



**European Language  
Resource Coordination**  
*Connecting Europe Facility*

## **Deliverable D3.2.28**

### **Task 3**

# **ELRC Workshop Report for Latvia**



<b>Author(s):</b>	Aivars Bērziņš, Andrejs Vasiljevs (Tilde)
<b>Dissemination Level:</b>	Public
<b>Version No.:</b>	V2.3
<b>Date:</b>	2023-01-16



## Contents

<b><u>1</u></b>	<b><u>Executive Summary</u></b>	<b><u>3</u></b>
<b><u>2</u></b>	<b><u>Workshop Agenda</u></b>	<b><u>4</u></b>
<b><u>3</u></b>	<b><u>Summary of Content of Sessions</u></b>	<b><u>6</u></b>
<b>3.1</b>	<b>Welcome and introduction</b>	<b>6</b>
<b>3.2</b>	<b>European Language Equality - Latvian Language in the Digital Age</b>	<b>6</b>
<b>3.3</b>	<b>Future Directions: the Common European Language Data Space</b>	<b>6</b>
<b>3.4</b>	<b>eTranslation: Machine Translation under the Digital Europe Program</b>	<b>7</b>
<b>3.5</b>	<b>HUGO.LV – Language Technology Platform for Latvian State Administration</b>	<b>7</b>
<b>3.6</b>	<b>Virtual Assistants for Support Services</b>	<b>7</b>
<b>3.7</b>	<b>Application of Speech Technology</b>	<b>8</b>
<b>3.8</b>	<b>Machine Translation Solutions</b>	<b>8</b>
<b>3.9</b>	<b>Take-home message and conclusions</b>	<b>9</b>
<b><u>4</u></b>	<b><u>Synthesis of Workshop Discussions</u></b>	<b><u>10</u></b>
<b><u>5</u></b>	<b><u>Country Profile: Language data creation, management and sharing</u></b>	<b><u>11</u></b>
<b><u>6</u></b>	<b><u>Workshop Participants</u></b>	<b><u>12</u></b>

## 1 Executive Summary

The 3rd ELRC Workshop in Latvia took place in the office of the European Commission Representation in Latvia (ES Māja, Aspāzijas bulvāris 28, Rīga) on November 30, 2022. It was organized locally jointly by Tilde and the Culture Information Systems Centre that is the host institution for the public national anchor point of ELRC. The onsite workshop succeeded to summon decision makers and representatives from the key public sector bodies – Parliament of Latvia (Saeima), State Chancellery, Bank of Latvia, Ministry of Education and Science, Ministry of Foreign Affairs, National Electronic Mass Media Council, the State Language Agency, State Language Centre, and others. To further extend the reach of the seminar a Facebook online streaming was provided. The workshop clearly demonstrated some major achievements in adopting language technology in the public sector institutions of Latvia. National language technology platform Hugo.lv plays a seminal role in making LT services and tools easily and freely usable and adaptable for the needs of different institutions. Integration of eTranslation services in Hugo.lv further increases its scope of language support. Representation at the workshop of the key public institutions laid a strong foundation for future cooperation on national and European level in Language Data Space and EDIC initiatives. The workshop also created a broader resonance attracting an interest from the influential political magazine “IR” which published an extensive interview based on the workshop topics.

## 2 Workshop Agenda

12:30 13:00	-	Registration
13:00 13:15	-	<b>Opening addresses</b> <i>Andrejs Vasiljevs, ELRC</i>
13:15 13:35	-	<b>European Language Equality - Latvian language in the Digital Age</b> <i>Inguna Skadiņa, University of Latvia</i>
13:35 14:05	-	<b>Future Directions: the Common European Language Data Space</b> <i>Philippe Gelin, G3 Head of Unit: Accessibility, Multilingualism and Safer Internet, DG CONNECT, EC</i>
14:05 14:25	-	<b>eTranslation: Machine Translation under the Digital Europe Program</b> <i>Christopher Scott, Directorate-General for Translation (DGT) — European Commission</i>
14:25 14:45	-	<b>HUGO.LV – Language Technology Platform for Latvian State Administration</b> <i>Jānis Ziediņš, Culture Information Systems Centre</i>
14:45 15:05	-	<b>COFFEE BREAK</b>
		<b>Artificial Intelligence Technologies in Daily Assistance to the Public Administration</b> <b>Moderator: Andrejs Vasiljevs, ELRC</b>
		<b>Virtual Assistants for Support Services</b>
15:05 16:55	-	<i>Helmutš Ancāns, Bank of Latvia</i> <i>Kristīne Ilgaža, Rural Support Service</i> <i>Mētra Aržanovska, State Chancellery</i>
		<b>Application of Speech Technology</b>  <i>Ieva Kalderauska, National Electronic Mass Media Council</i> <i>Indulis Bikše, Parliament of the Republic of Latvia</i>

**ELRC Workshop Report for Latvia**

---

**Machine Translation Solutions**

*Edgars Bondars, Ambassador Extraordinary and Plenipotentiary*

*Ģirts Šatats, Latvian Standard*

*Vents Armands Krauklis, Valka Municipality*

---

16:55 -  
17:10

**Summary and Conclusions**

---

## 3 Summary of Content of Sessions

### 3.1 Welcome and introduction

The Latvian ELRC Workshop was opened by **Andrejs Vasiljevs** who represented the ELRC consortium and the local organizer Tilde. This presentation set the scene for the ELRC activities in the broader context of AI trends and EU policy in promoting language equality in the digital age.

Overall aims and objectives of the third ELRC workshop in Latvia and its agenda were explained.

### 3.2 European Language Equality - Latvian Language in the Digital Age

In her talk prof. **Ingūna Skadiņa** (University of Latvia) presented the key findings of the European Language Equality (ELE) project. She started with an overview of the META-NET Study “European Language in the Digital Age” published 10 years ago. This seminal study concluded that at least 21 of European languages are at risk of digital extinction and Latvian language is among this group at risk. Latvian language was included in the 4<sup>th</sup> – lowest – category that listed languages with weak or no support in the key language technology resources and services. This study helped to drive political support that culminated in the European Parliament resolution on Language equality in the digital age adopted on September 11, 2018.

A number of projects and initiatives were triggered to boost technological support for Latvian. The results of these activities are summarized in the recent study of the ELE project. Prof. Skadina highlighted some of the most notable results. For language resources she noted online thesaurus tezaurs.lv, Latvian WordNet, national terminology portal Termini.gov.lv and EuroTermBank.eu, national language corpora Korpuss.lv, Latgalian spoken language corpus<sup>1</sup>, data provided by the platform Hugo.lv, parliamentary data corpus ParlaMint. As notable examples of tools she mentioned Latvian NLP Pipeline as a Service NLP-PIPE, WMT award winning machine translation systems by Tilde, EU Presidency Translator solution, speech transcription services RUTA:MED and Tilde Transcribe, and the broad use of virtual assistants and chatbots. She emphasized the importance of LT infrastructures, namely, HUGO.LV, CLARIN.LV and NLTP (National Language Technology Platform).

She concluded that major progress has been achieved in the last decade. However this field develops very rapidly and more activities and investments are needed to catch up. It is critical for the long-term sustainability of the Latvian language to provide a focused long term support for its technological development.

### 3.3 Future Directions: the Common European Language Data Space

The session Common European Language Data Space was presented by **Philippe Gelin** (Head of Unit G3: Accessibility, Multilingualism and Safer Internet, DG CONNECT, EC). In his online talk he presented the latest developments in the area of language data planned in the framework of the Digital Europe Programme. The creation of EU-wide common interoperable data spaces in strategic sectors aims to overcome existing legal and technical barriers to data sharing and, as such, unleash the potential of data-driven innovation. Emphasis was put on the Language Data Space and its future role in European LT.

Also, the Centre of Excellence in LT (CELT) and Centre of Excellence in LT Plus (CELT+) were presented, where the CELT will be formed by Member States ministries representatives and CELT+ will comprise research and industry representatives.

---

<sup>1</sup> <https://mularkorpuss.rta.lv/#/>

## ELRC Workshop Report for Latvia

The new legal framework to support Multi Country Projects - European Digital Infrastructure Consortium (EDIC) – was presented. It has numerous advantages in supporting national and European activities in language data and technologies such as facilitating coordination of funding between member states, speedy establishment and flexible implementation, ensuring common standards and interoperability, and ensuring sustainability of activities.

### 3.4 eTranslation: Machine Translation under the Digital Europe Program

CEF eTranslation platform/CEF LT services were presented by **Christopher Scott** (DG Translation, European Commission). In his talk he presented the current state of the eTranslation system and CEF LT services. The original eTranslation system, that was developed for the 24 official languages of the EU plus Icelandic and Norwegian, was expanded to include other important languages for Europe such as Japanese, Arabic, Russian and Ukrainian. He explained the intended role of the service and presented various usage scenarios. Furthermore, he explained how the target user base for these services has been expanded from public administrations to European SMEs, universities and NGOs, in order to assist in the development of a data-based economy by easing the online access with instant machine translation of contents that could never be translated by humans. The role of language data and the value of data collected by ELRC initiative was also presented.

### 3.5 HUGO.LV – Language Technology Platform for Latvian State Administration

The Language Technology Platform for the Latvian State Administration Hugo.lv was presented by **Jānis Ziediņš** (Culture Information Systems Centre). It is freely accessible to every resident of Latvia and provides automatic translation, speech recognition and speech synthesis, as well as a range of tools for supporting multilingual features in e-services.

Hugo.lv is customized to the Latvian language and state administration documents, thus, its translation quality is much higher than in other online translation services. More than 400 million words are translated yearly on the platform services. Furthermore, users of Hugo.lv can enjoy the services of the translation assistant for more convenient translation. Hugo.lv can be integrated into all state administration platforms and homepages to facilitate multi-lingual access to e-administration data and e-services.

Hugo.lv is also unique in providing a powerful platform for creating and running advanced virtual assistants i.e. chatbots. More than 60 public institutions have integrated Hugo.lv powered chatbots on their websites.

### 3.6 Virtual Assistants for Support Services

This session was dedicated to virtual assistants and their use in the public administration of Latvia.

**Helmūts Ancāns**, Head of Strategic Development at the Bank of Latvia, presented a use case of a virtual assistant in the central bank of Latvia. Chatbot Mona is a new channel for communication with society for several departments of the Bank – Statistical Department, Communication Department, IT Department, Cash Department. The knowledge base of the chatbot that is powered by Hugo.lv includes 280 answers and 4500 information requests.

**Kristīne Ilgaža** from the Rural Support Service showcased the positive impact that the introduction of the virtual assistant has on the client support operations. The client service operators have acquired new skills and now can work as trainers for their virtual assistant. This is a great motivation for employees to get new skills and avoid redundant work answering the same questions again and again. Now they see how chatbots can successfully deal with numerous basic questions on a 24/7 basis. Since

## ELRC Workshop Report for Latvia

2020 when the chatbot was first introduced it has become much more knowledgeable thanks to advanced conversation scenarios created by the trainers.

**Mētra Aržanovska**, the State Chancellery of the Republic of Latvia showcased virtual assistant Zintis for public administration. The uniqueness of this chatbot is its ability to serve multiple institutions by adapting to the specific information pertinent to each institution.

This chatbot is also created on the Hugo.lv platform. Its knowledge base is comprised of the specific information of every institution as well as information common for several institutions and state administration. The chatbot is integrated with Government information systems and automatically acquires information from the open data portal.

### 3.7 Application of Speech Technology

In this session the application of speech technologies in public sector institutions was discussed.

**Ieva Kalderauska**, Member of the National Electronic Mass Media Council (NEMMC), talked about the importance to work on accessibility of mass media. In particular she highlighted the need to greatly increase subtitling in public media. Surveys show that 43% of the Latvian population use subtitles when consuming audiovisual media. 61% read subtitles with translation of foreign language media, 16% use subtitles to watch content without sound, for 2% subtitles are important due to hearing disabilities. Currently only 11% of the content is provided with subtitles. This is why NEMMC is promoting the work on automated subtitling tools that are essential to tackle the challenge.

**Indulis Bikše** from the Parliament of the Republic of Latvia presented the use case of automatic speech recognition (ASR) tool. A specialized ASR system is created for the transcription service of the Parliament. It speeds up creation and reduces workload in preparing transcripts of Parliament sessions. The word error rate for current system is less than 7% and it decreases the time required for preparation of publication ready transcripts by up to 40%.

### 3.8 Machine Translation Solutions

In this session several use cases of machine translation application were presented.

**Edgars Bondars**, Ambassador Extraordinary and Plenipotentiary, talked about organization of The Three Seas Summit that greatly benefited from machine translation services. The Three Seas Initiative, known also as the Baltic, Adriatic, Black Sea Initiative or simply as the Three Seas, is a forum of twelve European Union states running along a north–south axis from the Baltic Sea to the Adriatic and Black Seas in Central and Eastern Europe. The Initiative aims to create a regional dialogue on questions affecting the Member States such as economic growth and resilience, attracting investments in the region, energy security, inclusive connectivity and digital transformation. Specialized machine translation services were developed to support document translation and multilingual communication for the Summit participants. Machine translation also enables access to the Summit website in all EU languages.

**Ģirts Šatats** from Latvian Standard (LVS) talked about the challenge of the Latvian national standardization body in adopting and promoting international standards. Specialized machine translation systems are used by LVS employees in the translation process. Machine translation systems were integrated in the SDL Trados computer aided tool to be used in the post-editing mode. Extensive use of custom machine translation systems and terminology resources have ensured high



## ELRC Workshop Report for Latvia

accuracy and consistency of translations and has greatly improved efficiency of the standard adaptation.

**Vents Armands Krauklis**, Mayor of Valka Municipality, talked about the specific situation of the city divided between two countries – Latvia and Estonia. Integration of machine translation on the website of Valka provides immediate access to all information published by municipality for Estonian speakers.

### 3.9 Take-home message and conclusions

In the concluding session of the ELRC workshop **Andrejs Vasiljevs** summarized the topics covered at the workshop. He praised impressive showcases presented by the workshop participants. He thanked everybody who has contributed to the success of the ELRC action. He concluded by highlighting the importance of participating in the next stage of building the Pan-European language data and language technology infrastructure – the Language Data Space and Language EDIC.

## 4 Synthesis of Workshop Discussions

The success of the ELRC Workshop was twofold – (1) in the promotion of current achievements and motivation to broaden the use of existing LTs, and (2) in motivation to cooperate and contribute to the creation of sustainable national and Pan European language data and technology infrastructures.

The workshop clearly demonstrated some major achievements in adopting language technology in the public sector institutions of Latvia. National language technology platform Hugo.lv plays a seminal role in making LT services and tools easily and freely usable and adaptable for the needs of different institutions. It has become an indispensable tool in everyday work of public sector institutions as well as SMEs and individuals. Integration of eTranslation services in Hugo.lv further increases its scope of language support.

A notable trend is the rapid increase in the adoption of virtual assistants in G2C (Government to Citizens) services. The extension of Hugo.lv with a platform for creating and running of virtual assistants / chatbots has enabled numerous institutions to adopt this technology in a very short timeframe. The workshop demonstrated convincing benefits of “employing” chatbots and served as a strong encouragement to further proliferate the use of this technology.

The workshop also showed a growing interest in speech technologies. It is driven both by the growing consumption of audiovisual media and by the necessity to ensure accessibility and inclusiveness of information and services.

Representation at the workshop of the key public institutions laid a strong foundation for future cooperation on national and European level in Language Data Space and EDIC initiatives.

## 5 Country Profile: Language data creation, management and sharing

An updated profile for Croatia with respect to language data creation and management has been recently published in the 2022 ELRC White Paper (see: <https://lr-coordination.eu/sites/default/files/LRB/LRB-12/ELRC-White-Paper.pdf> ).

## ELRC Workshop Report for Latvia

### 6 Workshop Participants

The workshop was mostly targeted at the public sector participants to provide a platform for sharing experiences in the use of LTs in the public context and drive support for the forthcoming initiatives of Language Data Space and EDIC. 84 invitations were sent and follow up phone calls were made to the up-to-date list of the key public sector stakeholders related to the language technology area. There were 32 onsite participants at the workshop. To further extend the reach of the seminar a Facebook online streaming was provided although we do not have statistics how many viewers it attracted.

The workshop succeeded to bring together decision makers and representatives from the key public sector bodies – Parliament of Latvia (Saeima), State Chancellery, Bank of Latvia, Ministry of Education and Science, Ministry of Foreign Affairs, National Electronic Mass Media Council, the State Language Agency, State Language Centre, and others.

Although only 16% of participants were from academic institutions they represented the key players in the field – the Latvian Science Council, University of Latvia and Institute of Mathematics and Informatics, Riga Technical University.