform @ work

Mr. Micha Jellinghaus, eTranslation Engines Project Manager, DGT, elaborated on the use of the eTranslation platform, its availability, users and benefits and gave practical examples of its actual use. Furthermore, Mr. Jellinghaus underlined the safety of the eTranslation service and its privacy as documents are deleted after 24 hours or upon demand immediately after delivery. He also explained the use of the eTranslation API in machine to machine scenarios, providing the examples of the N-Lex and the European Data portals. Mr. Jellinghaus also went on to describe in more detail the service provided saying that it performs best when trained on large volumes of text pairs (source-translated) in specific domains. Finally, Mr. Jellinghaus talked about Neural Machine Translation and explained how artificial neural networks are trained. He also said that

ELRC Workshop Report for Cyprus

the key to success for eTranslation are more language resources for all languages with better lexical coverage. Mr. Jellinghaus also underlined the need for future improvements such as: speed, transliteration, more formats (PDF output, JSON), named-entity recognition and new languages (AR, RU, TR, ZH).

3.5 The European Language Resource Coordination (ELRC) action

Mr. Stelios Piperides, from the ILSP/Athena R.C., ELRC, gave an overview of the ELRC Consortium and of the aims and objectives of the ELRC action, i.e. collecting language resources; identifying public services in need of multilingual functionalities; engaging the public sector in the identification and continuous sharing of such language resources; helping with legal and technical issues associated with the collection and/or provision of language resources; acting as observatory for language resources across all EU Member States and CEF-affiliated countries. Mr. Piperidis explained that the ultimate goal of ELRC is to help enhance the performance of eTranslation by collecting language resources (such as bilingual corpora, multi-lingual corpora, terminologies, etc.) that are used as “training data”. In-domain training data (i.e. translations from the target domain) are essential for achieving high-quality translation. Since the eTranslation system is reused by some of the CEF DSIs, the priority domains are consumers’ rights, social security, justice, commerce and business, culture etc. He further elaborated the types of data that are useful for training eTranslation, i.e. any piece of text in a natural language and its equivalent in another language. Data residing in local public organisations, produced in-house or outsourced, e.g. reports, communication, news, web content, policies, terminologies, archives, forms, and FAQs are useful for improving the platform. He also presented a short walkthrough to the ELRC-SHARE repository, i.e. how to register, upload and describe language resources. Mr. Piperidis provided examples of processing services applied to the contributed data in order to convert them to MT-ready training datasets and invited the participants to use the freely available ELRC services, i.e. technical and legal helpdesk and on-site assistance.

3.6 National initiatives for digital public services

The Director of the Press and Information Office, Mrs. Sophie Michaelides, said that CEF eTranslation is a significant tool in making digital public services multilingual and contributing to “Language Equality in the digital era”. Mrs. Michaelides referred to the European Parliament vote of 11/9/2018 on language equality which stresses the importance of promoting the digitization of smaller and minority languages in order to avoid digital extinction. Mrs. Michaelides said that the second ELRC workshop aims to raise more awareness about the ELRC work within the public sector and create a network of partners to help promote the collection of more language resources. Furthermore, Mrs. Michaelides said that the Press and Information Office shall serve as a coordinator between the public services in order to facilitate best practices in the collection of language data in Cyprus. Ending her speech, Mrs. Michaelides also expressed the will that the PIO assist in any possible way host the next ELRC Conference in Cyprus. (Mrs. Michaelides full speech (in Greek) is available [online](#)).

3.7 ELRC in Cyprus

The Technology NAP, Mrs. Fryni Kakoyianni welcomed the participants and said that the aim of the workshop is to continue the discussion on collecting and managing of linguistic data in the public services and present related legal matters and also technical issues which shall be of importance when the language data volume has increased.

ELRC Workshop Report for Cyprus

The Public Services NAP, Mrs. Natassa Avraamides stressed the importance of the sustainability of this project of language resources and that it is equally important to maintain the network of contacts within the public services. She also referred to the fact that the Press and Information Office has until today contributed around 400.000 words to the ELRC repository.

3.8 Multilingualism and digital services in Cyprus

The panel presentations were moderated by Natassa Avraamides NAP for Public Services. The speakers came from departments of ministries that handle digital services (open data, eGovernment, digital champion) and gave presentations on the work done in their departments and the digital services offered.

The eGovernment representative stressed the need to include languages and translation in the process of restructuring and modernizing the public sector and facing new challenges, whereby special mention was made to the electronic platform ARIADNE, through which citizens can conduct their transactions with the public administrations electronically on a 24/7 basis.

The need to use machine translation as a tool for enabling multilingualism services, such as eCommerce and eGovernment was stressed out. The Open Data representative and the Digital Champion additionally underlined the need for infrastructure update in the public sector, which is considered a priority for efficient eGovernment and modernization of the administration.

It was made clear by all speakers that a plethora of language data can be identified within their services with both monolingual and bilingual content which can be a useful contribution to the whole ELRC project. However, this brought up the need for coordinated action, i.e. to have a designated public administration agent serving as a coordinator between the public services in order to facilitate best practices in the collection of language data in Cyprus. Furthermore, the Deputy Director of the Department of Public Administration and Personnel, and Open Data, expressed the will to support the project of language resources and cooperate closely with the Press and Information Office in this direction.

3.9 Legal Aspects in language data contribution

Ms Tatiana Synodinou presented the legal framework for the further use of public sector data and the EU Directive 2013/37/EE) and its application in the Cyprus legal system, the different categories of public data, the exceptions in the reuse of data as well as the contribution of data and the protection of personal data. Mrs. Synodinou made reference to the Opinion 06/2013 regarding the further use of open data and information of the public sector and also to the Guidelines on recommended standard licences, datasets and charging for the reuse of documents (2014/C 240/01).

3.10 Tracking Greek Terminology

Mr. Ioannis Bersos, from the Terminology Coordination Unit of the European Parliament referred to TermCoord, its work and the challenges it faced, namely: to highlight the importance of terminology during the translation process, and to improve the cooperation between the EU Institutions. He also referred to EUR Term, the cooperation with academic bodies and to IATE and its latest version which was introduced in 2018. As regards the contribution of language data from Cyprus these should be of major importance in enhancing the Greek language resources in terms of different terminology used in the Cypriot administrative system.

ELRC Workshop Report for Cyprus

3.11 Identifying and managing your data: Questions & Answers

In this section which was moderated by the Technology NAP, Mrs. Fryni Kakoyianni-Doa, the participants were given the floor to ask questions relevant to the procedure, methodology or other technical matters regarding the identification and contribution of language data for the purposes of the ERLC and eTranslation.

The participants raised the issue of existence or non-existence of multilingual data in their departments and the need to start a contribution in this direction. Another topic was how to identify data and the form in which these should be, in order to correspond to the ELRC aims. Furthermore, the participants talked about the importance to continue this discussion and their willingness to contribute.

3.12 Conclusions

In this session the Public Services NAP gave an overview of the Workshop which was about presenting language data, where and how to identify and collect them, the forms in which these are to be found, what are open data, which public bodies produce, have or use language data in monolingual or bilingual form and also the new generation of automated translation “eTranslation” and how it can be improved to serve the needs of the national public administrations and the EU citizens in general. Mrs. Avraamides underlined the need to maintain a network of contacts so that this project can be sustained, with the main goal to preserve the Greek language in the digital age.

4 Synthesis of Workshop Discussions

The main issues extracted from the participation in the discussion were mainly the type of data needed, in which form and how to handle these. The main concerns of the Cypriot participants regarding language data was their identification, i.e. what is actually expected, in which form, in how many languages, whether also monolingual data are appropriate or can be accepted. They also expressed lack of awareness as regards language data sharing. Some questions were asked on the legal constraints regarding personal data and their anonymization. It was also mentioned that the work of language data collection should better address the Cypriot administrative jargon.

4.1 ELRC and Open language Data in

The open data policy in Cyprus is based on the PSI directive (2003/98/EC) which was implemented in 2006. In 2015 the law N.205(I)/2015 was passed based on the reviewed Directive. In November 2018 the portal data.gov.cy was rated as the best open data portal in the EU. The Open data portal shall start a cooperation with the PIO for the sharing of open language data. This cooperation will start very soon and we shall discuss the types of documents to be shared between the two organizations. To facilitate sharing of language data, the Cyprus Open Data Portal network of approximately 140 representatives from the public sector will be exploited.

4.2 Success stories and lessons learnt

- The prompt contribution of language data to the ELRC depository within six months after the PIO's involvement in the action is an important fact and a significant stepping stone. Also important is the fact that the PIO serves as the Official Translation unit for the entire public sector and can contribute large volumes of bilingual data to the repository. By establishing a sustainable network of contact persons throughout the public sector we create a best practice example.
- The ELRC could consider setting up a direct communication channel from which it can draw officially acquired language data from the participant countries. Setting up a centralized dedicated direct channel in each country could be easier and faster in the collection and dispatch of data directly to the ELRC, rather than expecting from each organization separately to set up their own language data management operations for this purpose.

5 Workshop Presentations

All presentations are available online on the ELRC website: http://www.lr-coordination.eu/l2cyprus_agenda.