



**European Language
Resource Coordination**
Connecting Europe Facility

Deliverable D3.2.15

Task 8

ELRC Workshop Report for Bulgaria

Author(s):	Ministry of Transport, Information Technology and Communications
Dissemination Level:	Public
Version No.:	<V1.0>
Date:	2018-12-07



Contents

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	National initiatives for open data	8
3.4	Connecting public services across Europe: ambition and results so far	9
3.5	CEF in Bulgaria: an outlook into current and future challenges – Panel session	9
3.6	The CEF eTranslation platform @ work	12
3.7	The European Language Resource Coordination (ELRC) action	13
3.8	ELRC in Bulgaria	13
3.9	Can language data be shared and how? National and European legal framework	14
3.10	Preparing and sharing data with the ELRC repository – and what happens next	15
3.11	Presentation of the project eTranslation Termbank (eTTB)	15
3.12	Identifying and managing your data: Questions & Answers	15
3.13	Conclusions	16
4	Synthesis of Workshop Discussions	18
4.1	ELRC and Open language Data in Bulgaria	18
4.2	Success stories and lessons learnt	19
5	Workshop Presentations	20

1 Executive Summary

This document reports on the European Language Resources Coordination (ELRC) workshop which was held on October 24, 2018, at the Representation of the European Commission in Bulgaria (House of Europe; address: 124 G.S. Rakovski Str., Sofia).

The national ELRC workshop was attended by over 150 participants from the Administration of the President of the Republic of Bulgaria, the National Assembly and the Council of Ministers, 17 ministries, 17 agencies, different committees, state companies and institutes, the Bulgarian Academy of Sciences, academic community, companies active in the field of professional translations, information and language technology and publishing, etc. The event was attended by two representatives from the ELRC Consortium and a representative of the European Commission. The workshop was jointly organised with the Bulgarian Language Institute of BAS and the EC Representation in Bulgaria. The availability of such a large number of representatives from different institutions is a statement of the major achievement in the process of disseminating of information on the benefits of sharing language resources for the ELRC repository and the needs of the CEF eTranslation platform.

The official opening of the event was attended by several officials as the Minister of Transport, Information Technology and Communications, Mr Rossen Jeliakov, the Deputy Minister for the Bulgarian Presidency of the Council of the EU 2018, Ms Monika Panayotova, and the Head of the EC Representation in Bulgaria, Mr Ognyan Zlatev.

Ms Mariya Gabriel, European Commissioner for the Digital Economy and Society, sent a special video message for the participants of the workshop. The Deputy Prime Minister of the Republic of Bulgaria, Mr Tomislav Donchev sent a greeting address, which stated support for the idea and goals of sharing of language resources.

The participants in the workshop had the opportunity to gather information about connecting public services in Europe and the national context of the Connecting Europe Facility, CEF Digital, the eTranslation platform, and Bulgaria's open data policy. The functionalities of the new National Open Data Portal were presented during the event, as well as the practical implementation of the system for electronic delivery "eDelivery". The panelists presented the legal framework of public sector data sharing, as well as the Directive on the reuse of public sector information that provides a common legal framework for a European market for government-held data.

2 Workshop Agenda

- 08:00 - 08:30 **Registration**
- 08:30 - 08:35 **Welcome and introduction**
Hristina Dobрева, Chief Expert at Ministry of Transport, Information Technology and Communications (MTITC) and National Anchor Point from the Public Administration
- 08:35 - 09:05 **Welcome by the EC**
Mariya Gabriel, European Commissioner for Digital Economy and Society - video address
Tomislav Donchev, Deputy Prime Minister - greeting address
Rossen Jeliakov, Minister of Transport, Information Technology and Communications
Monika Panayotova, Deputy Minister for the Bulgarian Presidency of the Council of the EU2018
Ognian Zlatev, Head of the EC Representation in Bulgaria
- Session 1. Connecting a multilingual Europe: European context and local needs**
- 09:05 - 09:30 **National initiatives for open data**
Reni Antonova, Chief expert at Data Department, the State e-Government Agency (SEGA)
- 09:30 - 09:50 **Connecting public services across Europe: ambition and results so far**
Aleksandra Wesolowska, EC, Directorate-General for Communications Networks, Content and Technology - Video Presentation
- 09:50 - 10:40 **CEF in Bulgaria: an outlook into current and future challenges - Panel session**
Moderator:
Eli Kaneva, State Expert at MTITC and National Representative in the Telecom Sectoral Committee under the Connecting Europe Facility
Panellists:
Tsvetanka Kirilova, Expert at the Information Systems and Interoperability Directorate at SEGA, Presentation of the e-Delivery System
Stanimir Minkov, Senior Expert in the Modernization of the Administration Directorate, Administration of the Council of Ministers, Presentation of the National Open Data Portal
- 10:40 - 11:05 **Coffee Break**
- 11:05 - 11:30 **The CEF eTranslation platform**
Markus Foti, eTranslation Programme Manager, European Commission, Directorate-General for Translation
- Session 2. Engage: hands-on data**
- 11:30 - 11:50 **The European Language Resource Coordination (ELRC) action**
Stelios Piperidis, ILSP/Athena R.C., ELRC
- 11:50 - 12:10 **ELRC in Bulgaria**
Prof. Svetla Koeva, Director of the Institute for Bulgarian Language of the Bulgarian Academy of Sciences (BAS) and National Anchor Point for language technologies
- 12:10 - 12:30 **Can language data be shared and how? National and European legal framework**

Jeneta Rogova, Head of Legal Affairs in Transport, Communications and electronic governance Directorate at MTITC

12:30 - 13:30

Lunch

13:30 - 14:00

Preparing and sharing data with the ELRC repository - and what happens next
Maria Giagkou, ILSP/Athena R.C., ELRC

14:00 - 14:15

Presentation of the project eTranslation Termbank (eTTB)

Andraž Repar, Representative of the Consortium eTranslation Termbank (eTTB)

14:15- 14:45

Identifying and managing your data: Questions and Answers

Moderators: Hristina Dobрева

14:45 - 15:00

Conclusions

Moderator: Hristina Dobрева

3 Summary of Content of Sessions

3.1 Welcome and introduction

Ms Hristina Dobрева, Chief Expert at Ministry of Transport, Information Technology and Communications (MTITC) and National Anchor Point from the Public Administration (and a member of the Language Resources Board of the ELRC action) opened the second National ELRC workshop. She briefed the participants on the topics of the workshop and thanked everyone for their presence. Ms Dobрева gave a short overview of the registered participants, highlighting the large number of participants from different institutions and pointing out that only one ministry of all 18 ministries did not attend the workshop. Ms Dobрева presented the official guests, as well as the Commission representative, Mr Markus Foti, and the representatives of the European Language Resources Coordination Consortium (ELRC), Ms Maria Giagkou and Mr Stelios Piperidis. She also brought to the attention of the participants the video message sent by the Commissioner for Digital Economy and Society, Ms Mariya Gabriel, and the greeting address from Mr. Tomislav Donchev, Deputy Prime Minister of the Republic of Bulgaria. Ms Dobрева thanked all participants for their attendance at the workshop and asked them to complete the feedback forms and engagement forms that they received in the information folder.

3.2 Welcome by the EC

Mr Rossen Jeliakov, Minister of Transport, Information Technology and Communications pointed out the importance of initiatives that promote cohesion, mutual understanding and justify the financial resources invested. “It is good if we are taking initiatives in our common European home that promote cohesion, mutual understanding and at the same time justified allocated resources.”, said Minister of Transport, Information Technology and Communications, Mr Rossen Jeliakov. The Minister pinpointed to the fact that there are 24 official languages spoken in Europe and 3 different alphabets are used. In this multilingual context the eTranslation platform helps overcome the language barriers. The Minister recognized eTranslation as a state-of-the-art automated translation system, different from relevant “competitors” in terms of:

- the level of security for the user’s sensitive and personal data;
- Its output quality.

Minister Rossen Jeliakov stressed that eTranslation will help not only administrations, but also all citizens in the Union who need to use quality multilingual public services. “I am sure that it will address not only the differences and language barriers, but it will also contribute to a better understanding of the ways and forms of functioning of public institutions within the European Union”, he added.

Minister Jeliakov clarified that the Ministry is the national coordinator for the Transport and Telecommunications sectors of the Connecting Europe Facility, by which the eTranslation platform is financed. He shared with the participants that the Ministry of Transport, Information Technology and Communications (MTITC) and the Bulgarian Institute of Sciences (BAS) had already provided a substantial dataset for training eTranslation in the Bulgarian language. In this regard, Mr Jeliakov appealed to the participants to make more language resources available to the platform so that it can be enhanced and exploited by even more consumers, helping citizens and businesses to use e-services

in Europe. “Here is our request to provide resources and documents, so that the platform is increasingly building up capacity and potential for professional use,” said the Minister.

The Deputy Minister for the Bulgarian Presidency of the Council of the EU2018, Ms Monika Panayotova and the Head of the European Commission Representation in Bulgaria, Mr Ognian Zlatev greeted all participants.

Ms. Monika Panayotova - Deputy Minister for the Bulgarian Presidency of the Council of the EU2018:

Ms. Panayotova thanked for the invitation and welcomed the workshop participants. She stressed that the event is an example of how European and national institutions, together with the academia and research community and innovative business, pool their efforts to overcome language barriers in Europe, contributing to a well-functioning Digital Single Market (DSM). She reminded that one of the key priorities of the Bulgarian Presidency was *digital economy and skills of the future*. According to her, new ICT technologies will have a strategic impact on European policy and economy and that, in order to use the potential of the Digital Transformation, we need to adapt national policies and create the necessary communication and internet infrastructure. The expectations towards the Trio Presidency of the Council of the EU — Estonia-Bulgaria-Austria — are high on the themes of the digital economy and transformation by building the DSM and promoting digital and communication connectivity in Europe. Ms Panayotova stressed that one of the main problems for the functioning of the DSM is the language barrier that the institutions and citizens are facing on a daily basis. According to Deputy Minister Panayotova, such projects strengthen our feeling of belonging to a European common space.

Ms Panayotova briefed the participants on the key dossiers that had been completed as a way to achieve the DSM objectives:

- Audio-visual Media Services (AMS) Directive;
- The European Electronic Communications Code;
- The Body of European Regulators for Electronic Communications (BEREC);
- The Regulation on a framework for the free flow of non-personal data in the European Union.

These steps will ensure easier and more secure access to online goods and services and the overcoming of language barriers which will help place Europe at the forefront of global information society. Ms Panayotova concluded that in the European Year of Cultural Heritage, which runs under the motto: “Our legacy: where the past meets the future”, one can precisely realize the value of digital tools in preserving linguistic diversity, while also facilitating cross-border multilingual communication.

Mr Ognian Zlatev - Head of the EC Representation in Bulgaria:

Mr Zlatev welcomed the workshop participants as a host of the House of Europe. He pointed out that the event took place at a very important time — the last year of the current European Parliament and the achievement of the main priority — Digital Single Market and Digital Economy, and this project is a significant step towards these goals. Mr Zlatev paid special attention to the major achievement of the European Commissioner for Digital Economy and Society — Ms Mariya Gabriel — the new Digital Europe Programme, which would cover five pillars: Digital skills, Cybersecurity, High performance computing (HPC), digitalisation of administration and artificial intelligence. According to him, the realisation of this project would enable the Bulgarian administration to continue to accumulate and to pass on positive experiences to the other Member States.

Ms Mariya Gabriel, European Commissioner for Digital Economy and Society - video address

Commissioner Mariya Gabriel sent a special video message to the participants, in which she congratulated the organisers of the event on the strong message they pass that linguistic diversity is one of the greatest treasures of Europe, to which all citizens and businesses should have access. For this purpose, it is necessary to collect language resources in Bulgarian for the needs of the eTranslation platform, in order to enable public administrations across Europe to provide their services in Bulgarian, as well as the Bulgarian public administration to be able to offer their services in all EU languages. Ms Gabriel noted that the EC launched the ELRC in April 2015, as this infrastructure is part of the Connecting Europe Facility and provides public digital multilingual services in order to facilitate contact between citizens, business and public administrations.

The objective of the coordination of language resources in Europe has three dimensions:

- To share resources;
- To continuously improve the existing machine translation service;
- To reuse collected data.

Ms Gabriel noted that the first language resource in Bulgarian language was provided in 2016 by the Ministry of Transport, Information Technology and Communications, and subsequently resources were added from the Ministry of Justice, the Institute of Justice, the National Revenue Agency, BAS and a number of other institutions. The Bulgarian language is the fifth largest resource in the repository.

Ms Gabriel stated that when there is a common cause, then better results can be achieved.

Mr Tomislav Donchev, Deputy Prime Minister - greeting address

Mr Donchev sent a greeting address to the workshop participants. He expressed his sincere wishes for the successful implementation of the workshop and apologised for his inability to attend. He noted that it is of utmost importance that communication between Member States is effective and accessible to both EU institutions and citizens. He expressed the view that the theme of the forum — the opportunity to communicate through digital-innovative technologies — would open up new horizons and the opportunity to achieve a digital revolution.

3.3 National initiatives for open data

Ms Reni Antonova, Chief expert at Data Department, the State e-Government Agency (SEGA):

Ms Antonova presented the legal framework for open data, the authority and responsibilities of the State e-Government Agency (SEGA), the open data initiatives, the different open data portals in Europe and some quantitative features of the data hosted in the European open data portal. She explained that in Bulgaria the information reused in the public sector is stored and provided in open and machine-readable format and according to the provisions of the Access to Public Information Act (APIA) and the Electronic Government Act. Ms Antonova noted that the State e-Government Agency develops and maintains the Open Data Portal and adopts a list of datasets on an annual basis. The SEGA also produces a synthesis report on the availability of public sector information every three years. Ms Antonova explained what open data initiatives are available in Bulgaria, namely:

- The Administration of the Council of Ministers — the existing National Open Data Portal — <http://opendata.government.bg> and the new Open Data Portal — <http://data.egov.bg>;
- MTITC;
- The Institute for Public Administration — regularly organises Open Data Training;
- SEGA;

She also brought to the attention of the audience the European Open Data portals:

- The EU Open Data Portal;
- European Data Portal;
- Persistent URI;
- The EU Web Archive.

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

Ms Aleksandra Wesolowska, EC, Directorate-General for Communications Networks, Content and Technology - Video Presentation

Ms Wesolowska's video presentation aimed to raise awareness of the Connecting Europe Facility (CEF Digital) and the implementation of the CEF eTranslation platform, and in particular, how administrations could benefit from it. Her presentation thoroughly described the digital service infrastructures (DSIs), the building blocks and how they are implemented at European and national level. The video also gave a brief overview of how administrations can benefit from the platform, how it can be implemented in different systems, and how it can be used as a building block.

The presentation took note of some of the applications of the eTranslation platform carried out so far:

- The Internal Market Information System (IMI) uses eTranslation to translate text messages automatically;
- In the Tenders Electronic Daily (TED) eTranslation is used to translate tender notices;
- eTranslation will strengthen Digital Public Procurement (eProcurement) infrastructure, linking 3 e-procurement platforms respectively in Slovakia, Slovenia and Croatia;
- eTranslation will be integrated as a translation tool to be available to Member States holding the EU presidency.

Ms Wesolowska called on the participants to support the ELRC's cause by sharing language resources.

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

Ms Eli Kaneva (Moderator), State Expert at MTITC and National Representative in the Telecom Sectoral Committee under the Connecting Europe Facility

Ms Kaneva took note of the role of the Sectoral Committee, which is to assist the European Commission by, for example, agreeing the list of projects approved for funding under CEF. She explained that MTITC is a coordinator in the telecommunications and transport sectors and actively promotes participation opportunities by organising information days and sessions, and assists potential participants in preparing projects. Ms Kaneva noted that since its launch in 2014 under the CEF Telecom sector, many calls were conducted and there was a sufficient Bulgarian participation in

them – in digital service infrastructures and building blocks. She mentioned examples of projects in the first type of projects: The e-Justice portal, the European platform for the interconnection of European business registers, the Electronic Data Exchange, the Online Dispute Resolution Platform, the Digital Library of Europe, and the Safer Internet Platform.

Ms Kaneva noted that Bulgaria has successful projects as part of the Electronic Exchange of Social Security Platform, Europeana and on safer internet. The panellist explained what constitutes the main building blocks, namely basic infrastructures that can be reused. There are 5 main building blocks — electronic delivery of documents, digital signature, electronic identification, e-invoicing and eTranslation. Ms Kaneva explained that eTranslation requires user identification via EULogin and so indirectly, the platform is linked to electronic identification. She noted that the Bulgarian Language Institute is actively involved in eTranslation and already had two successful projects. In 2019 there will be a new eTranslation call.

Concerning the challenges ahead, Ms Kaneva made a short presentation on what is going on after 2020, namely that the CEF will continue to function, but that it will be more network-oriented and the infrastructure will be funded and developed under the Digital Agenda for Europe.

Ms Kaneva noted that the purpose of the panel was to present national online services that integrate and/or use the building blocks.

Ms Tsvetanka Kirilova (Panellist), Expert at the Information Systems and Interoperability Directorate at SEGA, Presentation of the e-Delivery System

Ms Tsvetanka Kirilova thanked the organisers for the possibility to participate in this prestigious forum and for including one of the building blocks of the Bulgarian eGovernment, namely the Secure Electronic Delivery System “eDelivery” to the workshop program. The system is also part of the single application model for payment and provision of electronic public services at national level. The system is built in accordance to European and national legislation, e.g. Regulation 910/2014 on electronic identification and trust services for electronic transactions in the internal market, as well as in compliance with all the requirements of GDPR and, at national level, Electronic Government Act, the Electronic Signature and Trust Services Act, the Electronic Communications Act and any sub-regulations thereto.

Ms Kirilova noted that all e-government systems are in compliance with the European Interoperability Framework principles, which involve multilingual, reusable building blocks and interoperability by default. This system shall be identical to electronic registered delivery, providing all users with security, respecting personal data protection and privacy requirements, security for the integrity of transmitted messages and electronic documents security for the duration of the communication. The system is designed to secure interactions between citizens, businesses, and public administrations. The system is also in line with one of the main principles of Regulation 910/2014 that the electronic documents received and sent by electronic means are the correct ones as each message is signed with a time stamp, which guarantees the time of sending and receiving, and that each communication is stamped with the electronic seal of the SEGA as a third party who ensures the flow through a secure environment.

Ms Kirilova clarified that the eDelivery system information is available on the SEGA website. The system has a user interface and can be used after registration as a legal representative of the organisation concerned. Ms Kirilova remarked that there is a possibility to integrate the system into the IT systems of relevant organisations for both public administration and businesses. She also noted that the

multilingual principle was respected, but for the time being the system is bilingual and offered in Bulgarian and English. The system will be further redeveloped and enriched at a later stage. The system will be included in the European eDelivery infrastructure as one of the ways to use electronic administrative services at national level. Ms Kirilova clarified that the possibility would be extended to all European citizens, who have an electronic identity as required by Regulation 910/2014, to create their user account in the system via the eIDAS node, which has already been built and is operational. The integration of eTranslation to further enhance the system could be one of the future goals.

Ms Kirilova presented several video clips:

- How to request, i.e. to form a system message. At the moment, the recipient may be merely an administrative organisation and the message is formed by an individual person. Registration in the system is currently via PIC (personal identification code) from NSSI and electronic identification but is extended in order to be able to use all the identifiers authorised by a formal act.
- How to enter the profile of the user.

Question: Having PIC (personal identification code) from NSSI can one send documents only to NSSI or to any other organisation/administration?

Answer (Ms Tsvetanka Kirilova): There are currently more than 10000 registered citizens in the system. The registration of administrations is on voluntary basis at the moment. The user can send messages to all administrations that are registered in the system. Then the deadline for replies starts. All administrations and persons referred to in Article 1 (1) of the Electronic Government Act (notaries, private bailiffs, persons providing public services) will be obliged to make their own account in the system by means of a Decision of the Council of Ministers.

Mr Stanimir Minkov (Panellist), Senior Expert in the “Modernization of the Administration” Directorate, Administration of the Council of Ministers, Presentation of the National Open Data Portal

Mr Minkov is the administrator of the National Open Data Portal (new and old). He explained that the (new) portal is still under development and still not finalised. Mr Minkov briefed the participants on the old Open Data Portal, which had been operational since the end of 2014 (only public bodies publishing data in open formats have contributed to the old open data portal). Mr Minkov pointed out that the new portal would allow for the registration of civil and business organisations that have their own datasets, meaning that it would build on the very open data policy at national level and would offer a wider range of information.

The portal is built in accordance with the requirements of the Access to Public Information Act, and the information it publishes is set out in the Regulation on standard terms and conditions for reuse of public sector information. Mr Minkov demonstrated the functionalities of the new open data portal and how information from the old portal will be transferred. The new portal was set up in 2018 and the data published on it is divided into thematic areas, following the European Data Portal subject areas categorisation. The new portal will allow different data sets to be interconnected. Mr Minkov noted that the portal is now only bilingual in Bulgarian and English, but it is foreseen to be multilingual. He also explained that a Guide to the Portal will also be published. Mr Minkov noted that the portal is based on open source software, i.e. any functionality could be further worked or reused in the creation

of other platforms and/or products. The data itself can be reused by other applications, which is the essence of the open data concept.

Question: How is the long-term storage of the data published on the Portal ensured?

Answer (Mr Stanimir Minkov): The data is stored in an archive on a secure server of the Portal owner, which is located in the SEGA.

Ms Kaneva made concluding remarks, pointing out that two national services were presented in the panel, making use of CEF DSIs and called for greater involvement in all CEF initiatives.

3.6 The CEF eTranslation platform @ work

Mr. Markus Foti, eTranslation Programme Manager, European Commission, Directorate-General for Translation

Mr Foti briefed the participants on the development of the automated translation platform so far:

- MT@EC - the European Commission's Machine Translation system was launched in June 2013 and was Moses-based;
- eTranslation - the successor to MT@EC, is streamlined, cloud-based neural machine translation system.
- CEF Automated Translation - the platform which includes eTranslation but is also intended to integrate other NLP tools (e.g. named entity recognition, transliteration, etc. based on user needs). Related projects and grants (e.g. ELRC, Translator for the Council presidency, generic services projects) are also under the umbrella of the CEF Automated Translation building block.

The eTranslation platform is available for individuals and machine-to-machine use. Users are Digital Service Infrastructures (EESSI, ODR, Open Data Portal, Europeana, etc.), System suppliers (EURLex, N-Lex, Internal Market Information system) and individuals in public administrations. Benefits of using the platform are: increase speed and productivity, reduce costs and facilitate information exchange.

The eTranslation user interface is available since November 2017. Some of its features include:

- User interface – All EU Languages,
- Supports common document formats (Office formats, ODT, PDF, html, xml),
- Quality output

The N-Lex portal is an example of machine-to-machine eTranslation service. N-Lex offers a single entry point to national law databases in individual EU countries. Let's assume for instance that the user needs to retrieve relevant legislation for social security issues in Lithuanian. The European Data Portal (EDP) is a second example of services enhanced with eTranslation. The descriptions (metadata) of each dataset retrieved are machine translated and available in all EU languages.

Mr. Foti went on to discuss the recent shift to neural machine translation (NMT). Large Artificial Neural Networks are trained on existing translations. NMT training induces hidden representations for textual data and exploits hidden generalisations. It constitutes a radical departure from the phrase-based SMT approaches, where subcomponents are separately optimized. NMT can generate more fluent and grammatical translations and it improves its performance on previously unseen data. CEF.AT brings reliable and trustworthy translation for EU and National Public Administrations, support for languages with fewer speakers, opportunities for private sector, through grants and generic services,

higher quality language technologies, thus fostering demand and public availability of data collected by ELRC.

3.7 The European Language Resource Coordination (ELRC) action

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

Mr Piperidis briefed participants on the ELRC, which is an action funded by CEF. CEF is a key EU funding instrument that supports the development of high performing, sustainable and efficiently interconnected trans-European networks in the fields of transport, energy and telecommunications. The Telecom strand, among other things, funds the deployment of digital service infrastructures (DSIs) – this part of CEF Telecom is called CEF Digital. One of these Digital Service infrastructures is eTranslation, basically this is the Machine Translation system of the European Commission. And here is where ELRC comes to play, as an action supporting the eTranslation DSI, by coordinating the collection of language resources that are necessary to enhance eTranslation.

The ELRC Consortium is: DFKI – the German Research Centre for Artificial Intelligence; ELDA – the Evaluations and Language Resources Distribution Agency; TILDE – a Latvian Language Technology and Services Provider and ILSP – the Institute for Language and Speech Processing with 30 ELRC Technological NAPs (one per CEF affiliated country), 30 ELRC Public Services NAPs (one per CEF affiliated country) and relevant legal experts.

The aims and objectives of the ELRC action include first and above all: Collecting language resources; identifying public services in need of multilingual functionalities and their corresponding multilingual needs; 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 went on to explain that the way to enhance the performance of eTranslation is with the corresponding “training data” that is “fed” into the system (i.e. language resources, such as bilingual corpora, multi-lingual corpora, terminologies, etc.). In-domain training data (i.e. translations from the target domain) are essential for achieving high-quality translation.

3.8 ELRC in Bulgaria

Prof. Svetla Koeva - Director of the Institute for Bulgarian Language of the Bulgarian Academy of Sciences (BAS) and National Anchor Point for language technologies

Ms Koeva started her presentation with a clarification of the term ‘language resource’. Ms Koeva explained that, depending on their content and structure, language resources can be: corpora; semantic assets; description of language rules (grammar) and language processing programmes.

Prof. Dr. Svetla Koeva delivered a detailed explanation to the participants on the purpose of collecting Bulgarian language resources within the ELRC framework. The resources that are collected in the project repository are used for the “learning” of the eTranslation platform in order for it to translate

into and from Bulgarian, to improve the quality of automatic translation tailored to the needs of public administrations.

Ms Koeva, briefed the participants on the results of the first workshop held on 18 March 2016, explaining that 130 people attended and provided detailed information on the various institutions that had been represented.

According to Ms Koeva, the Bulgarian resources in ELRC-SHARE are 5% of the total resources hosted in the repository. In total, Bulgarian is encountered in 23 resources (6 monolingual, 13 bilingual and 4 multilingual). ELRC-SHARE currently hosts 104.504 Bulgarian terms and 21.659.096 tokens. Ms Koeva additionally provided an overview of the distribution of resources by thematic areas: 6 resources in the law domain, 3 policy resources, 2 European Union topics, 2 in the field of science, 2 on social issues, etc. Ms Koeva made a reference to the contributing institutions, e.g. the Ministry of Transport, Information Technology and Communications, the Ministry of Justice, the Institute for Bulgarian language for the Bulgarian Academy of Sciences and others.

Prof. Dr. Koeva discussed the main obstacles that prevent sharing of data. Some of them are, according to Dr. Koeva: identifying and convincing high-level officials to authorize sharing, lack of resources, no established procedures for translation of documents on administrative level, technical difficulties relating to data processing and last but not least, resistance to new technologies.

As the main achievements, Prof. Dr. Koeva identified: the procedure established by ELRC for making resources available, the raised awareness among public sector representatives for the value of language resources, the fact that a considerable number of resources, covering a wide range of thematic areas has already been collected, etc.

As steps to be taken hereafter, Prof. Dr. Koeva highlighted the need to further promote the value of language data for language technologies. Language data sharing should be embraced by a wider audience, while the data contribution processes should be fine-tuned. Prof. Dr. Koeva stressed the importance of using translation memories in the translation process.

3.9 Can language data be shared and how? National and European legal framework

Ms Jeneta Rogova - Head of Legal Affairs in Transport, Communications and electronic governance Directorate at MTITC

Ms Rogova presented the Directive on the reuse of public sector information, the national legal framework transposing the Directive, in particular the Access to Public Information Act. Ms Rogova noted that the Ordinance on standard terms and conditions for the reuse of public sector information and its publication in an open format are the normative acts that define the standardised public sector conditions for the use of information in terms of copyright which we may have in the information we hold. She noted that for sharing language resources, it is appropriate to join forces with the National Open Data Portal. She explained that the reuse of public sector information for private or commercial purposes with minimal or no legal, technical or financial constraints is important for economic growth and innovation.

The speaker noted that laws and their official translations, administrative rulings, documents containing ideas and concepts, news, facts, and informative material are not copyright-protected data.

She clarified that the rules for reuse are at the top of the other legal regimes and assured that, when all other legal conditions are met, public sector information should be available for reuse by all, as easily as possible. She pointed out that when sharing resources, it is necessary to avoid internet content, as it is unclear to whom the rights belong or to contact the ELRC department for legal assistance.

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

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

Ms Giagkou started her presentation with a discussion of the value of data. She emphasized that, like numerical data, language data are a valuable capital as well, especially for language technologies, such as machine translation. She went on to explain 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 that is managed for several languages, Policies, Terminologies, Archives, Forms, and FAQs are useful for improving the platform. She explained ways to provide language resources and appropriate formats about this. She also presented the basic steps through which language resource can be shared through the ELRC-SHARE repository. Dr. Giagkou 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 of technical and legal helpdesk and on-site assistance.

3.11 Presentation of the project eTranslation Termbank (eTTB)

Mr Andraz Repar, Representative of the Consortium eTranslation Termbank (eTTB)

Mr Repar made a presentation on the eTranslation Termbank (eTTB) project, which aims at providing new Terminological Resources (TRs) (i.e. monolingual glossaries, term lists, dictionaries, etc.) to the language resource pipeline of CEF eTranslation for three sector-specific domains: health, business legislation and consumer protection to support the respective CEF DSIs: eHealth, e-Justice and Online Dispute Resolution. A minimum of 150 TRs covering all official languages of the EU plus Norwegian and Icelandic will be provided by the end of the project. The eTTB consortium consists of 8 members each of them responsible for several languages. Jožef Stefan Institute covers Slovenian, Croatian and Bulgarian. He called for the active sharing of Bulgarian language resources.

3.12 Identifying and managing your data: Questions & Answers

Ms Hristina Dobрева

The National Anchor Point of the Bulgarian Public Administration presented some practical tips related to the identification of language resources. The information was summarized in the form of questions and answers. The proposed questions are collected by the ELRC Consortium and arose on the basis of their daily work. Ms Dobрева highlighted the importance of a data management plan for any organisation. The data management plan should specify how data is processed during the production

process and thereafter. It covers the entire life cycle of the data and defines the data policy in order to effectively manage it and ensure its sustainability.

According to the Ms Dobрева it is necessary to envisage all potential legal issues concerning the ownership of the data. According to her, data must be legally “cleared”, i.e. the intellectual property rights should be identified. External translation services providers should respect the “ownership” rights of the contracting entities. It is recommended that all produced intermediate documents (e.g. translation memories) be handed over to the assignor and remain the assignor’s property. Personal privacy issues should be checked beforehand and anonymised, if necessary. The data should be delivered in an appropriate format (not only PDFs, but also TMX / XML / etc.).

If a new resource is built from several existing resources, all these resources must be released. Licenses should allow reallocation of derived products. If an institution is the owner of the resources and distribution rights have been granted to the third party, the distribution agreement must allow distribution of resource through an additional channel using CC licenses. In her presentation Mrs. Dobрева drew attention to the Creative Commons - the licenses and the rights which they provide.

Ms Dobрева described personal data issues and when they should be anonymised. In case of any doubts about how to handle data sets containing personal data, Ms Dobрева recommended consultation with the ELRC team, which offers legal assistance, as well as a service to anonymize shared resources. The consortium also offers language processing services to public administrations (data conversion, label removal, reformatting, cleaning, alignment, validation of metadata, etc.). Technical assistance is also provided. These services are free of charge.

3.13 Conclusions

In the context of the final panel Ms Hristina Dobрева summarized the Second National ELRC Workshop. She reviewed all the topics discussed within the framework of the day such as: Connecting Europe Facility, the eTranslation platform and the ELRC project, good governance activities and practices in Bulgaria, open data, multilingualism and language technologies, review of the legal framework for data sharing in the public sector, Directive on the reuse of information, necessary data types and how they can be shared with the ELRC repository, the ways in which ELRC can support identification and sharing of language resources.

Ms Dobрева once again appealed to the workshop participants to share their data (multilingual databases, terminology or even large monolingual databases) with the eTranslation platform through ELRC. In case of concerns in the sharing process, Ms Dobрева stressed the possibilities of technical and legal support from ELRC. She clarified that assistance could be provided in the preparation, processing and anonymization of data.

In conclusion, Ms Dobрева highlighted that not only administrations will benefit from the eTranslation platform, but also businesses and all European citizens who will have the opportunity to use the EU and national digital services in their own language.

Questions concerning the eTranslation platform, legal and technical questions about sharing of data and etc. were raised by the audience during the final panel discussion. Mr Stelios Piperidis and Mr Markus Foti were invited to answer these questions.

After the conclusion of the discussions, participants were invited to go on with discussion in an informal environment, during the second coffee break.

4 Synthesis of Workshop Discussions

Some of the participants in the workshop pointed out that the public institutions have a huge amount of data but they cannot provide it on their own decision. They can share their resources only after a sanction from a higher representative of the political cabinet in their organization or after a formal order from the Minister.

Attention was drawn to the considerable terminology datasets available in the field of defence. This data cannot be shared or made publicly available, even under the condition to be used only for the purposes of the CEF eTranslation platform as these data is part of the national defence.

Many participants expressed an interest to use CEF eTranslation platform within their public administration body.

In the panel discussion for Connecting Europe Facility there were several questions concerning open data portal, open data policy and the eDelivery system.

Some participants who had attended the first ELRC workshop two years ago shared their experience with the ELRC repository. Several participants contacted the organizers during the breaks and shared problems with uploading resources. They were invited to try to upload them once again as the tool is much more developed.

The question of the quality of the provided language resources has been raised from the audience on different occasions. There was a question whether it would be appropriate to provide translations automatically generated by CEF eTranslation and proofread and manually edited by experts.

Ms Dobrova informed the participants that they would receive detailed information on the possibilities of using the automatic translation platform and how to get access to the platform translation services. Mrs Dobrova added that at the beginning of September 2018 the Ministry of Transport, Information Technology and Communications requested authorisation to register the domain “@mtitc.government.bg” and received it within one day. All the MTITC users who have EU Login can use the platform now.

4.1 ELRC and Open language Data in Bulgaria

When presenting the legal framework for sharing language resources, special attention was paid to the Directive on the reuse of public sector information, the legal framework on the reuse of information and in particular the [Access to Public Information Act](#), which introduced the Directive through a special regime substituting the older access to information regime.

In her presentation, Ms Rogova pointed out that the sharing of language resources can be linked with the activities on National Open Data Portal which now mainly collects numerical datasets, but not language resources.

One important note was that according to the national legislation, the reused information in the public sector should be in open and machine-readable format.

The new National Open data portal will be available soon. The big difference with the old one is that the new portal would allow registration of civil and business organisations that have their own datasets

they wish to share, thus expanding the range of datasets hosted in the portal. It is expected that the new portal will be launched at the same address: <https://opendata.government.bg/> and it is foreseen to be bilingual and maybe multilingual in the future.

4.2 Success stories and lessons learnt

The following points can be considered success stories for Bulgaria in the context of ELRC:

- We are confident that the widespread dissemination of information on the ELRC workshop as well as the presentation of the ELRC project objectives will lead to an increase in the number of shared resources from and into Bulgarian. The high level of officials who attended our national workshop: The video message sent by EU Commissioner for Digital Economy and Society, Ms Mariya Gabriel, was a proof of the importance of sharing language resources, supported by the Deputy Prime Minister of the Republic of Bulgaria, Mr Tomislav Donchev, in his welcoming speech. ELRC was also supported by the Minister of Transport, Information Technologies and Communications, Mr Rossen Jeliazkov, the Deputy Minister of the Bulgarian Presidency of the Council of EU 2018, Ms Monika Panayotova and the Head of the European Commission's Representation in Sofia, Mr Ognian Zlatev.
- Sending invitations to all central administrations widely promoted the initiative, e.g. the need for a large number of language resources in order to train the eTranslation platform from and into Bulgarian.
- The good organization of the workshop, the serious preparation, the choice of appropriate speakers, moderators and presenters.
- Some of the participants have already provided information about the links to their organisations where suitable materials could be found.
- Broad interest in the event was also manifested by the event's media coverage. As an example, the ELRC PS-NAP was invited to give an interview for the national radio. (See also Appendix 8.3.)

Some of the major conclusions from the Bulgarian workshop are the following:

- Some contributors believe that copyright issues can be a major problem in sharing language resources.
- The Bulgarian National Anchor Points are planning to reach out to the municipal authorities in Bulgaria, which also have appropriate materials for the needs of the eTranslation platform.

5 Workshop Presentations

All presentations are available on the MTITC website (<https://www.mtirc.government.bg/bg/category/237/vtoriyat-nacionalen-workshop-za-spodelyane-na-ezikovi-resursi-premina-s-golyam-interes>) and on the ELRC website: (http://www.lrc-coordination.eu/12bulgaria_agenda).