



**European Language
Resource Coordination**
Connecting Europe Facility

Deliverable D3.2.8 Task 8

ELRC Workshop Report for Slovakia



Author(s):	Miroslav Zumrík
Dissemination Level:	Public
Version No.:	<V1.0>
Date:	2018-06-22



Contents

Contents	2
1 Executive Summary	3
2 Workshop Agenda	4
3 Summary of Content of Sessions	6
3.1 Welcome and introduction	6
3.2 Welcome by the EC	6
3.3 Connecting public services across Europe: ambition and results so far	6
3.4 National initiatives for digital public services and (open) data	6
3.5 CEF in Slovakia: an outlook into current and future challenges – Panel session	7
3.6 The CEF eTranslation platform @ work	9
3.7 The European Language Resource Coordination (ELRC) action	9
3.8 ELRC in Slovakia	9
3.9 Can language data be shared and how?	10
3.10 Preparing and sharing data with the ELRC repository – and what happens next	10
3.11 Identifying and managing your data: Questions & Answers	10
3.12 Conclusions	10
4 Synthesis of Workshop Discussions	12
4.1 ELRC and Open language Data in Slovakia	12
4.2 Success stories and lessons learnt	12

1 Executive Summary

This document reports on the second ELRC Workshop in Slovakia, which took place on May 30, 2018 at the Representation of the European Commission in Bratislava, Slovakia.

The goal of the event was to raise awareness both about the ELRC action, the eTranslation service and the CEF mechanism in general, and to link the European context to the Slovak national strategy for informatisation of public services, Slovak national open data portals and related legal and technical issues. The ongoing and rather complex informatisation in Slovakia is mainly coordinated by the Deputy Prime Minister's Office for Investments and Informatisation, while the national open data portals are maintained by the National Agency for Network and Electronic Services (NASES). The EESSI system in Slovakia is being implemented by the Social Insurance Agency. The workshop had representatives for all these crucial stakeholders. The workshop showed that the Slovak public bodies could profit from the massive CEF Telecom funding in the near future, be it with regards to better data management and/or better and multilingual public services. The participants were interested in the eTranslation service, which they had been mostly unaware of until then. The eTranslation has meanwhile proven useful for health/social digital services, as well as for the development of eProcurement. We can conclude that even though there still is lack of awareness about the possibilities of EC funding mechanism amongst Slovak public bodies, and even though full digitization of (multilingual) Slovak public services is far from complete, Slovakia has already done some major steps to achieve it and should continue to do so in the future.

The workshop was attended by representatives mainly from Slovak public bodies, but also from academia, translation agencies and other LT companies.

The remainder of this report includes the agenda of the event (section 2) and briefly informs about the content of each individual, interactive and panel workshop session (sections 3 & 4). The dedicated event page can be found at: <http://www.lr-coordination.eu/l2slovakia>.

2 Workshop Agenda

08:30 – 09:00 **Registration**

09:00 – 09:10 **Welcome and introduction**

Miroslav Zumrík, *ELRC Technology Anchor Point*

09:10 – 09:15 **Welcome by the EC**

Emília Andrejová, Representation of the EC in Slovakia

Session 1. Connecting a multilingual Europe: European context and local needs

Connecting public services across Europe: ambition and results so far

09:15 – 09:35 **Aleksandra Wesołowska**, Project Officer, Directorate-General Communications Networks, Content and Technology, European Commission (live video link, interpretation in Slovak)

National initiatives for digital public services and (open) data

09:35 – 09:55 **Martina Slabejová**, Deputy Prime Minister's Office for Investments and Informatisation of the Slovak Republic

CEF in Slovakia: an outlook into current and future challenges – Panel session

Moderator: Jana Levická

09:55 – 10:40

- Michal Ohrablo (UPVII)
- Marek Vandák (ANASOFT APR)
- Jakub Začka (Social Insurance Agency)
- Vladimír Bednár (NASES)

The CEF eTranslation platform @ work

10:40 – 11:00 **Markus Foti**, [MT@EC](#)/eTranslation Project Manager, Directorate-General for Translation, European Commission (live video link, interpretation in Slovak)

11:00 – 11:30 *Coffee Break*

Session 2. Engage: hands-on data

11:30 – 11:55 **The European Language Resource Coordination (ELRC) action**

Stelios Piperidis, ILSP/"Athena" R.C., ELRC

11:55 – 12:15 **ELRC in Slovakia**

Miroslav Zumrík, ELRC Technology Anchor Point

12:15 – 12:40 **Can language data be shared and how? National and European legal framework**

Ronald Strehár, Ministry of Finance of the Slovak Republic

12:40 – 13:40 *Lunch Break*

13:40 – 14:15 **Preparing and sharing data with the ELRC repository – and what happens next**

Maria Giagkou, ILSP/"Athena" R.C., ELRC

14:15 – 14:45 **Identifying and managing your data: Questions & Answers**

European Language Resource Coordination (ELRC) is a service contract operating under the EU's Connecting Europe Facility SMART 2015/1091 programme.

Miroslav Zumrík, ELRC Technology Anchor Point

14:45 – 15:00

Conclusions

Miroslav Zumrík, ELRC Technology Anchor Point

15:00 – 15:30

Coffee Break

3 Summary of Content of Sessions

3.1 Welcome and introduction

As Technology NAP for Slovakia, Miroslav Zumrík from the Ľudovít Štúr Institute of Linguistics, Slovak Academy of Sciences, welcomed the participants of the second Slovak ELRC workshop by saying that he was glad the event was again held at the Slovak EC Representation in Bratislava, with some known and some new participants. This, as he said, could be seen as an expression for both continuation and renewal of the ELRC action. He then expressed his thanks to all who made the workshop possible and sketched the basic message, as well as an outline of the agenda ahead.

3.2 Welcome by the EC

As representative for the European Commission, the Slovak DGT Field Officer at EC Representation, Mrs. Emília Andrejová, welcomed the participants by stressing that 2018 is proclaimed the European Year of Cultural Heritage. While in the beginning there were 4 EU languages, nowadays with 24 official languages we face 46 times more language combinations. Every year, more texts need to be translated and they need to be translated ever faster, so the automatic translation becomes a necessity. Therefore, the European Commission makes long term investments into the new technologies. In her affiliation with the EC (since 2007), Mrs. Andrejová has observed that the number of software solutions increases every year, which makes translation terminology checks easier. In 2017 alone, 2 million pages were translated, 300.000 of them last month. The amount of pages that need to be translated per day is thus between 13 and 20. Mrs. Andrejová concluded that if we want to succeed in building a functioning Digital Single Market, we need to improve services like eTranslation that would serve all authorized users, and in order to do that, we need language resources to train the translation systems.

3.3 Connecting public services across Europe: ambition and results so far

A live video presentation from Aleksandra Wesolowska (DG CONNECT) was transmitted to the audience with live interpretation into Slovak. Mrs. Wesolowska presented the Connecting Europe Facility, with special emphasis on the Digital Service Infrastructures (DSIs) and particularly the CEF Automated Translation building block. She concluded with the need for the involvement and connection of the national public administrations with eTranslation and the current funding opportunities.

3.4 National initiatives for digital public services and (open) data

After the presentation of Aleksandra Wesolowska, Mrs. Slabejová from the Deputy Prime Minister's Office for Investments and Informatisation (UPVII) gave a more local-oriented view on how the digitalization and informatisation of Slovak public sector proceeds. Mrs. Slabejová aimed at presenting the ongoing cooperation with European institutions. She started her presentation with the question how many of the attendees plan to submit a proposal for the CEF Telecom Call in September: there were only three participants' hands raised. Mrs. Slabejová used this as an example of how little Slovakia uses the cohesion funds offered by EU, which also has its consequences, like significant budget cuts (approx. 1,3 billion euros), compared with the previous period. Mrs. Slabejová then focused on possibilities of Slovakia in this area. Since 2016, she has been working as a Slovak Digital Leader, which is a step towards centralization of informatisation

in Slovakia. Mrs. Slabejová also presented operational programmes created by the Slovak government which shall implement digitalization of public administration infrastructures, lead to a functioning business sector and society and help building the Digital Single Market. Mrs. Slabejová stressed some basic priorities at the Deputy Prime Minister's Office for Investments and Informatisation: value for money, professional and transparent approach. All relevant strategic documents are made available to those interested. Mrs. Slabejová focused on the topic of data management, which is inevitable for building eGovernment services, and that, in turn, requires good (cleaned, structured, categorized) data. Several initiatives have been made with the aim of cleaning public data and connecting public bodies (this project started in December 2017), developing management of personal details (since April 2018) and modernization of national public administration portal, with "client portfolios", which are recordings of interactions between state and citizens in compliance with the GDPR regulation. Mrs. Andrejová admitted that this would be "a long-term run". There is also a call for projects aimed at the public administrations in need of cleaning their data, which could then be sent to a centralised system. Another operational programme has been mentioned, dealing with making public administration more effective. Mrs. Slabejová said that there are approximately 100 projects that are crucial for the Deputy Prime Minister's Office's, whose aim is to offer better e-services.

Regarding the issue of Open Data, Mrs. Slabejová pointed out that the position of Slovakia in this aspect is relatively good (11th place in the EU), and went on to present the national open data portal, Data.Gov.Sk, which serves public bodies and citizens. Slovakia has also declared support to the Free Flow of Data initiative. Mrs. Slabejová emphasized that collecting open data has to harmonise with raising awareness of how public bodies could use these data, and presented two recent legal propositions/drafts, the first one regulating information systems within public administrations, the second one regulating data. The second document would also relate to current trends and initiatives in data management, like the AI Declaration, which Slovakia has supported since April this year. There are possibilities of using machine learning in ordering and clearing data at Slovak public bodies, which would make predictability of future developments in society easier. Mrs. Slabejová has also mentioned Block Chain Technologies, currently used by the Slovak NASES Agency, and the so-called Behavioural Office, focused on raising behavioural innovations and digital skills of the public. Mrs. Slabejová mentioned 37 partners working with this issue. Then there are projects of integrating Slovak public bodies into the governmental cloud (with both private and public part), whereas the public part could store open data material, usable in eTranslation training procedures. Another project focuses on how to bring small improvements into eGovernment services, and the third project is dealing with clearing the data. Mrs. Slabejová repeatedly encouraged the audience not to ignore the ongoing CEF calls and apply for mutually profitable CEF projects.

3.5 CEF in Slovakia: an outlook into current and future challenges – Panel session

In the panel session, Jana Levická from the Ľudovít Štúr Institute of Linguistics welcomed the four participants. The aim of the panel was to both present the respective institution of the participant and to discuss the ways in which CEF programme is being deployed in Slovakia.

Vladimír Bednár has presented NASES, the National Agency for Network and Electronic Service. This agency maintains two main national open data portals, Slovensko.sk and Data.Gov.sk. E-services provided by Slovensko.sk are quite complex and aim to address the needs of citizens in all major life situations. As of today, the number of e-services is around 1700, whereas by December 2017, the portal had had 480 thousand registered users, mainly Slovak, as well as from

other EU countries. The Data.Gov.sk portal includes all state/public institutions and all open data that are not confidential. The aim is to provide all data structured, but as of today, there remains a significant part of them that still are plain (translated) texts.

Marek Vandák has shortly presented the cross-border eProcurement notification solution developed by the company ANASOFT APR, which should encourage businesses from Slovakia, Croatia and Slovenia to enlarge their future customer basis abroad. So far, the company is satisfied with the results of the eTranslation service.

Jakub Začka from the Social Insurance Agency has presented a complex and significantly funded project of EESSI system implementation in Slovakia. As such, the project has direct impact on the whole public health sector. Since 2001, there is an obligation to introduce electronic exchange of social security information. The project has been granted nearly 2 million euros (CEF Telecom) in order to spare national expenses. Since then, the process of digitisation has begun, with the Social Insurance Agency as the national coordinator. Mr. Začka has informed the audience about the requirements of the system, which were high due to the scope of public employment-related institutions (ministries, regional agencies, insurance agencies) engaged. The exchange concerns 570.000 procedures annually, which is equal to 2,4 million reports that are being exchanged. Whereas the usual waiting time with the procedures may take one month, this time can be electronically reduced to some mouse-clicks, so to say. Nowadays, the system is being tested, beginning with the Czech-Slovak information exchange (September 2018). Since the majority of the exchanged data is structured, translation applies mostly to added information or similar fields. An apt foreign project partner in this aspect is still searched for. The results with eTranslation were relatively favorable, with smaller texts translated within 1 minute, larger texts 5 minutes.

Michal Ohrablo, who works for Deputy Prime Minister's Office for Investments and Informatisation, continued with more general presentations given by Aleksandra Wesolowska and Martina Slabejová. He informed on the characteristics of CEF mechanism, its functioning, the role of the INEA agency, which is responsible for the life cycle of calls and projects, as well as for maintaining a viable degree of bureaucracy. Mr. Ohrablo accentuated the importance of some preliminary considerations when applying for a project. Firstly, one has to be aware on the functioning of the funding mechanism, secondly, the project has to have a manageable implementation scheme. The role of the Deputy Prime Minister's Office with respect to the INEA is an advisory one: the Office also provides consultations and monitors calls for projects.

The questions in the panel concerned mainly performance quality of eTranslation. Mr. Začka from the Social Insurance Agency commented on their experience with translation of approximately 6.000 paragraphs, where there was room for improvement and the first version of the text had to be translated again. However, this was a legal text with significant demands on term accuracy. In the end, the case showed the need for feeding the translation machine with documents in the domain. The experience of ANASOFT APR with the eTranslation was described as good. The discussion then focused on general situation with regard to CEF implementation in Slovakia. Mr. Bednár from the NASES used the expression «breakthrough period», and he stated the need for more multilingual digital services in Slovakia as the one of the biggest challenges. The eTranslation will thus have an ever-growing importance for Slovak public sector. Slovakia, however, has begun an ambitious path of public sphere informatisation with remarkable results so far. On the other hand, the completemultilingual coverage of digital public services will still take some more time. Mr. Ohrablo accentuated the importance of multilingual services, mainly in the border regions of Slovakia.

3.6 The CEF eTranslation platform @ work

A live video presentation from Markus Foti (DGT) was transmitted to the audience with live interpretation into Slovak. Mr. Foti presented the eTranslation platform, its principles and mode of operation, its intended and current users (in terms of DSIs), and the steps required to connect with the platform and the benefits of using MT@EC/eTranslation.

3.7 The European Language Resource Coordination (ELRC) action

As a representative for ELRC, Stelios Piperidis focused on putting the ELRC action in the broader context of EU politics, thus answering the question what the ELRC action stands for, what is its place within the CEF mechanism, what are the three pillars of the CEF mechanism itself, etc. Then he presented basic DSIs within CEF Digital, and, more specifically, eTranslation as the one DSI that supports all remaining ones. He also explained the vital need for feeding the translation systems with domain specific documents and terminology. ELRC is responsible for delivering language resources that enable the training of these translation systems. Mr. Piperidis commented on the role of National Anchor Points and specified what kind of data ELRC is mainly looking for: translation memories, parallel documents, large monolingual (Slovak) texts, terminologies. He also emphasized that the ELRC mission includes identification of multilingual public services in respective countries, consulting services regarding technical and legal issues of reusing the documents. The ELRC is interested in getting to know what kind of language tools there are in use. Mr. Piperidis summarized the history of LR acquisition in the ELRC so far: the consortium was able to identify 400 data holders in the timespan of three years (April 2015 – 2018), and as of April 2017, 225 LRs have been gathered, 176 of them labelled as open data or otherwise accessible, 39 other available under other terms of use).

3.8 ELRC in Slovakia

Miroslav Zumrík (T NAP) gave an overview of the steps that have so far been made in gathering multilingual LRs from Slovak public bodies. He emphasized the role of feedback forms which hinted possible ways of how to best proceed when trying to get access to the documents needed. As it seemed, there were at least two approaches: either bottom-up or top-down method, the latter aiming at the highest representatives within given institutions. This approach has been then employed, when 12 Slovak ministries were addressed by official letter to the Heads of respective ministries. The response, although somehow reserved from some of the ministries, resulted in providing at least some contact persons given the task to cooperate with the T NAP in gathering LRs. Miroslav Zumrík mentioned two public bodies where this approach had proven fruitful: Ministry of Justice and Ministry of Culture of the Slovak Republic. These institutions provided altogether more than 1 million tokens of raw mono- and bilingual texts in various language combinations, mainly English-Slovak. The materials covered the agenda of both ministries and a wide spectrum of genres: laws, law proposals, reports, directives, agreements, letters, brochures, invitations, etc. From this raw material, 2 parallel English-Slovak and 2 monolingual datasets were created (containing 100 000 tokens each), aligned at the sentence level and then delivered to the ELRC-SHARE repository. Miroslav Zumrík emphasized the attitude of respective contact persons at the ministries, which made the first collecting of Slovak LRs possible. He then briefly commented on the main obstacles and achievements, summarized lessons learned from the first round of workshops and sketched an outlook into methods of future gathering of LRs.

3.9 Can language data be shared and how?

In the section dedicated to legal aspects of data reuse, Mr. Strehár from the Ministry of Finance of the Slovak Republic focused on the general framework of PSI implementation in Slovakia including informatisation of public services at the level of local government. He dealt mainly with issues of financial compensation for gathering resources (or their free character), whether the documents can have a licence for use, and if yes, what types of licences there are. Several adjustments and new regulations within PSI have been added recently in order to make the reuse of documents more transparent. In Slovakia, an important legal document would be the Law on Free Access to Information. The data made available generally have to fulfill certain standards. The Deputy Prime Minister's Office works now with new legislation regarding data. Mr. Strehár also spoke on various types of possible licenses for these documents (CC, OGL, CC BY). One possible issue with the new GDPR regulation would concern the reuse of documents for other purposes (and by third parties), as originally stated in the licence. Mr. Strehár went further on details with respect to structured data, principles of gathering resources, and the "less is more" principle, which means that the focus should be on quality data first. Mr. Strehár also presented portals like Data.Gov.sk (and found the quality of the portal, as well as position of Slovakia within EU in this aspect, quite good, ranking 3rd as of 2018). The progress has been made mainly thanks to local self-government initiatives and development of civil society. As the main challenges for future digitalization work, Mr. Strehár identified awareness raising on open data, making them more accessible, and solving the relevant technical issues. In the conclusion, he also mentioned some interesting open data portals (like Crime Map, based on criminal statistics of the Slovak police), and stressed the possible use of open data within the business sector.

3.10 Preparing and sharing data with the ELRC repository – and what happens next

As the second representative for ELRC, Maria Giagkou defined elementary concepts used within the ELRC action (language resource, metadata, etc.), and then presented the ELRC-SHARE portal. She described what types of documents EC needs/prefers (XMLs, spreadsheets, plain texts), what the Commission already has, and what is particularly lacking. Then she went through practical issues and details on how the document should be formatted (what alignment, with what naming of files, from which domains should the texts be), in order to be of use for training of the translation systems. She also dwelt some time on the issue what the LR providers should/could do, if they have something to share, what is the procedure of registration at the ELRC-SHARE, and she presented several ways of step-by-step LR uploading. Another area she concentrated upon was what happens with provided data after being fed into the repository (anonymization, cleaning, converting).

3.11 Identifying and managing your data: Questions & Answers

3.12 Conclusions

These two sections have more or less unfolded as one discussion, covering a wide range of issues. Firstly, it was the restriction of access rights to the eTranslation services, then the procedure of getting the access, and then a question concerning the commercial or non-commercial character

of the ELRC consortium. Another interesting question was raised on whether there is an interdependency between providing LRs to eTranslation and using the service. One participant asked about the experiences of ELRC with LR harvesting in other countries. Maria Giagkou replied that the obstacles are more or less similar in all countries. Another question focused on the persistent role of National Anchor Points in the resource delivery. Stelios Piperidis stressed the openness of future scenarios for the ELRC related cooperation. The ELRC representatives also described the encryption/decryption procedures, applied to documents fed to translation machines. This ensures that the information contained in documents would only be read by machines. An important issue was raised toward the end of discussion, concerning how to motivate institutions to donate data and how to build a functioning pipeline for sustainable LR delivery in the future. The participants were additionally interested in getting to know whether providing LRs to ELRC is a prerequisite in order to use the eTranslation service. It was explicitly stated that it is not; eTranslation is open to public bodies regardless of their involvement with ELRC. Another point of interest was the workflow of ELRC action, registration of eTranslation users, processing the LRs, the type and characteristics of content that the consortium is mainly looking for.

At the very end of the workshop, Mr. Začka repeated that the need for eTranslation only would grow in time, and that it would spare thousands of man-hours annually. Local organizers then thanked all speakers, participants, the EC Representation, as well as interpreters and colleagues from the Ľudovít Štúr Institute of Linguistics for their contribution to making the workshop as mutually useful as possible.

4 Synthesis of Workshop Discussions

4.1 ELRC and Open language Data in Slovakia

Based on information from the legal presentation (S2.3 Can language data be shared and how?), from the Panel and from the Open Data representative's presentation, it is possible to conclude the following:

The Slovak government and its bodies has in the recent years shown a strong wish to proceed with the centralized digitalization of public services (with Deputy Prime Minister's Office for Investments and Informatisation as the crucial public body), creating a governmental cloud for better data management, maintaining the Open Data portals, and continuation of PSI Directive implementation. The government also encourages representatives from various public administration offices and businesses to apply for open CEF Calls. The main challenges, as the questions from audience showed, is the sparse awareness about the range of technological solutions and financial mechanisms that the public bodies could take advantage of (in tasks like data cleaning). On the other hand, it is also important to point at already running and successful complex projects (within DSIs like eProcurement and EESSI) and comprehensive national open data portals (Slovensko.sk and Data.Gov.sk). These portals contain more than 1700 e-services provided by the central or by local/regional government throughout the country (the City of Prešov being one of the finest examples with most provided datasets) and covering all major areas of life. There is a great deal of structured data in these datasets, and the main part of the Open Data compilations remains numerical, but there are at least regions (like Southern Slovakia on the borders to Hungary) where multilingual digital services would be (and partly are) of high relevance. The fact that Slovakia ranks, as Mr. Strehár (during the legal presentations) pointed out, as the 3rd country in the EU, is also a very promising one. The panel discussion showed that it is mainly in the public health sector, where eTranslation would prove, given the amount of information exchanged annually, as an apt instrument for making the social security procedures more effective, user-friendly and faster. The participants were on their side mainly interested in getting to know whether eTranslation service requires providing LRs to ELRC before letting the public bodies to make use of the service. Another point of interest was the very workflow of ELRC action, registration of eTranslation users, processing the LRs, the type and characteristics of content that the consortium is mainly looking for.

4.2 Success stories and lessons learnt

We have seen that the ELRC initiative can be embraced by representatives of Slovak public bodies, thanks to individuals at those bodies willing to engage. The first datasets provided could show the way for representatives of other public bodies, in that they would understand that providing at least some highly valuable resources is in principle possible, and that the processing of these resources takes not much effort in advance. We have also seen that the amount of data that are already open to the Slovak public is quite big, and that there were participants in the audience who let us know that they already knew about possible resources to be gained at their, or affiliated, institutions. The first two workshops can be seen as two first steps towards creating a functioning pipeline for future LR delivery, whereas personal networking is highly important, even though it might be a little hesitant in the beginning. The good thing is to know that the Slovak government already has invested a substantial effort into maintaining major national open data portals and that more projects are being prepared and/or already running. It is also good to know that informatisation of society has one centralized institution, which is mainly responsible for this

European Language Resource Coordination (ELRC) is a service contract operating under the EU's Connecting Europe Facility SMART 2015/1091 programme.

task (the UPVII Office). This would enable a more effective coordination of digital public services. As is seems, the amount of Slovak datasets to be gathered without extra effort is not insignificant – this we have seen based on information provided in the engagement forms. The follow-up to the workshop (collecting of the resources) should be done as soon as possible, while the appeal of information provided at the event is still relatively vivid amongst its participants. What also showed as a quite effective door opener was to use official letters and invitations signed by high EC officials.