

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



Deliverable D3.2.27

Task 8

ELRC Workshop Report for Latvia



Contents

<u>1</u>	<u>Executive Summary</u>	<u>3</u>
<u>2</u>	<u>Workshop Agenda</u>	<u>4</u>
<u>3</u>	<u>Summary of Content of Sessions</u>	<u>6</u>
3.1	Welcome and introduction	6
3.2	Connecting public services across Europe: ambition and current results	7
3.3	National initiatives for Language technologies	10
3.4	Overview of the European Language Resource Coordination action	13
3.5	National LT infrastructure - HUGO.LV	15
3.6	CEF eTranslation platform	18
3.7	Language technologies and language data in national administration	21
3.7.1	Volume of translation services in public procurement	21
3.7.2	Panel discussion	23
3.8	Official ceremony for acknowledging the contribution of large LR datasets by HUGO.LV	28
<u>4</u>	<u>Workshop Participants</u>	<u>31</u>
<u>5</u>	<u>Workshop Evaluation – Participant Feedback Analysis</u>	<u>32</u>
5.1	Participant Feedback Forms	Error! Bookmark not defined.
<u>6</u>	<u>Dissemination and publicity</u>	<u>33</u>
<u>7</u>	<u>Appendix</u>	<u>40</u>
7.1	Localised agenda	40
7.2	Full participants list	Error! Bookmark not defined.

1 Executive Summary

The second European Language Resources Coordination Activities (ELRC) seminar in Latvia took place in Riga on September 26, 2019. The seminar was attended by 37 participants - language and translation specialists, officials who use foreign languages in their work, as well as experts from the European Union. Participants discussed the use of automatic translation by public administration, defined daily work requirements, and identified relevant sources of multilingual data, as well as reviewed technical and usability issues related to the use of data for automated translation.

In his welcome speech Mr. Rolands Lappuķe, the advisor on Smart Technology to the President of Latvia, affirmed the need for a political will to increase support for the development of automatic translation and promised the involvement and support of the office of the President.

Representatives from the European Commission and the ELRC consortium presented developments in automated translation services at the European Commission, their benefits for public administrations and the importance of multilingual data to improve the quality and usability of machine translation. They encouraged participants to contribute their data to further develop the CEF eTranslation building block.

Representatives from key language policy and eGovernment institutions discussed various aspects related to the automation of translation work in the public sector, management of language data, and national initiatives in language technologies. In particular, participants stressed the synergy and collaboration of ELRC with the Latvian national language technology infrastructure HUGO.LV.

The ELRC seminar culminated with a formal ceremony during which Ms. Eva Juhņēviča, the Parliamentary Secretary of the Ministry of Culture, officially contributed the open language data generated by the Latvian state administration language technology platform Hugo.lv to the European Commission.

This contribution demonstrates the successful cooperation between Latvia and the European Commission in the development of automated translation systems. This cooperation promotes equal opportunities for every citizen of the European Union when communicating with and seeking information from government entities.

The seminar received very high evaluation scores with average scores in the range 4.67 – 5.0 on a 5 point scale for all evaluation categories.

The seminar results were extensively disseminated in mass media, on governmental websites, and through social media channels.

2 Workshop Agenda

10:00 - 10:30	Registration
<hr/>	
Opening	
10:30 - 10:45	Rolands Lappuķe , the External Advisor on Smart Technology to the President of Latvia)
	Ainars Freimanis , Programme officer at the European Commission DG CONNECT Unit G3 Accessibility, Multilingualism and Safer Internet
	Andrejs Vasiljevs , ELRC Consortium
<hr/>	
Connecting public services across Europe: ambition and results so far	
10:45 - 11:10	Ainars Freimanis , Programme Officer at the European Commission DG CONNECT Unit G3 Accessibility, Multilingualism and Safer Internet (Presentation)
<hr/>	
11:10 - 11:35	National initiatives for language technologies Andrejs Veisbergs (University of Latvia) (Presentation)
<hr/>	
11:35 - 11:50	Coffee break
<hr/>	
Overview of the European Language Resource Coordination action	
11:50 - 12:15	Andrejs Vasiljevs , ELRC Consortium (Presentation)
<hr/>	
National LT infrastructure - HUGO.LV	
12:15 - 12:40	Jānis Ziediņš , the Centre for Culture Information Systems, (Presentation) (Video)
<hr/>	
CEF eTranslation platform	
12:40 - 13:00	Ainars Freimanis , Programme Officer at the European Commission DG CONNECT Unit G3 Accessibility, Multilingualism and Safer Internet (Presentation)
<hr/>	
13:00 - 14:00	Lunch
<hr/>	

Language technologies and language data in national administration
The volume of translation services in public procurement

Ēriks Ajausks, Project NEC TM, ([Presentation](#))

Panel

14:00 - 14:50

Ainars Freimanis, Programme Officer at the European Commission DG CONNECT Unit G3 Accessibility, Multilingualism and Safer Internet

Gatis Ozols, *Head of Electronic Service Division and the Deputy Head of the Public service department at the Ministry of Environmental Protection and Regional Development*

Māris Baltiņš, Director of the State Language Centre

Inita Vītola, Head of Division at the Latvian Language Agency

14:50 - 15:00

Official ceremony recognizing the contribution of large LR datasets by HUGO.LV ([Presentation](#))

3 Summary of Content of Sessions

3.1 Welcome and introduction

Mr. Andrejs Vasiljevs, the local ELRC representative, opened the second ELRC seminar by welcoming the audience and introducing the key persons, as well as underlining the importance of such workshops where state administration can meet technology developers and linguists.

The opening address was delivered by Mr. Roland Lappuķe, the External Advisor to the President of Latvia on Smart Technology. He reiterated the need for a political will to increase support to the development of automatic translation and promised the involvement and support of the office of the President.

Mr. Ainars Freimanis, Programme Officer at DG CONNECT, underlined the importance of language and language technologies to deliver digital solutions for public administration. “When we know the language, we feel at home everywhere,” he mentioned in his opening speech.



Figure 1 Mr. Andrejs Vasiljevs (Tilde/ELRC) introducing participants



Figure 2 Opening address by Mr. Rolands Lappuķe (the advisor to the President of Latvia)



Figure 3 Opening address by Ainārs Freimanis (DGT/eTranslation)

3.2 Connecting public services across Europe: ambition and current results

Mr. Ainārs Freimanis, Programme Officer at DG CONNECT, began the first session with the presentation "Connecting public services across Europe: ambition and current results". Mr. Freimanis described the importance of linguistic diversity in the EU, which is home to 24

official EU languages and about 60 “regional/minority” languages. Mr. Freimanis emphasized that the discussion on the importance of multilingualism and language technologies for bringing down language barriers in the European Digital Single Market started during the Riga Summit on the Multilingual Digital Single Market in 2015. Since then, many important results have been achieved, which highlights multilingualism as a key enabling factor in the development of the Digital Single Market.



Figure 4 Mr. Ainars Freimanis, European Commission DG CONNECT Programme officer for Multilingualism and Safer Internet

In his presentation he pointed to a few major challenges for the digital market:

- Shoppers shop 6 times more often on websites that are in their language;
- In the digital market, no EU language covers more than 20% of users in their mother tongue;
- Most EU languages cover less than 3% of the EU Single Market;
- 9 out of 10 Internet users browse the web in their own language, if it is possible;
- Only 1/3 of the national government portals offer information in other foreign languages, except English;
- Only about a half of Europe's population is able to hold a conversation in English;
- Language is a major barrier to cross-border growth of businesses;
- 1 out of 4 European SMEs believe that the language barrier is the biggest barrier to business expansion across borders;

- 55% mention “lack of language skills in foreign trade” as a major reason for SMEs inability to export;
- Traditional translation solutions can increase the potential turnover 20 times, but translation costs are still too high for SMEs.

Many of the listed obstacles could be removed by coordinated activities in all member states. A big step towards this goal was made by the European Parliament resolution of 11 September 2018 on language equality in the digital age, highlighting importance of joint action by all stakeholders, be it European Commission, municipalities or research institutions: all are equally responsible and should take immediate action. Mr. Freimanis presented good examples (eTranslation, ELRC actions, role of technologies in development of digital services infrastructures (DSI)) whose development and introduction were supported by the CEF Telecom programme, as well as those examples, which are planned in the Multiannual Financial Framework for 2021-2027.

VALODU RESURSI: panākumu atslēga

- Labāks visu valodu pārklājums
- Jomai/tematikai pielāgota mašīntulkošana
- Vairāk valodu tehnoloģiju rīku

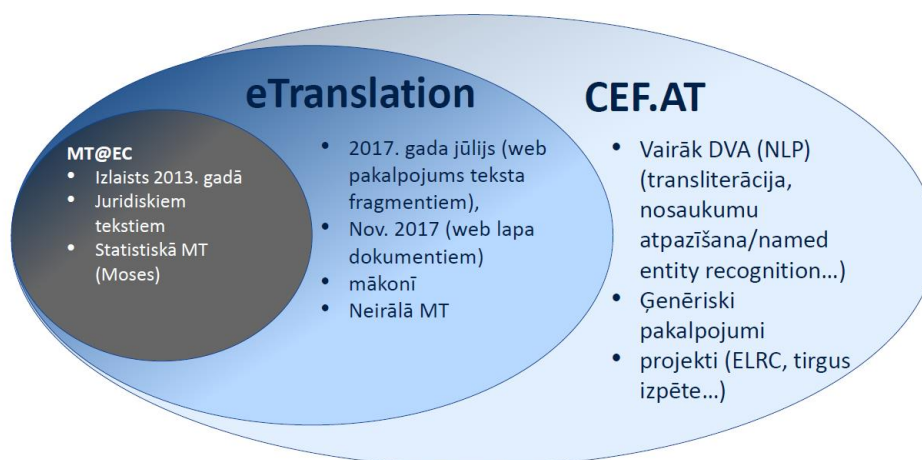
ELRC, Eiropas valodu resursu koordinācijas iniciatīvas ceļā Komisija uzrunā valodu resursu īpašniekus, kas vēlas dalīties ar saviem datiem, lai uzlabotu automatizētās tulkošanas risinājumus.

<http://www.lr-coordination.eu/>



Figure 5 ELRC activities and ELRC-SHARE deployed as part of CEF Telecom core services

CEF AT un eTranslation evolūcija



8



Figure 6 eTranslation evolution

3.3 National initiatives for Language technologies

In his presentation, professor Andrejs Veisbergs presented a view of the language as a “project” with a direct impact on the language itself (written language), Latvian identity and sustainability. He started his presentation by looking back at the development of the Latvian language, namely, the translation of the Bible into Latvian. Ernst Glick from the Lutheran Parishioners Congregation of Alūksne was the first to translate the entire Bible in 1689. We can consider that the beginnings of the written Latvian language. In his presentation, he explained that the development of the written language facilitated by this first translation had a significant impact on continued existence of the Latvian identity and the development of the language.



Figure 7 Professor Andrejs Veisbergs presenting “National initiatives for Language technologies”

He continued with a presentation of the results of a study by Kai Chan “The Power Language Index” as an illustration of how powerful and influential the Latvian language is in comparison to other languages.

five opportunities provided by language:

- ▶ 1. Geography: The ability to travel
- ▶ 2. Economy: The ability to participate in an economy
- ▶ 3. Communication: The ability to engage in dialogue
- ▶ 4. Knowledge and media: The ability to consume knowledge and media
- ▶ 5. Diplomacy: The ability to engage in international relations



Figure 8 Five opportunities provided by language “POWER LANGUAGE INDEX”

The Power Language Index (PLI) is an assessment of the influence of a language on the global stage, and demonstrates the ability of a language to unlock the opportunities provided by

language - geography, economy, communication, knowledge & media, and diplomacy. Latvian was ranked as number 50 of 124 languages.

Latvian 50. Native 7,8 ?

- ▶ 48 Icelandic
- ▶ 49 Bulgarian
- ▶ 50 Latvian
- ▶ 51 Belarusian
- ▶ 52 Vietnamese
- ▶ 53 Azerbaijani



Figure 9 Latvian language in rank of 124 languages

In his presentation, Prof. Veisbergs provided an overview of the various challenges and potential solutions, as well as an insight into some of the important activities that allow Latvian to develop in the digital environment. The advancement of language is linked to a number of important challenges that should be considered when thinking about the development of the language:

- Small number of language users > market > return > investments;
- Demographics and Russian-speaking citizens;
- No comprehensive corpus;
- Few language and computer linguist experts;
- Limited cooperation among EU institutions, Latvian institutions, among themselves;
- Limited availability of data.

At the end of his presentation, Prof. Veisbergs provided a comparison of translations of Samuel Coleridge's words "Language is the armoury of the human mind, and at once contains the trophies of its past and the weapons of its future conquests." The comparison entailed translations of the above by Google translator, Tilde translator, HUGO.lv and back-translation by HUGO.

Tulkojumi

- ▶ "Valoda ir arsenāls no cilvēka prāta, un uzreiz satur tās pagātnes trofejas un tās nākotnes iekarojumu ierocus". **google translate**
- ▶ "Valoda ir cilvēka prāta armoārijs, un tajā uzreiz ir savas pagātnes trofejas un tās nākotnes varoņu ieroči". **Tildes tulkotājs HUGO**
- ▶ The language is an arsenal of the human mind, and it simultaneously contains both the trophies of the past and the weapons of its future conquest. **HUGO**

Figure 10 The comparison of translations

At the end of his presentation, Prof. Veisbergs reformulated Samuel Coleridge's words as "Language resources are an arsenal of artificial intelligence containing past trophies (texts) and weapons of the future."

3.4 Overview of the European Language Resource Coordination action

"62% of Europeans do not speak English well enough to maintain a conversation" and "42% of Europeans never buy products or services other than in their mother tongue": this data from Eurobarometer was used to open the fourth session of the day. This session was presented by Mr. Andrejs Vasiljevs, coordinator of ELRC Data project and CEO of Tilde.

In his presentation, Mr. Vasiljevs gave an overview of the spotlights and progress of the language technology issue, escalated by the META-NET "Language Whitepaper Series". The aim of the study was to assess the current situation with language technology and prospects for future development of all official EU languages. The most worrying conclusion was that more than 20 European languages face digital extinction due to a lack of technological support. This study catalysed a series of activities and projects (Riga Summit on the Multilingual Digital Single Market, Document SRIA Language Technologies for Multilingual Europe, CEF Initiative and many more) which led to an important resolution by the European Parliament, namely, the Resolution of Sept 11, 2018 on "Language equality in the digital age". In the resolution, the European Parliament calls on the Member States to develop comprehensive language related policies and to allocate resources to ensure easier access and use by SMEs and start-ups of language technology (LT) tools and services, as well as all levels of administration to improve access to online services and information in different languages.



Figure 11 Mr. Andrejs Vasiljevs presenting ELRC-SHARE repository

To continue, Mr. Vasiljevs presented the European Language Resource Coordination Action and eTranslation. He explained the ELRC goals and why surmounting language barriers is important for the European Digital Single Market. eTranslation is one of the building blocks of the Connecting Europe Facility program, and, accordingly, the European Commission's intention of creating cross-border digital services that automate and streamline many aspects of public administration activities. Mr. Andrejs Vasiljevs explained how ELRC works, what organisations participate in the initiative and what is the value of textual data for effective and accurate translation tools and services.

3.5 National LT infrastructure - HUGO.LV

This session looked at the use of various language technology tools to cross language barriers in Latvia's e-services. Unlike many other countries, Latvia already has an machine translation (MT) service used by the public sector, making it a true success story in multilingual Europe.



Figure 12 Janis Ziediņš HUGO.LV project manager (Culture information System centre)

Mr. Jānis Ziediņš, HUGO.LV project manager (Culture Information System Centre), presented the HUGO.LV platform and service, main objectives and the reason why such project was initiated. Mr. Ziediņš was proud that this project has grown from a vision to provide equal opportunities to all citizens of Latvia regardless of the language they have mastered to one of the success stories for language technology applications in public administration.

HOGO.LV is a language technology platform, administered by the Latvian State administration that is freely available to everyone. It provides machine translation, an online CAT tool, speech recognition and speech synthesis services, as well as a range of tools for supporting multilingual features in e-services.

Kultūras informācijas
sistēmu centrs

Pirmā platforma Eiropā

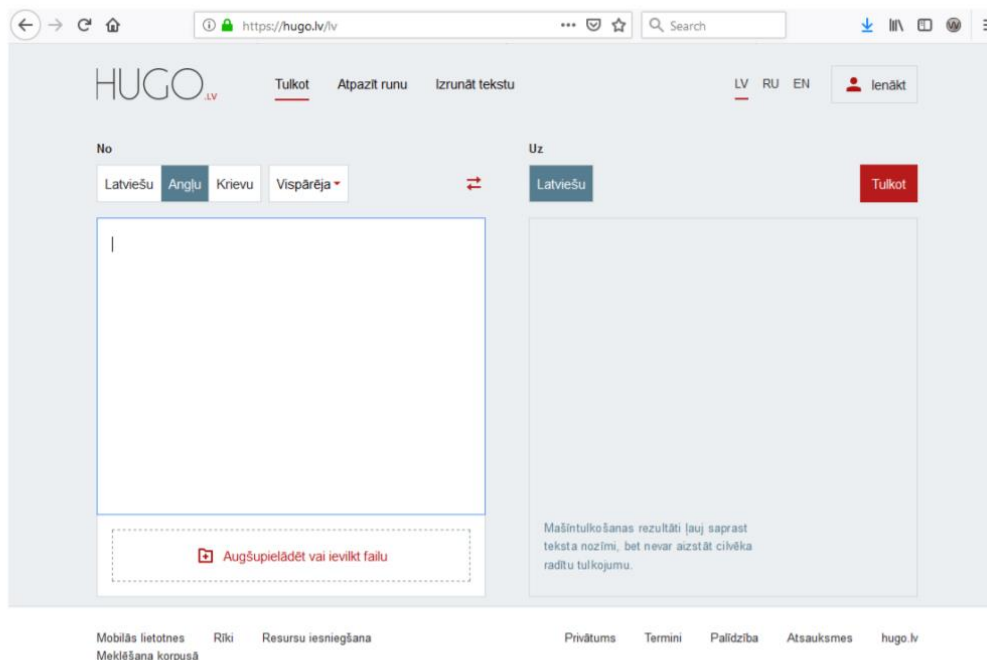


Figure 13 HUGO.LV

Hugo.lv is customized to the Latvian language and state administration documents, thus, its translation quality is significantly higher than other online translation services. Furthermore, users of Hugo.lv can enjoy the services of a translation assistant for more convenient translation.

Statistical data for the first 9 months of the platform usage show that it had processed more than 12 million translation requests and translated more than 212 million words, and that the most used language pairs were EN-LV. All of the above was generated by more than 700 000 visitors.

Mr. Ziediņš presented one of the most recent services offered by the platform – a resource management feature with two main functionalities – “Submit resource” (TMX) and “Search

the database". This service collects and manages translation memories owned by state institutions in one place and trains HUGO.LV systems with the latest datasets.

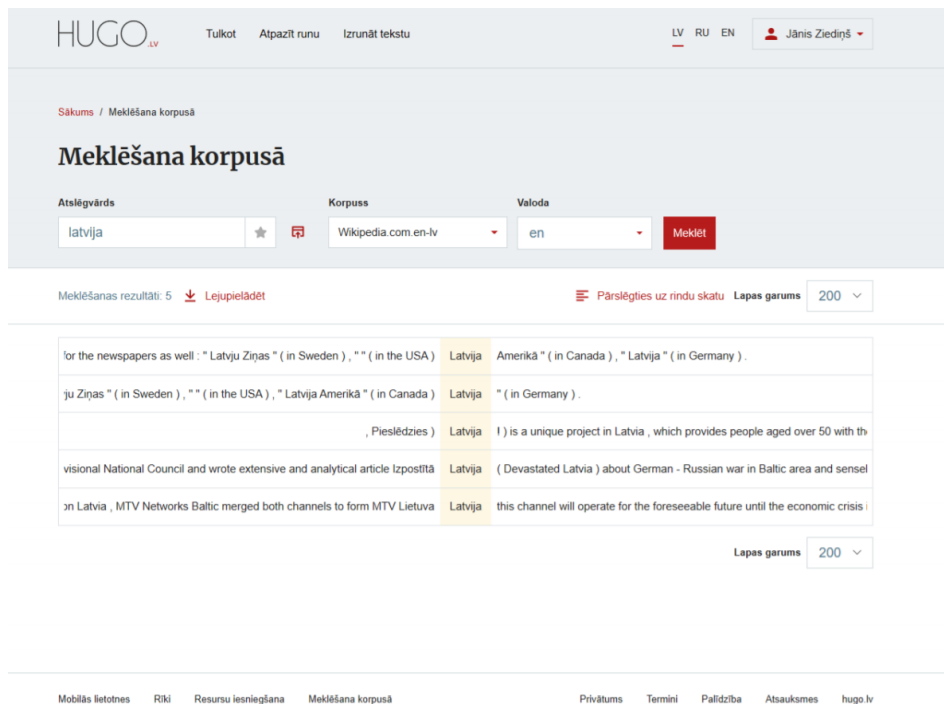
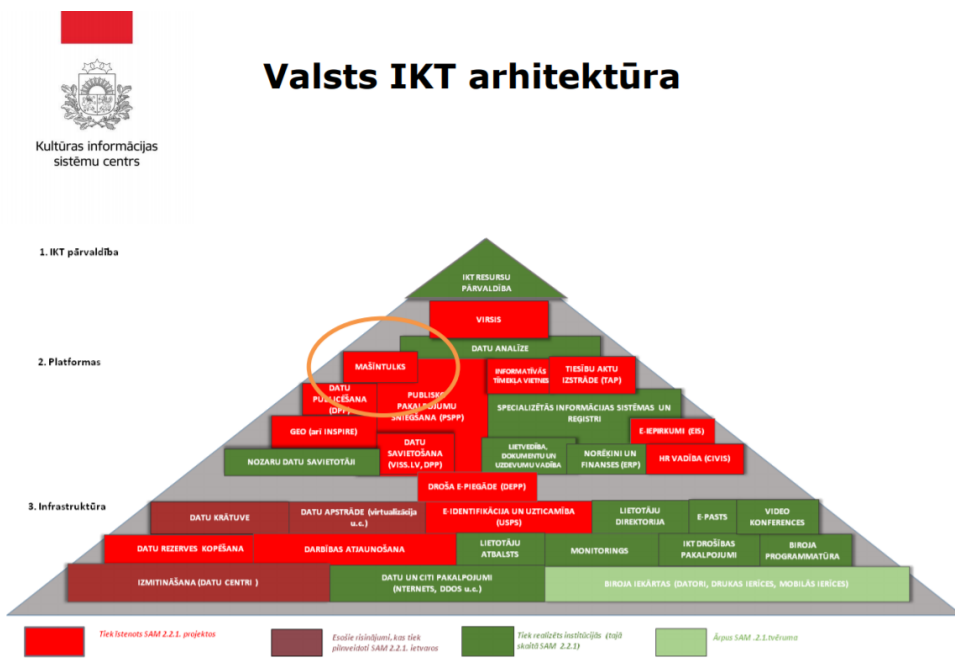


Figure 14 Search feature in TMX repository - service by HUGO.lv

At the end of his presentation, Mr. Ziediņš explained that the Machine translation service and Machine translation platform holds a stable position in the State ICT system.



http://www.varam.gov.lv/lat/darbibas_veidi/e_parv/valsts_ikt_arhitektura/?doc=25467 1
2

Figure 15 Place of Machine translation platform in hierarchy of State ICT infrastructure

3.6 CEF eTranslation platform

Mr. Ainars Freimanis presented the eTranslation platform and services. eTranslation is an MT platform with two major user groups – EC internal tools to support DGT translation teams and a CEF Building block and an MT infrastructure for all public administrations in all CEF Countries, as well as in Norway and Iceland.



2017. gada 15. novembris

**Connecting Europe Facility
modulis (Building Block)**

Tulkošanas ĢD un CNECT

<https://webgate.ec.europa.eu/etranslation>

Figure 16 eTranslation as CEF Building Block officially launched on 15 November 2017

In his presentation, Mr. Freimanis gave insights into accessing the eTranslation platform and using the eTranslation platform for translation purposes. Mr. Freimanis explained that translations could be completed using either the online interface for translation of text snippets, or by uploading documents to be translated and receiving the translated documents back by email or downloading within in 24 hours. This offline feature provides an opportunity to translate large text files. The multilingual eTranslation service is available to all EU, Icelandic and Norwegian governing bodies in order to provide multilingual support for translation from/to 24 official EU languages, and the Icelandic and Norwegian languages.

Main services offered by eTranslation platform are:

- Translation of documents and text snippets;
- Translation of multiple documents at the same time;
- Translation into multiple languages at the same time;
- Neural machine translation engines for many languages;
- Domain tailored systems.

Currently, the eTranslation platform is used by 1 572 active users from various public administrations in EU, with 9 users in Latvia.

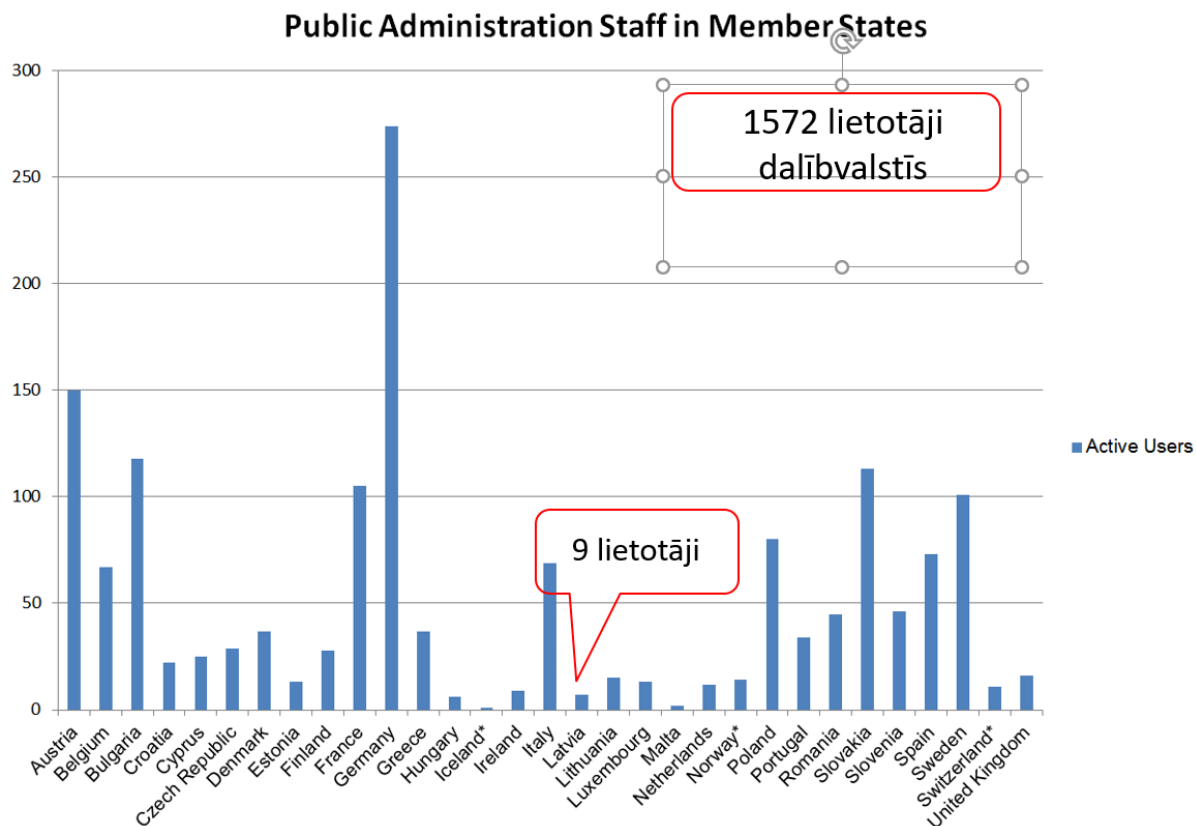


Figure 17 Users of eTranslation in CEF countries

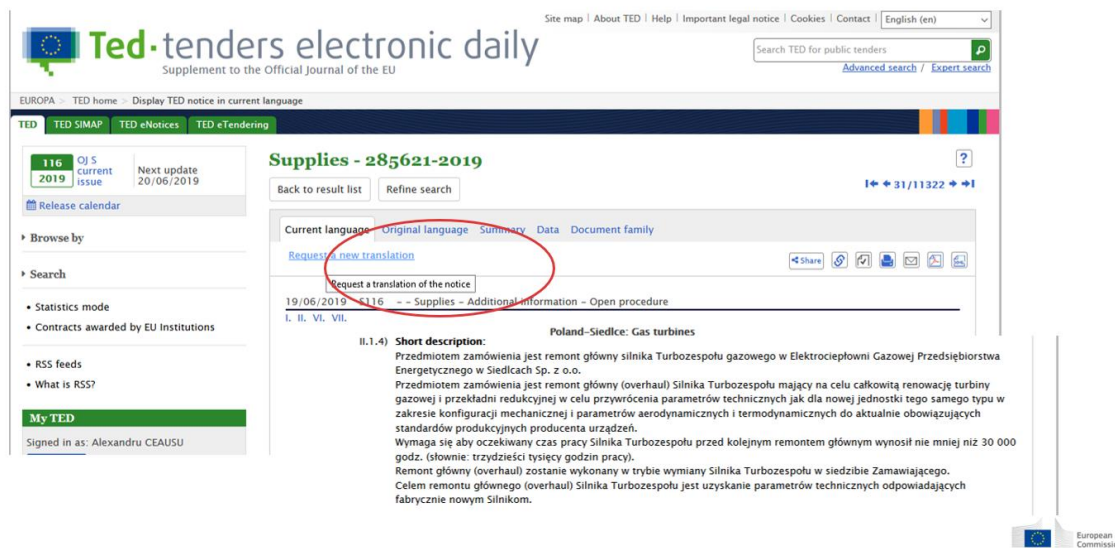
There are institutions with close cooperation with eTranslation. These partners provide data and receive tailored systems for their use:

- Ministry of Finance (France)
- Federal Bank (Bundesbank, Germany)
- Language service (Bundessprachenamt, Germany)
- Prime Minister's Office (Finland)

A number of services play an important role in cross-border cooperation. These are a few examples mentioned,

- Danube River Metadata
- eProcurements (Slovakia)
- Network of the Supreme Courts of the Union
- Railway regulatory

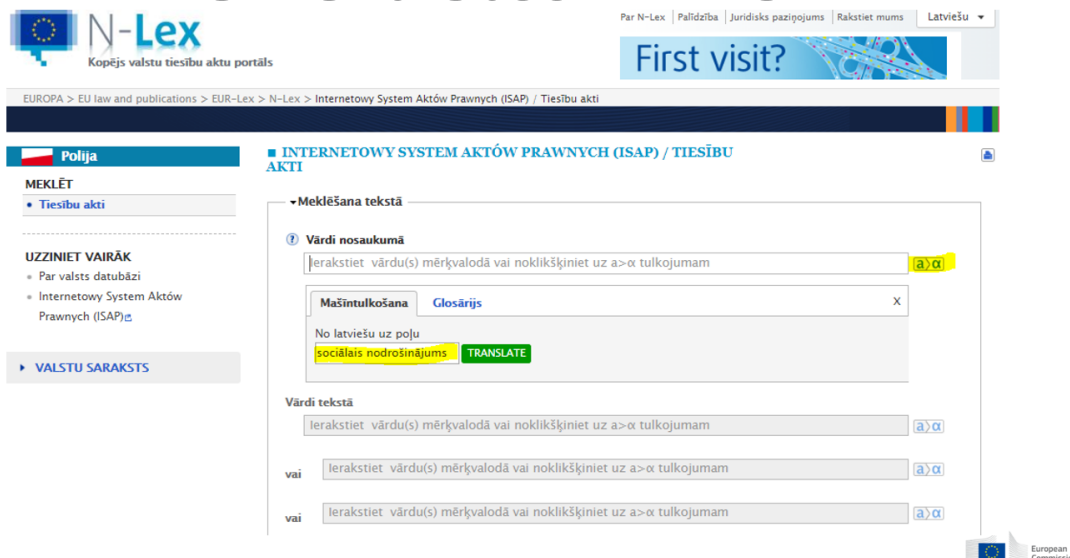
CEF eTranslation in TED



The screenshot shows the TED website interface. At the top, there's a navigation bar with 'Ted tenders electronic daily' and 'Supplement to the Official Journal of the EU'. Below that, there are tabs for 'TED', 'TED SIMAP', 'TED eNotices', and 'TED eTendering'. The main content area displays a tender notice for 'Supplies - 285621-2019'. A red circle highlights the 'Request a new translation' button in the 'Current language' section. The notice details include the date '19/06/2019', the tender number '116', and the title 'Poland-Siedlce: Gas turbines'. The description is in Polish, mentioning a gas turbine overhaul.

Figure 18 The illustration of example of eTranslation in Tender electronic system

CEF eTranslation in N-Lex



The screenshot shows the N-Lex website interface. At the top, there's a navigation bar with 'N-Lex' and 'Kopējs valsts tiesību aktu portāls'. Below that, there are tabs for 'Par N-Lex', 'Palīdzība', 'Juridisks paziņojums', 'Rakstiet mums', and 'Latviešu'. The main content area displays the 'INTERNETOWY SYSTEM AKTÓW PRAWNYCH (ISAP) / TIESĪBU AKTI' section. A red circle highlights the 'TRANSLATE' button in the 'Vārdi nosaukumā' section. The interface includes search bars and a list of legal acts.

Figure 19 The illustration of hoe eTraslation is available in N-Lex system

In 2018, eTranslation users translated 41 million pages (on average, 2 million pages are translated by professional translators from the Commission's Directorate-General for Translation).

Mr.Freimanis mentioned several important advantages of the platform in comparison to publicly available systems:

- High quality machine translation: rooted in decades of EU translation. Excellent translation of EU policy-related and legal documents.
- Confidential and secure: Intellectual property rights belong to the owner of the original text sent for translation.
- Service quality: Quick translation into multiple languages at the same time.
- Formats: Compatible with a series of formats (docx, xlsx, pdf, etc.).
- Financial support to the CEF for cross-border projects is available.

At the end of his presentation, he reminded participants that each of them could take part and help to improve eTranslation by joining ELRC. Through the European Initiative for the Coordination of Language Resources (ELRC), the Commission is seeking language resource holders who want and are able to supply data to improve translation systems.

Mr. Freimanis finished his presentation with an example from the eTranslation system that shows that the extensive base of official terminology used by eTranslation while training MT systems facilitates more accurate translation than other publicly available MT systems. As an example he showed the translation of the phrase “lesser devil ray” from English to Latvian. eTranslation provided the most accurate translation of the name of this species of fish



Figure 20 Translation of - lesser devil ray

3.7 Language technologies and language data in national administration

3.7.1 Volume of translation services in public procurement

As an introduction for the panel discussion, information about the volume of translation subcontracts in public administration was presented by Mr. Ēriks Ajausks (Tilde). Mr. Ajausks provided insight into the most recent findings of the study done in NEC TM project.

In order to understand the approximate public spending for translation services, the NEC TM consortium working on market research carried out an EU-level exercise to assess the expenditure for translations outsourced by public authorities in each member state.

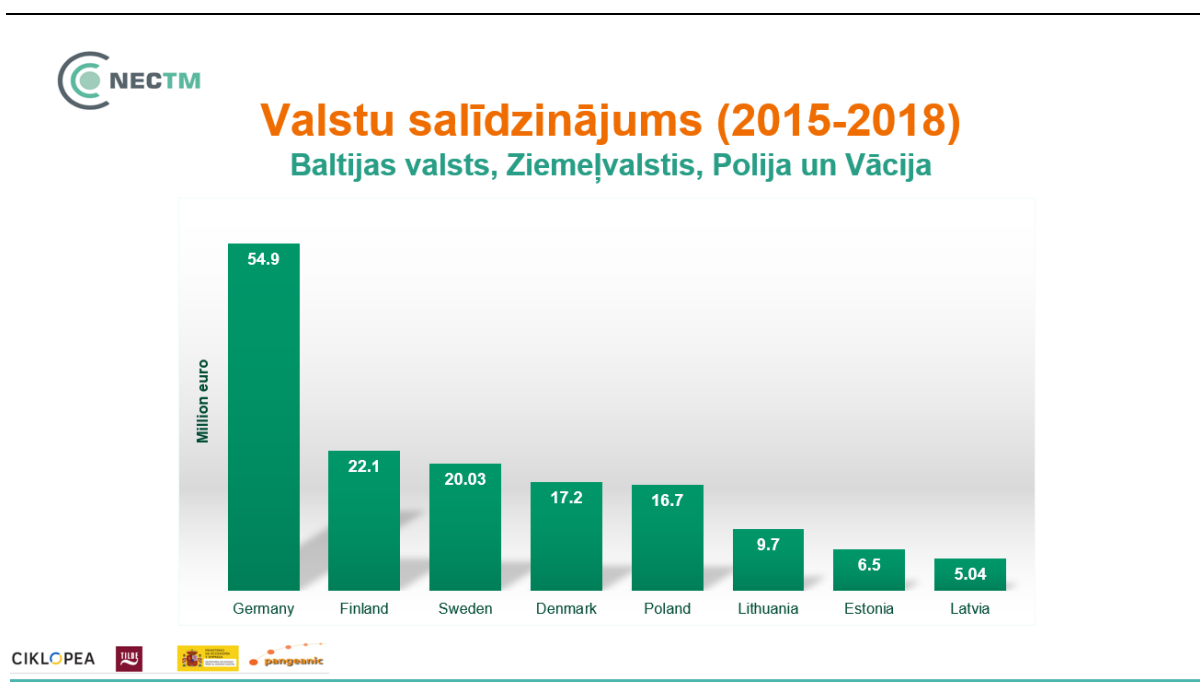


Figure 21 EUR spent on translation subcontracts from 2015 till 2018

Below are the main findings presented to the audience:

- EU Member States and EU institutions outsourced translations for up to €800 million euro from 2015-2018.
- Analysing total expenditure over the four years (2015-2018) in Baltic countries (Latvia, Lithuania, Estonia), we concluded that around 5 million euro were spent on translation services in Latvia, 6,5 million in Estonia and approximately 9,76 million euro in Lithuania.
- By analysing translation procurement in Latvia in greater detail over the last four years and by classifying domains according to *EuroVoc*, it can be concluded that the largest translation costs in Latvia are incurred in the following domains: law (30%), international relations (18%), finance (18%), politics (13%), social questions (2%), education and communications (2%).

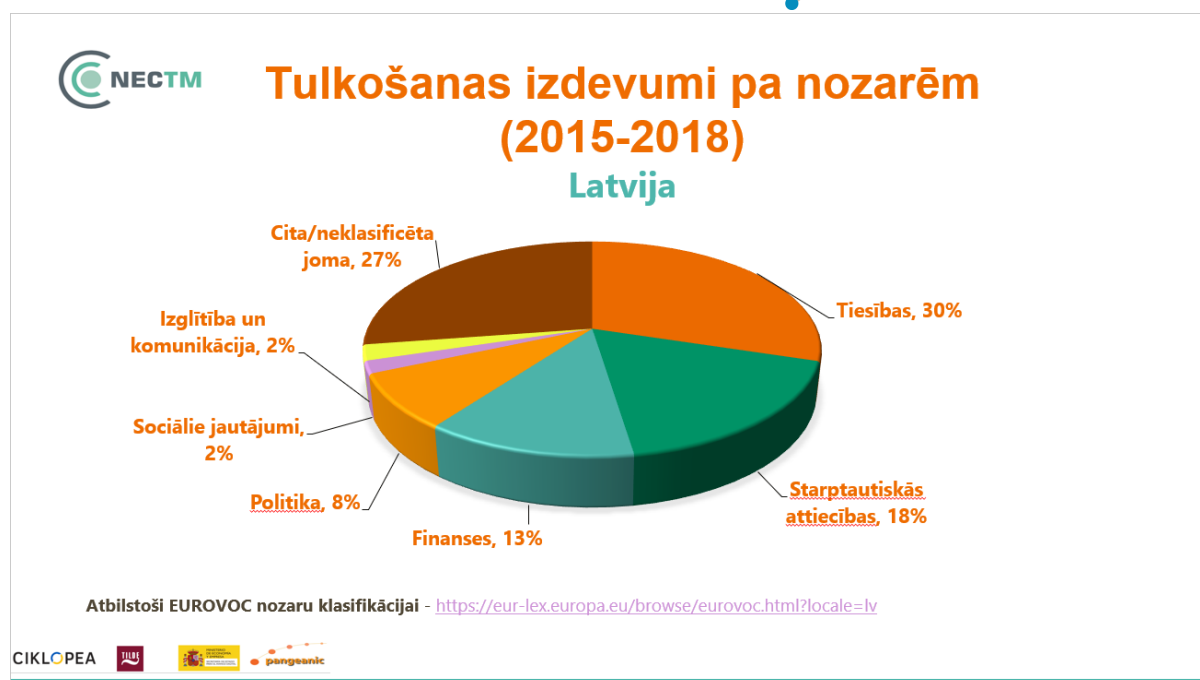


Figure 22 Chart illustrating distribution of translation tasks by domain, largest proportion go to translations related to law/home affairs

Some policy recommendations on how to improve translation procurements and reduce translation costs nationally (and on the EU level) were provided at the end of the NEC TM dissemination presentation:

1. Create industry-specific repositories (repositories) of translation memories and terminology so that resources can be managed in a uniform way.
2. Consolidate engagement and industry-wide translation purchases.
3. Maximize the use of automated translation tools that reduce translation costs by more than 30%.
4. Ensure that outsourcing of translation services also includes translation memories.
5. Ensure availability of translation memories on public administration translation platforms.
6. Include public translation memories in open data repositories.

3.7.2 Panel discussion

This panel discussion focused on language technologies in contemporary public administration – examining the status of “digital extinction” for Latvian and the technologies that are available to support public administration. The panel consisted of four speakers: **Mr.Ainars Freimanis** (Programme Officer at DG CONNECT), **Mr.Gatis Ozols** (Deputy Head of the Public Service department, Head of the Electronic Service Division, Ministry of Environmental Protection and Regional Development), **Mr.Māris Balčiņš** (Director of the State Language Centre) and **Ms.Inita Vītola** (Head of Division at the Latvian Language Agency). The panelists represented different institutions covering all stakeholders related to language technologies in public administration.

The panel was moderated by Mr.Andrejs Vasiļjevs.



Figure 23 Panelists (from left): Ainars Freimanis, Gatis Ozols, Inita Vītola, Māris Baltiņš, Andrejs Vasiljevs.

The first question was addressed to Mr. Gatis Ozols - how does language and language technologies fit into the overall national eServices Development Strategy?

Gatis mentioned the following key factors that are directly linked to language technologies in the development of eServices.

Public administration looks at public services within the framework of the European Digital Single Market. Access to cross-border services is currently a priority in the framework of state services; state services are no longer just local services. They are services that must be provided across the borders of the European Union, so that residents of other countries can also receive services from Latvian authorities and vice versa, i.e. Latvians can receive equivalent services in other countries. And in this respect, language technologies play a key role in the accessibility of services.

Emerging technology trends offer vast opportunities: artificial intelligence and new methods in the field of language technologies open up new opportunities to ensure greater access to services for people with disabilities. The services can be implemented at a new level. A practical example is the new website accessibility directive, which requires that content on a website be accessible to all. Such condition often includes the requirement that video materials must be provided with subtitles and you should be able to listen to the written text. Without language technologies, this would be difficult to supply. It is a time-intensive and financial resource-intensive process. Language technologies allow this process to be simplified and automated and can help to optimize costs.

Another trend and the role of language technologies within the common framework of public administration e-services is the opportunities stemming from new methods and machine learning techniques that allow improvement of productivity in the delivery of services and

customer support. A good example is chatbots and virtual assistants already employed in public administration in customer support services.

In cross-border cooperation, it is critical to ensure that information about products and services is provided in multilingual. Often companies don't have the resources to create product and services descriptions in many languages. Public administration can create preconditions for promoting content accessibility in many languages through the use of the latest technologies. The solutions created by public administration can be further exploited for the development of services and thus multilingual content contributes to cross-border cooperation.

All the points listed above are relevant and will be incorporated into digital development strategy for Latvia, explained Mr. Ozols.

The second question was about the open data policy, that mainly refers to statistical numerical data. Mr. Ozols was asked how language data fits into this open data policy framework.

Language data is considered to be one of the dimensions of open data. The Culture Information Systems Centre has already published data as open data on Latvia's open data portal. It should be acknowledged that the State Administration is currently focusing on opening certain priority datasets. The general framework for data opening is established and an obligation to open data remains, but often such a general framework is not sufficient. Technical expertise and knowledge of data opening requirements are often lacking. In order to address these issues and to identify more efficient patterns of action, an open data strategy for Latvia was published in August. In the development of this strategy, public administration and industry worked together with the main aim of improving data opening procedures and data availability at large. An important part of this process is the identification of priority datasets and establishing of best practices for opening of priority datasets. As part of this process, the industry has appealed to the Ministry of Environmental Protection and Regional Development to open the translation memories in the possession of the State Language Centre and translations at the disposal of the State administrative authorities. However, it should be acknowledged that opening public administration translations is not a simple process and involves a number of measures, such as changing procurement policies, translation and data administration habits in public administrations. It is an important task for the ministry and industry to establish best practice and create preconditions to make language data more accessible and available as open data.

The State Language Centre Director **Mr. Māris Baltiņš** was asked about the perspective of the National Language Centre, how the process of promoting access to language data is assessed and what else could be done in this process?

The textual data is content generated by all public administrations and this is perceived as self-evident and is not considered as data or even more as open data. It is essential to understand how public administration data is created. In his response, Mr. Baltiņš referred to a first session of the panel on public administration subcontracts and a conclusion or recommendation to create a specific repository for language data (translation memory and terminology). He agreed that there may be local data storage for translation memories, but terminology data must have one single terminology repository to ensure uniform use of

terms, regardless of who uses it – translators or public employees. Such a common solution for terminology management has already been created in cooperation between the National Language Centre and the Cultural Information Systems Centre. If more than one storage of terminology data exists, there will be a problem to synchronize terms. As an example, he mentioned translations made for NATO purposes, where it was particularly important to maintain the fairness and uniform use of terminology, regardless of who carried out these translations. Therefore, one single terminology database played an important role in this process.

As far as translations as data are concerned, it should be noted that this cannot only apply to public administration translations, but we also need to look more broadly at how the process is organized between publishers and libraries. The National Library of Latvia plays an important role in this process, which would have an interest and the need to ensure that books that have received State funding are delivered as electronic versions to the National Library and researchers. Of course, all aspects and circumstances must be considered when establishing such a system, so that economic operators are not harmed.

If we look at the process of opening up translation memories, it should be stressed that the translation memory of the National Language Centre will not be the only one that will resolve the availability and openness of language data from the public administration. In this context, we must avoid the wish to create local translation memories within one institution. We need to look at the process and aims to create a single national translation memory.

What is the roadmap to ensure that there is a single term database and how can public administrations be involved in this process and also in the process of creating new terms? This question was asked by moderator Mr. Andrejs Vasiljevs, referring to the huge work done under Mr. Baltiņš leadership to ensure the exchange of terminology data, the correctness of terms and the creation of a single term database.

Mr. Baltiņš outlined a number of important aspects that need to be considered when organizing the process of collecting terminology. It is essential to understand that the term repository is the place that aggregates previously created terms. The term repository does not create terms. Accordingly, there should be a clear strategy on how to invest in targeted identification of areas that would require a review of terminology and its accuracy, and clarification as needed. A strategy to ensure that the identification and updating of these terms would be carried out by specialists in the field. Dialogue between scientists, field experts, language experts and industry experts needs to be promoted, in order to develop precise and correct terminology.

It is necessary to establish such a strategy to promote the collection and availability of correct terminology through a single term repository and to prevent the accumulation and use of incorrect terminology data sets that are not specialist-designed and do not correspond to the specific nature of the field.

The terminology aspect is a significant factor if we talk about opening translations of public administrations, because often translation is done without using the correct terms. Using such data facilitates the use of incorrect terms and, later, when solutions are created, these solutions contain wrong terminology and generate new data in which incorrect or inappropriate terminology is used.

The next question was addressed to the Latvian Language Agency, which works with language issues every day, both through the consultation service desk and by promoting different projects. What challenges and opportunities do you see, what can be provided by language data and language technologies?

The Latvian Language Agency is less involved in translation processes, although there are also projects to translate linguistic literature from foreign languages into Latvian and vice versa.

If it concerns updates and the addition to any database, the question here is not only whether the work created has been financed through State funding, but there is also a question of copyright. All works that have been used by the agency are copyrighted and the Latvian Language Agency has bought the copyright to the respective work.

In regard to Latvian in the digital environment and accessibility with the help of different databases and apps, this issue will always be an important topic. In recent years, the agency has focused more on creating and ensuring the availability of different databases. A good example is the foreign personal name rendition database, which summarizes all current and available information from different sources. This database is used by State administrations and is available through a web interface where you can search for words and see the correct rendition in Latvian. The previously presented Hugo.lv platform for public administration is an excellent means of making such a database available to a large number of users. This service enables specialists to see the most frequently used rendition and the principles applied. It is important to emphasize that when creating these databases, they are created using databases or dictionaries compiled by specialists and such works are protected by copyright. This prevents such collections from being opened freely, even though the database has been created using public funding. The explanation was provided by **Ms. Inita Vītola**.

The whole audience was involved in the discussion. From the audience, a question about the accessibility and use of eTranslation was raised, as well as whether the creation of specific tools and solutions could be supported by the European Commission. As an example, a questioner mentioned the creation of a machine-translator that would be needed for a specific pair of languages and in a specific field, for example between the Estonian and Latvian languages.

In response to this question, Mr. Ainars Freimanis recalled that eTranslation could be used by all public entities and their employees. He stressed that eTranslation can be used and tested to understand whether eTranslation helps to address those challenges that the institution is willing to solve. Mr. Freimanis also explained that there are different funding programs for different cross-border projects, as well as for integration of eTranslation into different solutions. He explained that an eTranslation support service is available and can provide answers to a variety of questions, ranging from who can access the system to what technical parameters are to be considered for integration of the eTranslations system into specific solutions for public administration.

From the audience side, a question was asked to all of the panelists. The questioner stressed the need for a variety of changes, including regulatory adjustments, in order to establish a system that ensures that derivatives of translations (e.g. terminology glossaries, TMX files etc.) are returned. Questioner drew attention to the fact that the volume of translations is measured in millions of euros annually and that the derivatives of these translations are not

returned to public institutions. In such circumstances, derivatives have been lost as an open data object and as a means of creating a knowledge base for the commissioning parties themselves and promoting the creation of a uniform translation memory for the public administration. In this context, a question was raised to all panel participants: what tools and legislative initiatives should be created to initiate the process of returning these derivatives to the commissioning party?

Mr. Baltiņš said that for more than 15 years he has been working on issues that have focused on arranging a regulatory base that governs the field of terminology. This situation can be addressed through specific Cabinet regulations to regulate how such orders are carried out. In addition, a broader view should be taken, and at the same time, it is also necessary to address related issues, such as copyright to certain types of article works, which should be included in different databases and a uniform Latvian language body.

Ms. Vitola stressed that the issue of copyright is crucial for the opening of existing databases held by them. Her institution may open data of which they are the authors themselves or the opening of which has been agreed with the other institutions. However, there are also different situations where copyright is difficult to resolve, despite the amount of financial resources available.

In response to the question, Mr. Ozols mentioned the existence of a general framework on data opening and the obligation to open data. However, if such processes are not controlled, they result in chaos. This creates a situation where some datasets are opened, others are not and there even are situations where the databases themselves do not meet a uniform standard, because there has been no common understanding about what is open and how this is to be done. This is all about a clearly defined management process at national level and the level of the institutions, which then determines who does it, how it is done, who are acting recklessly and what the standards are. And it is also important that there is one responsible authority that is able to put it all together and monitor it. Latvia has an open data portal that monitors the process of opening these data and this is already a step in the right direction. In order to facilitate the opening of language data, the field specialists and the holders of this open data portal should meet and should go to ministries and explain the processes, the dimensions of the data and their relevance. It is necessary to make this process more controlled and comprehensible to data holders.

3.8 Official ceremony for acknowledging the contribution of large language resource (LR) datasets by HUGO.LV

The workshop culminated with an official ceremony where Ms. Eva Juhņēviča, the Parliamentary Secretary of the Ministry of Culture, contributed the open language data generated by the Latvian state administration language technology platform Hugo.lv. The open language data deliverable includes **318** million words of monolingual corpora, **15** million words of parallel corpora, and **19 000** Latvian terms.



Figure 24 The Parliamentary Secretary of the Ministry of Culture, Ms.Eva Juhņēviča, contributed the open language data to European Commission.



Figure 25 From left: **Ainars Freimanis (EC)**, **Eva Juhņeviča (Ministry of Culture)**, **Jānis Ziediņš (Culture information system centre)** and **Andrejs Vasiļjevs (ELRC Lot 3 coordinator, Tilde CEO)**



4 Workshop Participants

The workshop was well-attended – there were 37 participants from a wide range of ministries and public organizations, as well as from LSPs and academia. Before the workshop, the cold calling activity took place. In this way, the organizers were able to make sure that important public administration institutions were represented.



Figure 26 Workshop opening

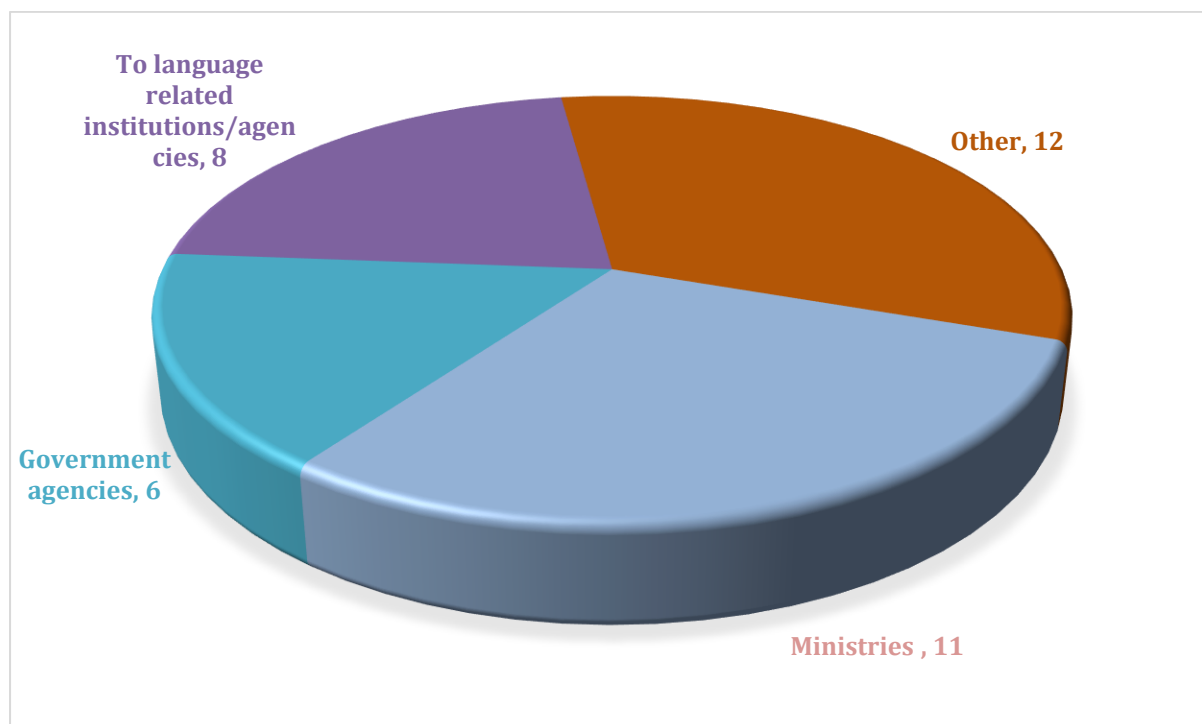
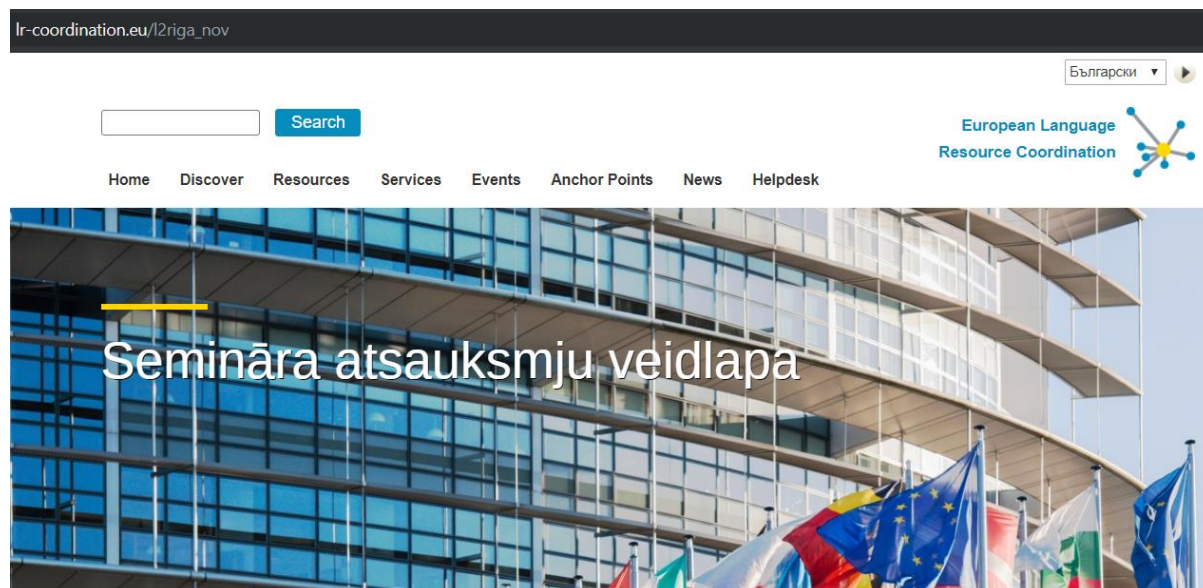


Figure 27 Participant distribution by sector

5 Workshop Evaluation – Participant Feedback Analysis

For the Riga workshop we created online evaluation forms and engagement forms accessible from Ir-coordination.eu Website (http://www.lr-coordination.eu/l2riga_nov).



Forms were accessible from the web site and by scanning a QR code. This allowed the participants to assess the workshop and participate in the discussion in a convenient way.



6 Dissemination and publicity



CEF Digital
Connecting Europe

About Building Blocks Sectors Success Stories Grants Monitoring Contact

CEF Digital > News

Latvia contributes language data for eTranslation



@ELRC

During a formal ceremony at the second European Language Resources Coordination Activities (ELRC) seminar in Latvia, the parliamentary secretary of the Ministry of Culture, Eva Juhņēviča, contributed the open language data generated by the Latvian state administration language technology platform [Hugo.lv](https://hugo.lv) to the representative of the

Latest

News

[Event calendar](#)

[Sectors](#)

[Media library](#)

[Success stories](#)

Follow us



Subscribe to our newsletter

SUBSCRIBE

Figure 28 <https://bit.ly/34d2ksF>



Ministrija

Kultūra

Integrācija un sabiedrība

Fondi un ES politika

Mediju politika

Kontakti

Vadība

Struktūra

Kontakti

Vakances

Budžets

Iepirkumi

Par ministriju

Dokumenti

Ētikas kodekss

Pārskats

Vēsture un ēka

Īpašumi

Pakalpojumi

Latvijas valsts simtgade

Sabiedrības līdzdalība

Jaunumi

KM Jaunumu arhīvs

Kalendārs

Ieguldījumi

Foto

KM iestādes un kapitālsabiedrības

KM logo

Viegli lasīt

Latvija Eiropas komisijai nodod apjomīgu atvērto valodas datu kopumu

27.09.2019

Drukāt

[Tweet](#)
[Share](#)


Eiropas Komisijas pārstāvis Ainārs Freimanis, LR Kultūras ministrijas parlamentārā sekretāre Eva Juhņēviča, Kultūras Informācijas sistēmu centra Hugo.lv projekta vadītājs Jānis Ziediņš, valodu tehnoloģiju uzņēmuma "Tilde" vadītājs Andrejs Vasiļjevs.

26. septembrī Otrajā Eiropas valodu resursu koordinācijas aktivitātes (ELRC) Latvijas seminārā svinīgā ceremonijā LR Kultūras ministrijas parlamentārā sekretāre Eva Juhņēviča nodeva valsts pārvaldes valodas tehnoloģiju platformā Hugo.lv uzkrātos atvērto valodas datus Eiropas Komisijas pārstāvim Aināram Freimanim. Tas ir būtisks Latvijas labas gribas solis, kas apliecina Eiropas Komisijas un mūsu valsts sekmīgu sadarbību automatizēto tulkošanas sistēmu izstrādē un pilnveidē, tādējādi nodrošinot līdzvērtīgas iespējas ikvienam Eiropas Savienības iedzīvotājam saziņā ar valsts pārvaldi un informācijas iegūšanā.

Atvērto valodas datu nodevums ietver valsts pārvaldes valodu korpusus latviešu-angļu valodās, tai skaitā valodu korpusus tieslietu un kultūras jomās. "Tas būs abpusējs ieguvums gan Eiropas Savienībai, gan Latvijas valstij, jo tādējādi tiks gan papildināta eTranslation platforma ar augstvērtīgiem valodas resursiem, gan nodrošināts

Figure 29 <https://www.km.gov.lv/lv/ministrija/jaunumi/latvija-eiropas-komisijai-nodod-apjomigu-atverto-valodas-datu-kopumu-4244>



LR Coordination
@LR_Coordination

Follow

Happy to see that language data sharing is successfully put into practice in Latvia 🙌
[#languagedatamatters](#)



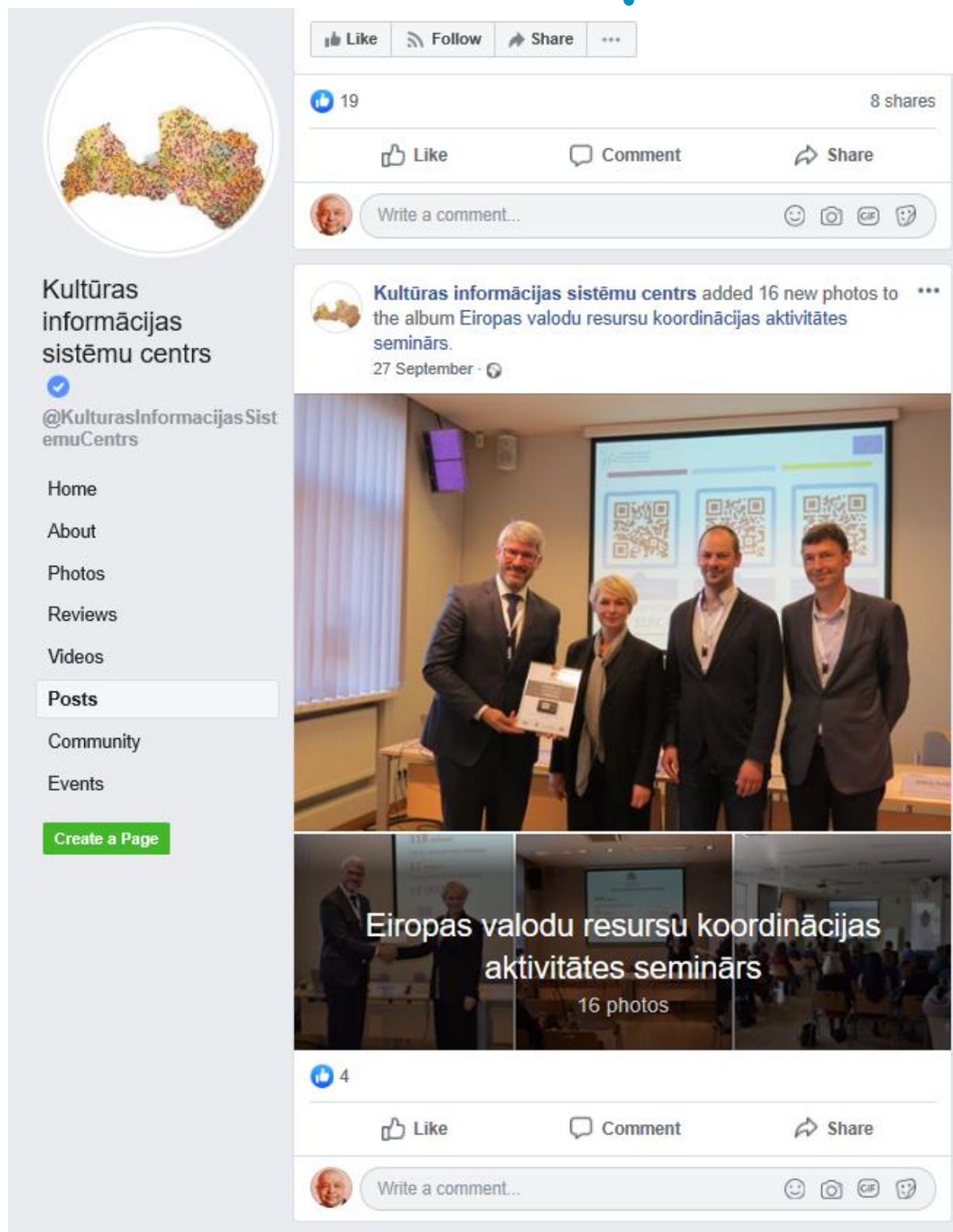
Tilde @TildeCom

At the second #ELRC seminar, the Parliamentary Secretary of the Ministry of Culture of the Republic of Latvia contributed language data to ELRC for #eTranslation.

6:17 AM - 30 Sep 2019

2 Retweets 9 Likes





Like Follow Share ...

19 8 shares

Like Comment Share

Write a comment...

Kultūras informācijas sistēmu centrs added 16 new photos to the album Eiropas valodu resursu koordinācijas aktivitātes seminārs.
27 September · 🌐

Home
About
Photos
Reviews
Videos
Posts
Community
Events
Create a Page

Eiropas valodu resursu koordinācijas aktivitātes seminārs
16 photos

4

Like Comment Share

Write a comment...

Figure 30 <https://www.facebook.com/pg/KulturasInformācijasSistemuCentrs/posts/>



Kultūras informācijas
sistēmu centrs

Par mums ▾

Projekti ▾

Starptautiskā sadarbība ▾

Viegli lasīt

TULKO AR HUGO.LV

[English](#) [Русский](#)

Meklēt ...

Latvija Eiropas komisijai nodod apjomīgu atvērto valodas datu kopumu

26. septembrī Eiropas valodu resursu koordinācijas aktivitātes (ELRC) Latvijas seminārā svinīgā ceremonijā LR Kultūras ministrijas parlamentārā sekretāre Eva Juhņēviča nodeva valsts pārvaldes valodas tehnoloģiju platformā Hugo.lv uzkrātos atvērto valodas datus Eiropas Komisijas pārstāvim Ainaram Freimanim. Tas ir būtisks Latvijas labas gribas solis, kas apliecina Eiropas Komisijas un mūsu valsts sekmīgu sadarbību automatizēto tulkošanas sistēmu izstrādē un pilnveidē, tādējādi nodrošinot līdzvērtīgas iespējas ikvienam Eiropas Savienības iedzīvotājam saziņā ar valsts pārvaldi un informācijas iegūšanā.



Attēlā (no kreisās):

Eiropas Komisijas pārstāvis Ainars Freimanis, LR Kultūras ministrijas parlamentārā sekretāre Eva Juhņēviča, Kultūras Informācijas sistēmu centra Hugo.lv projekta

Figure 31 <http://www.kis.gov.lv/2019/09/latvija-eiropas-komisijai-nodod-apjomigu-atverto-valodas-datu-kopumu/>



Eiropas Komisija

27. septembrī, 2019
Lasišanai: 5 minūtes

RUBRIKA: Relīze
TĒMA: Eiropas Savienība



Latvija EK nodod apjomīgu atvērto valodas datu kopumu



26. septembrī Otrajā Eiropas valodu resursu koordinācijas aktivitātes (ELRC) Latvijas seminārā svinīgā ceremonijā LR Kultūras ministrijas parlamentārā sekretāre Eva Juhņēviča nodeva valsts pārvaldes valodas tehnoloģiju platformā Hugo.lv

Figure 32 <https://lvportals.lv/dienaskartiba/308741-latvija-ek-nodod-apjomigu-atverto-valodas-datu-kopumu-2019>



Follow

At the second #ELRC seminar, the Parliamentary Secretary of the Ministry of Culture of the Republic of Latvia contributed language data to ELRC for #eTranslation.



2:17 AM - 30 Sep 2019

2 Retweets 7 Likes



1 comment 2 retweets 7 likes

PangeaMT @pangeamt · Oct 9

Replying to @TildeCom
Great Dissemination effort also for @nec_tm project!

1 comment 2 retweets 2 likes

Figure 33 <https://twitter.com/TildeCom/status/117859966588702258>

7 Appendix

7.1 Localised agenda

Otrais Eiropas valodu resursu koordinācijas aktivitātes (ELRC) Latvijas seminārs

- 10:00 - 10:30 Reģistrācija
- 10:30 - 10:45 Atklāšanas uzrunas
- Rolands Lappuķe** (Valsts prezidenta ārštata padomnieks viedo tehnoloģiju jautājumos)
- Ainars Freimanis** (EK, CONNECT ĢENERĀLDIREKTORĀTS Komunikācijas tīkli, saturs un tehnoloģijas G3 nodaļa: Pieklūstamība, daudzvalodība un drošāks internets Programmu koordinators)
- Andrejs Vasiljevs, ELRC/TILDE
- 10:45 - 11:10 Sabiedrisko pakalpojumu savienošana visā Eiropā: mērķis un līdzšinējie rezultāti
- Ainārs Freimanis** EK, CONNECT ĢENERĀLDIREKTORĀTS Komunikācijas tīkli, saturs un tehnoloģijas G3 nodaļa: Pieklūstamība, daudzvalodība un drošāks internets Programmu koordinators)
- 11:10 - 11:35 Valodu tehnoloģiju nacionālās iniciatīvas
- Andrejs Veisbergs** (Latvijas Universitātes profesors)
- 11:35 - 11:50 Starpbrīdis kafijai un sarunām
- 11:50 - 12:15 Eiropas *valodu* resursu koordinācijas darbības pārskats
- Andrejs Vasiljevs, ELRC
- 12:15 - 12:40 Nacionālā VT infrastruktūra — HUGO.LV
- Jānis Ziediņš** (Kultūras informācijas sistēmu centrs (KISC))
- 12:40 - 13:00 CEF eTranslation platforma
- Ainārs Freimanis** EK, CONNECT ĢENERĀLDIREKTORĀTS Komunikācijas tīkli, saturs un tehnoloģijas G3 nodaļa: Pieklūstamība, daudzvalodība un drošāks internets Programmu koordinators)
- 13:00 - 14:00 Pusdienas
- 14:00 - 14:50 Valodas tehnoloģijas un valodas dati valsts pārvaldes iestādēs
- Tulkošanas pakalpojumu apjoms publiskajos iepirkumos

Ēriks Ajausks (Projekts NECT TM)

Paneldiskusija

Ainars Freimanis EK, CONNECT ĢENERĀLDIREKTORĀTS Komunikācijas tīkli, saturs un tehnoloģijas G3 nodaļa: Pieklūstamība, daudzvalodība un drošāks internets Programmu koordinators)

Gatis Ozols (Vides aizsardzības un reģionālās attīstības ministrija, Publisko pakalpojumu departamenta Direktora vietnieks, E-pakalpojumu nodaļas vadītājs)

Māris Baltiņš (Valsts valodas centra direktors)

Inita Vītola (Latviešu valodas aģentūra, Daļas vadītāja)

14:50 - 15:00 HUGO.LV valodas datu nodošanas ceremonija Eiropas Komisijai

7.2 Table of Figures

Figure 1 Mr. Andrejs Vasiļjevs (Tilde/ELRC) introducing participants	6
Figure 2 Opening address by Mr. Rolands Lappuķe (the advisor to the President of Latvia) ...	7
Figure 3 Opening address by Ainārs Freimanis (DGT/eTranslation).....	7
Figure 4 Mr. Ainars Freimanis, European Commission DG CONNECT Programme officer for Multilingualism and Safer Internet	8
Figure 5 ELRC activities and ELRC-SHARE deployed as part of CEF Telecom core services.....	9
Figure 6 eTranslation evolution	10
Figure 7 Professor Andrejs Veisbergs presenting “National initiatives for Language technologies”	11
Figure 8 Five opportunities provided by language “POWER LANGUAGE INDEX”	11
Figure 9 Latvian language in rank of 124 languages	12
Figure 10 The comparison of translations	13
Figure 11 Mr. Andrejs Vasiļjevs presenting ELRC-SHARE repository.....	14
Figure 12 Janis Ziediņš HUGO.LV project manager (Culture information System centre).....	15
Figure 13 HUGO.LV	16
Figure 14 Search feature in TMX repository - service by HUGO.lv.....	17
Figure 15 Place of Machine translation platform in hierarchy of State ICT infrastructure	17
Figure 16 eTranslation as CEF Building Block officially launched on 15 November 2017	18
Figure 17 Users of eTranslation in CEF countries	19
Figure 18 The illustration of example of eTranslation in Tender electronic system	20
Figure 19 The illustration of hoe eTrasnlation is available in N-Lex system	20

Figure 20 Translation of - leser devil ray	21
Figure 21 EUR spent on translation subcontracts from 2015 till 2018.....	22
Figure 22 Chart illustrating distribution of translation tasks by domain, largest proportion go to translations related to law/home affairs	23
Figure 23 Panelists (from left): Ainars Freimanis, Gatis Ozols, Inita Vītola, Māris Baltiņš, Andrejs Vasiļjevs.	24
Figure 24 The Parliamentary Secretary of the Ministry of Culture, Ms.Eva Juhņēviča, contributed the open language data to European Commission.	29
Figure 25 From left: Ainars Freimanis (EC), Eva Juhņēviča (Ministry of Culture), Jānis Ziediņš (Culture information system centre) and Andrejs Vasiļjevs (ELRC Lot 3 coordinator, Tilde CEO)	30
Figure 26 Workshop opening.....	31
Figure 27 Participant distribution by sector	31
Figure 28 https://bit.ly/34d2ksF	33
Figure 29 https://www.km.gov.lv/lv/ministrija/jaunumi/latvija-eiropas-komisijai-nodod-apjomigu-atverto-valodas-datu-kopumu-4244	34
Figure 30 https://www.facebook.com/pg/KulturasInformacijasSistemuCentrs/posts/	36
Figure 31 http://www.kis.gov.lv/2019/09/latvija-eiropas-komisijai-nodod-apjomigu-atverto-valodas-datu-kopumu/	37
Figure 32 https://lvportals.lv/dienaskartiba/308741-latvija-ek-nodod-apjomigu-atverto-valodas-datu-kopumu-2019	38
Figure 33 https://twitter.com/TildeCom/status/1178599665887072258	39