# Deliverable D3.2.13
# Task 8

# ELRC Workshop Report for Estonia

| | |
|---|---|
| Dissemination Level: | Public |
| Version No.: | V1.0 |
| Date: | 2018-11-30 |

# Contents

# 1 Executive Summary

The Estonia ELRC Workshop took place in Tallinn, on the 3$^{rd}$ of October 2018, at the office of the European Commission Representation in Estonia.

The workshop was co-organized by the ELRC Consortium member Tilde, Estonian Language Resource Centre and Office of the European Commission Representation in Estonia. In general, the participants appreciated the workshop very much, finding a well-organised arrangement, a pleasant atmosphere and interesting presentations.

An event-dedicated web page in Estonian was set up at the ELRC website before the event, at http://www.lr-coordination.eu/l2estonia. The page was populated with the agenda (http://www.lr-coordination.eu/l2estonia_agenda) and the online registration form.

Invitations to the participants were sent out a couple of weeks before the event. The invitation text was a localized version of the respective template. The invitation was followed by a personal e-mail and calls to the persons to whom invitations were sent.

The event attracted 49 participants. All feedback forms that were handed in showed a positive experience.

## 2   Workshop Agenda

The workshop agenda was based on a "master" agenda, the agenda was adopted for the local needs.

http://www.lr-coordination.eu/l2estonia_agenda

| | |
|---|---|
| 09:00 | **Registration and coffee** |
| 09:30 – 9:40 | **Welcome by the local organisers** <br> *Pekka Myllylä, (Tilde EE)* |
| 09:40 – 10:10 | **Translation at DGT** <br> *Merit-Ene Ilja, (DG Translation)* <br> **Presentation** |
| 10:10 – 10:35 | **Connecting public services across Europe: ambition and results so far.** <br> Video presentation by representative from *DGCONNECT* |
| 10:35 – 11:00 | **Overview of the European Language Resource Coordination action** <br> *Andrejs Vasiļjevs (ELRC, Tilde)* |
| 11:00 – 11:30 | *Coffee Break* |
| 11:30 – 11:50 | **Multinguality Challenges in Estonia: Riigi Teataja case study** <br> *Jüri Heinla (Ministry of Justice)* <br> **Presentation** |
| 11:50 – 12:30 | **National initiatives for language technologies in Estonia** <br> *Kadri Vider, Kadri Vare, Keaty Siivelt* <br> *(Estonian Language Resources Center, Ministry of Education)* <br> **Presentation I  Presentation II** |
| 12:30 – 12:50 | **Legal aspects of sharing language resources** <br> *Aleksei Kelli (University of Tartu)* <br> **Presentation** |
| 12:50 – 13:50 | *Lunch* |
| 13:50 – 14:15 | **Translation Procurements in Estonian Public Sector** <br> *Inge Rätsep (ETBL, Interlex)* <br> **Presentation** |

| 14:15 – 14:40 | **The CEF eTranslation platform @ work**<br>*Ingrid Sibul (EC), Kelly Lilles (EKI)*<br>**Presentation I  Presentation II** |
| --- | --- |
| 14:40 – 14:50 | **Wrap-up** |

## 3  Summary of Content of Sessions

### 3.1  Translation at DGT

**Merit-Ene Ilja**, the Director of the DGT, opened the event by welcoming the audience.



**Image 1 Merit-Ene Ilja presenting Translation at DGT**

Ilja introduced to the audience the European Commission's translation activities, organization of DGT, translation volumes, translation outsourcing and translation tools used.

The annual cost of DG Translation is around €325 million and the latest figure for DG Interpretation is €119,5 million. In the EU total budget, translation costs are part of the administrative costs. They have been ±6% of the total budget for a long time. Translation in all EU institutions costs €1.7 per citizen per year

She also described the role of the eTranslation and outlined objectives and challenges for future, notably: continuos development of information technology (eTranslation, CAT and other tools), streamlining and, where feasible, automation of workflows, and optimisation of quality assessment, optimisation of translation management.

She discussed how DGT follows the evolution of AI, including NMT, and how the world is changing – translation is ubiquitous, hidden component of services and products and not a luxury good any more.

### 3.2  Overview of the European Language Resource Coordination action

**Andrejs Vasiļjevs**, the ELRC representative, welcomed the audience on behalf of the consortium.

**Image 2 Andrejs Vasiļjevs presenting Overview of the European Language Resource Coordination action**

He described the crucial role that language technology development has for smaller languages for them to survive in the digital era and the importance of cooperation in collecting language resources for building automated translation systems. He also gave an overview of the ELRC consortium's activities and concrete activities and services offered to data providers and donors.

## 3.3    Multilinguality Challenges in Estonia: Riigi Teataja case study

 **Jüri Heinla** from the Ministry of Justice introduced Estonia's information system for legal acts – Riigi Teataja (RT) where visitor can look up consolidated texts of English translations of Estonian legislation.



**Image 3 Jüri Heinla presenting Multilinguality Challenges in Estonia: Riigi Teataja case study**

Upon launching the English website of Riigi Teataja, the translations of nearly 260 Estonian laws (acts of Parliament) became available. Translations are kept up-to-date so that current Estonian legislation is always available in English. This means that whenever an act is amended, a translation of the new consolidated text containing the recent amendments are published usually no later than by the date of entry into force of the amendment.

Heinla also described the users of the legislation and their multilingual preferences and how these are met by translating all of the legislation into English. Each year, it costs EUR 63 000 to translate laws.

Translation of legal acts into English by numbers:

- A total of 3000 different translation versions have been published;

- A total of 566 translations are currently valid;

- 300 000 users per year representing 263 countries
  https://www.riigiteataja.ee/en/statistics/bycountry;

**Table 1 Top 10 countries of visitors**

| # | Country | Visitors |
|---|---|---|
| 1 | Estonia | 683063 |
| 2 | Latvia | 37585 |
| 3 | India | 33795 |
| 4 | United Kingdom | 30854 |
| 5 | United States | 29704 |
| 6 | Finland | 28326 |
| 7 | Germany | 25919 |
| 8 | Lithuania | 22111 |
| 9 | Russia | 22075 |
| 10 | Ukraine | 17364 |

One of the unique services provided by RT is "Time machine" – a feature to serve up-to-date acts at any given point of time, including past acts and acts that will take force in future.
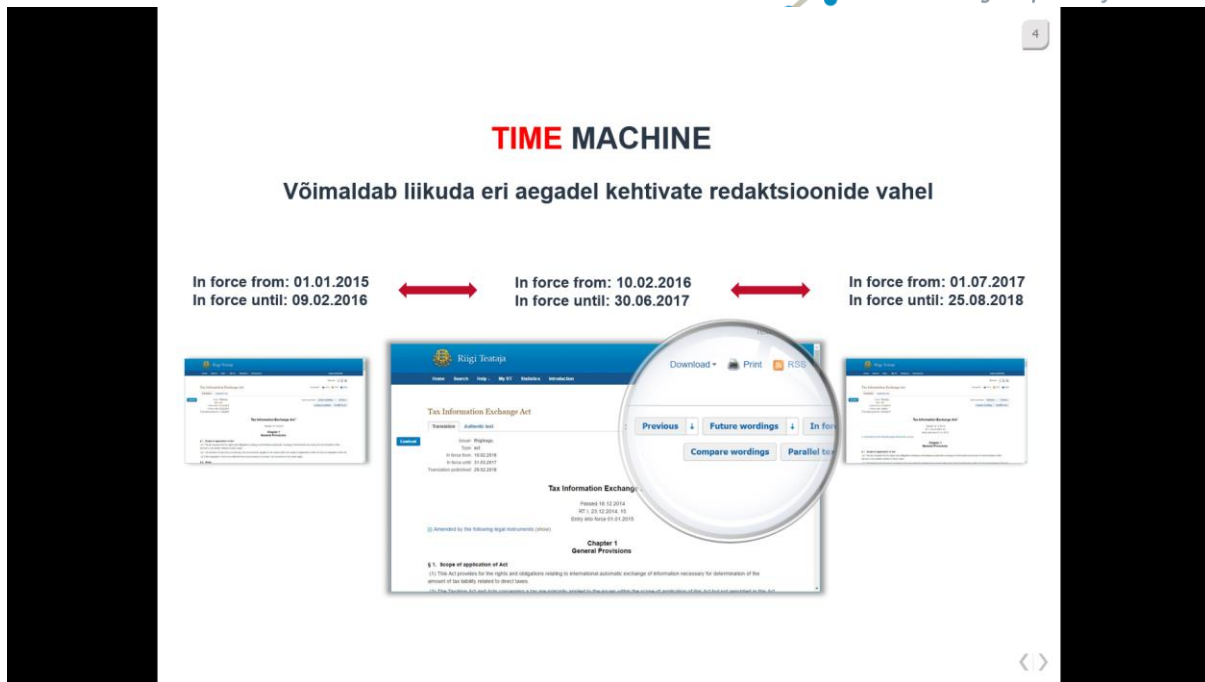
**Image 4 Illustration on the Time Machine principle of the Riigi Teataja**

The goals of translating Estonian legal acts are to introduce the Estonian legal framework to Estonian partners in Europe and worldwide, and to make Estonian legal acts accessible for foreign investors.
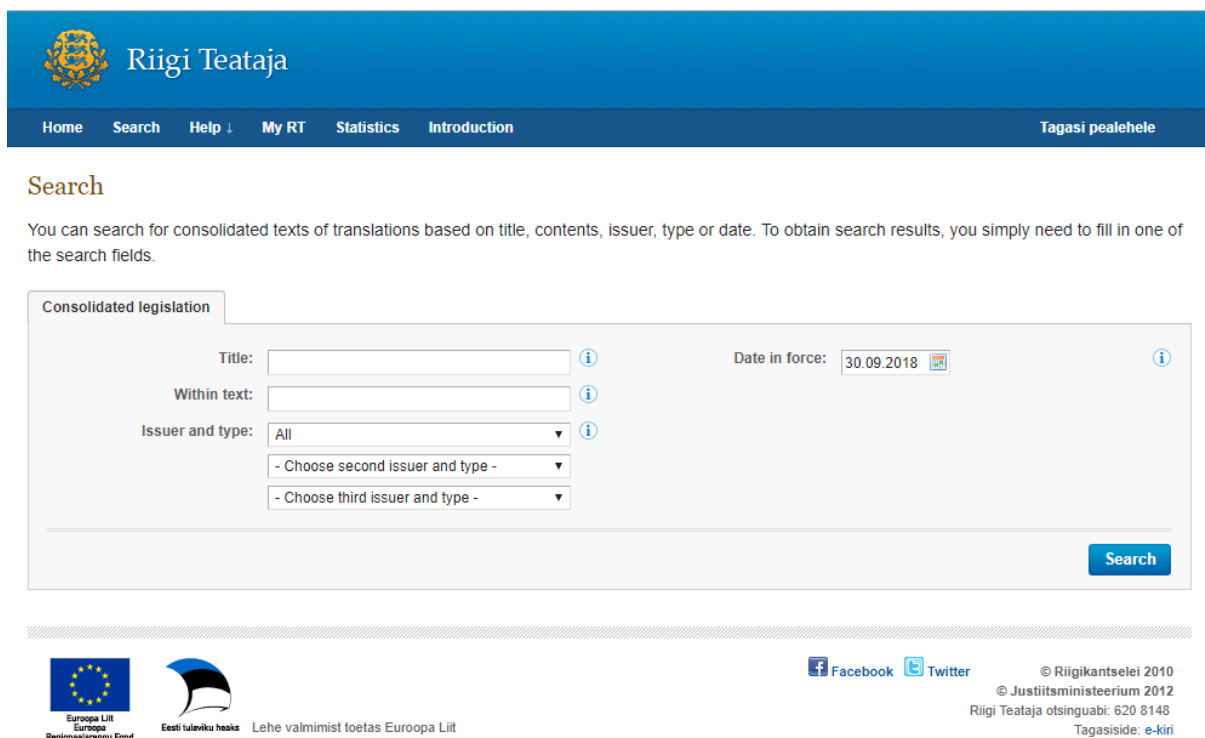


**Image 5 Riigi Teataja search feature**

Heinla also described translation policies and supporting tools (XML Vexpro), and the possibility to use custom machine translation with post-editing carried out by legal experts. But before applying MT the process of translating legal acts must be re-engineered: terminology work, translation methodology, and inclusion of editors.
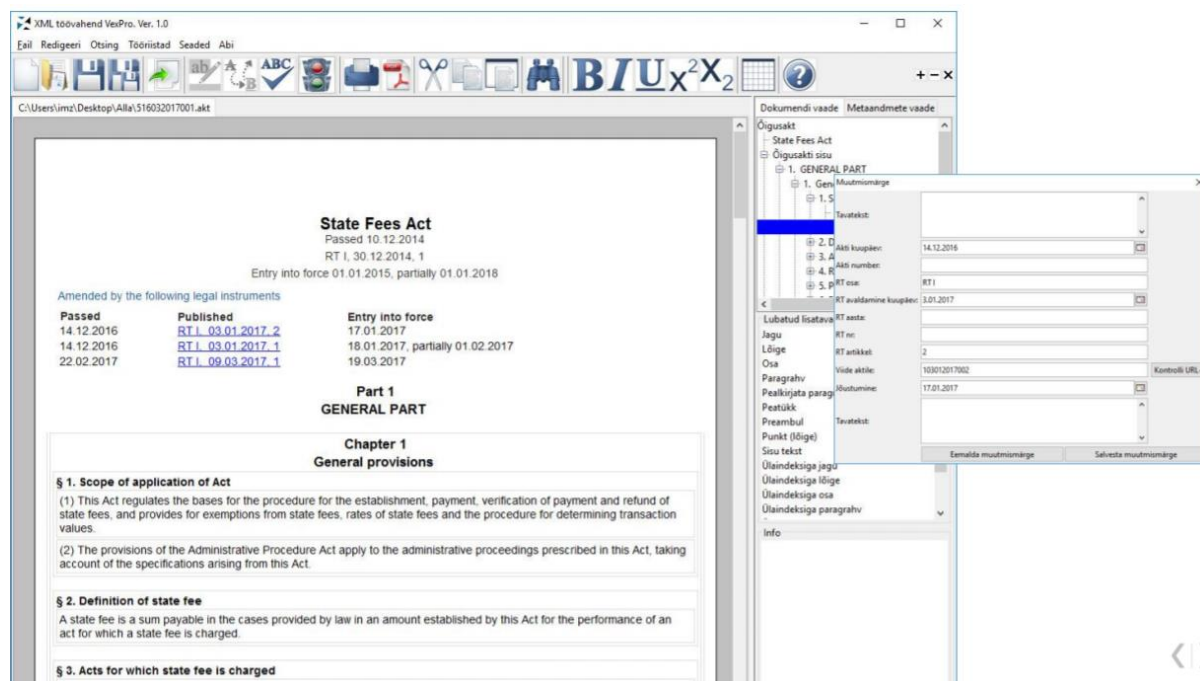


**Image 6 XML Vexpro**

## 3.4 National initiatives for language technologies in Estonia – Panel session

**Kadri Vare** from the Ministry of Education introduced the Estonian Language Technology Programme (cf https://www.keeletehnoloogia.ee/en) and it's outcomes, namely: speech recognition, machine translation, speech synthesizes, text analytics.

In the field of machine translation the Estonian LT Programme has funded several projects, including Estonian Open Parallel Corpora (Tilde) for collecting Estonian-English/Latvian/Finnish/ parallel corpora, general domain machine translation systems by University of Tartu and Tilde, and speech technology projects by Tallinn Technical University.

**Image 7 Introduction of the Estonian Speech Recognition Project**

The results of the projects financed under this programme are freely available by default to everyone.



**Image 8 Speech recognition results http://bark.phon.ioc.ee/**

**Kadri Vider**, the head of the Center of Estonian Language Resources, welcomed the audience on behalf of her organization, one of the co-organizers of the event. She explained to the audience what a language resource is, presented the Estonian language resource repository based on META-SHARE, talked about the metadata for LRs, term bases, and Estonian wordnet.

The current status of Estonian resources at ELRC at the time of reporting is 10 language resources. An up to date status could be followed at the https://www.elrc-share.eu/ site.
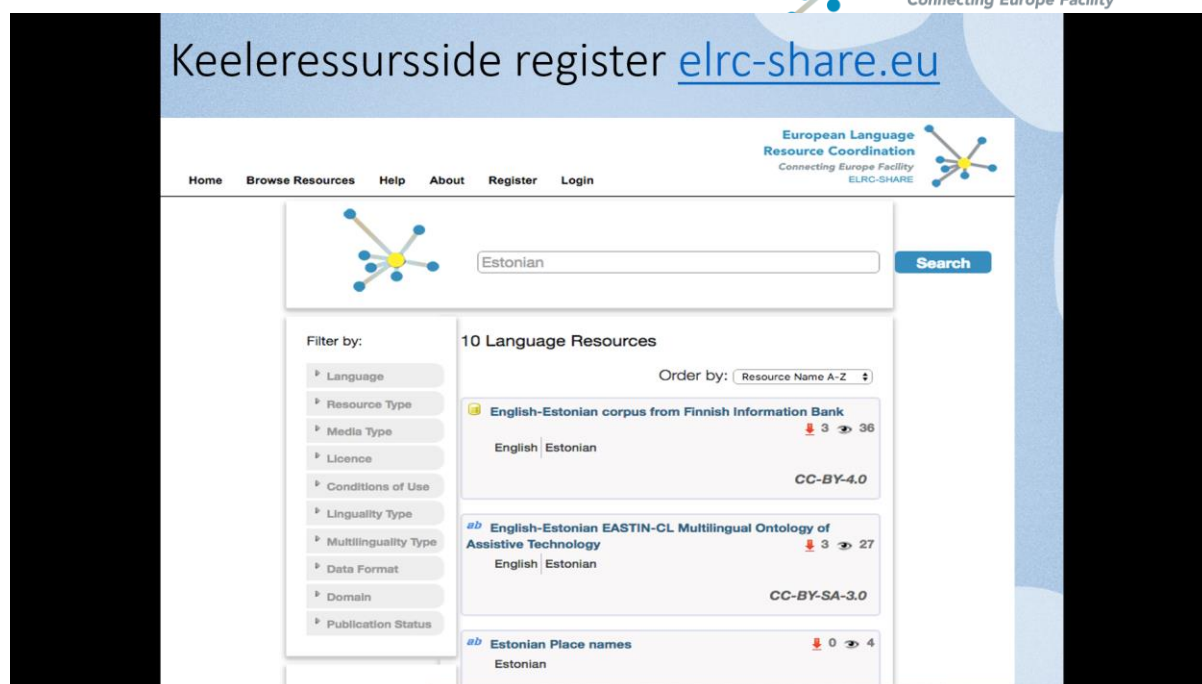
**Image 9 https://www.elrc-share.eu/ site showing some of the Estonian resources**

## 3.5 Legal aspects of sharing language resources

**Aleksei Kelli** from the University of Tartu discussed the legal frameworks and their implications on language resource creation, re-use and modifications. He also discussed personal data protection including GDPR and its implications on language resource creation and sharing.



**Image 10 Aleksei Kelli presenting Legal aspects of sharing language resources**

The grounding principle being *Nemo plus iuris transfere (ad alium) potest quam ipse habet (Ulpianus)* -- a Latin phrase meaning 'no one can transfer a greater right than he himself has'. He described the exceptions for re-using LRs, also private data in LRs.

## 3.6   Translation Procurements in Estonian Public Sector

**Inge Rätsep** discussed the as-is practices of translation procurements in public sector and gave suggestions how to improve it. She emphasized the role of TMs re-use, technical tools that support it, cat-grid.

She suggested to use the following questionnaire while preparing a translation tender document

- The goal of translating?
- Language pairs, source and target languages?
- Volumes?
- Who will be managing translation memories?
- Who is the owner of translation memories?
- How long is the contractual period?

She also discussed the benefits of language technologies in translation procurements.

For future public sector translation procurements, she suggested to:

1. Consolidate related (per domain) procurements;
2. Create domain-specific TMs and terminology repositories and manages these resources together;
3. In addition to considering word price consider the total cost of the whole contracting period.

## 3.7   The CEF eTranslation platform @ work

**Ingrid Sibul** outlined the features of eTranslate, its user types and usage scenarios, evolution from MT@EC, security aspects of the system. She compared the statistical and NN based approached to MT. The presentation was concluded with a roadmap ahead.

Sibul described how eTranslation is used in N-Lex portal and the possibilities to carry out searches over serveral languages.
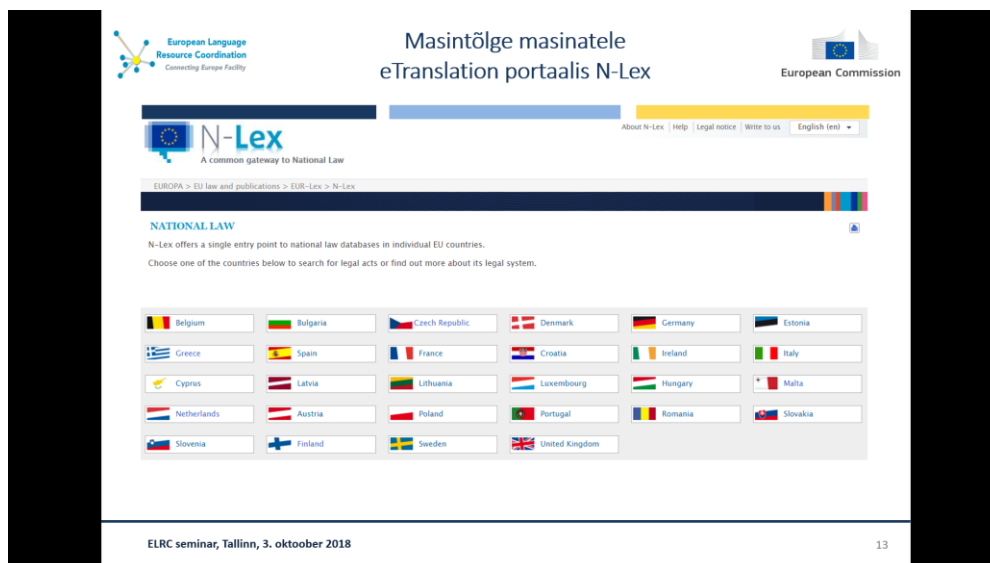


**Image 11 eTranslation for N-Lex portal**

**Kelly Lilles** from the Institute of Estonian Language introduced eTranslation Termbank project. The project is financed by CEF and Estonian Ministry of Education.

The goals of the project in Estonia is to collect and provide for eTranslation Estonian terminology resources in various domains with a focus on health, business legislation and consumer protection.
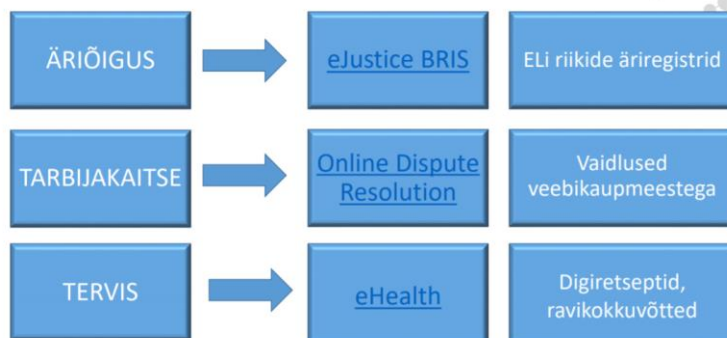


**Image 12 Termbases for the project focus domains**

So far the Institute of Estonian Language has collected over 53 termbases containing more than 18 000 entries, the majority of these are available with CC-BY license type.



**Image 13 license representation for 53 term collections**

Project dedicated web part in Estonian language has been set up under Institute of the Estonian Language main Web page and can be accessed through following link: http://portaal.eki.ee/tegevusvaldkonnad/2192-etranslation.html. This page describes consortium and major objectives of the project.

# 4  Synthesis of Workshop Discussions

Highlight:

- The automated translation for English-Estonian has gone through a tremendous quality leap due to Neural Machine Translation.
- The public sector translation procurements have evolved compared to the previous ELRC workshop, the benefits of re-using TMs have been recognized by the state authorities.
- National programme on developing Estonian language technologies among other activities supports creation of language resources that may be of use for eTranslation.

# 5   Workshop Participants

The Estonian ELRC workshop received 73 registrations spanning a wide range of ministries and public organizations, and from LSPs and academia. Before the workshop the cold calling activities were organized in this way organizers were able to make sure that important public administration institutions were represented.

The workshop was well-attended – there were 49 participants. Table 1 illustrate distribution of participants per sectors.

**Table 2 Distribution of participants per category**

| Sector | Number of participants | % |
|---|---:|---:|
| Public service administration | 21 | 43% |
| Participants from LSPs | 7 | 14% |
| Participants from technology eco-system | 17 | 35% |
| Organizers | 3 | 6% |
| EC | 1 | 2% |
| TOTAL | **49** | |

The full participants list can be viewed in the Appendix.



**Image 14 Video welcoming**

**Image 15 Workshop participants**

# 6 Appendix

## 6.1 Localised invitation letter

**Lugupeetud kaasamõtleja**

Kutsume Teid osalema Euroopa keeleressursside koordineerimise (ELRC) Eesti seminaril, mis toimub 3. oktoobril Tallinnas Euroopa Liidu Maja infokeskuses, Rävala 4 (sissepääs Laikmaa tänavalt)..

ELRC on Euroopa Komisjoni Euroopa ühendamise rahastu (CEF) algatus. CEF pakub korduskasutatavaid ja digitaalseid alusteenuseid, mis on Euroopa ühendamiseks hädavajalikud. Kogu Euroopa avalike teenuste ühendatuse ja koostalitlusvõime kindlustamine võib käegakatsutavalt mõjutada Euroopa avalikke teenuseid, ettevõtteid ja kodanike igapäevaelu.

CEFi teenuste hulka kuuluv eTranslation-süsteem aitab ületada keelebarjääre, mis on tänapäeval nii Euroopa kui ka kohaliku tasandi avalikus halduses ühed peamised teabevahetust takistavad murekohad. Seetõttu on ELRC eesmärk tuvastada ja koguda riikliku avaliku halduse ja riiklike teenuste seisukohast olulisi keeleressursse, mille abil saaks eTranslation-süsteemi kohandada riiklike haldusasutuste vajadustega ning muuta avalikud digitaalteenused mitmekeelseks.

ELRC seminaril soovime koos Euroopa Komisjoni valdkonnaspetsialistidega, Eesti avaliku halduse ja valitsuse kõrgetasemeliste esindajatega, keeletehnoloogidega ning keeleteenuste osutajatega pidada lahendustele orienteeritud arutelu, mille käigus saaksime üksteisega jagada tänapäevase mitmekeelse avaliku halduse kogemusi ja vajadusi.

Samuti loodame ühiselt välja selgitada asjakohased mitmekeelsed keeleressursid, mis hõlbustaksid CEFi eTranslation-süsteemi kohandamist Eesti riigi avalike teenustega, ning käsitleda automaattõlke jaoks andmete kasutamisega seostuvaid võimalikke tehnilisi ja õiguslikke murekohti.

Üksikasjaliku päevakorra leiate http://www.lr-coordination.eu/et/l2estonia_agenda.

Teie tagasiside ja osavõtt aitavad CEFi eTranslation-süsteemi Eesti avalikes teenustes sujuvamalt ja edukamalt kasutusele võtta. ELRC seminaril osalemine on tasuta, kuid üritusel osalemiseks on vaja ennast registreerida veebilehel http://www.lr-coordination.eu/l2estonia.

Vastame meelsasti teie küsimustele ning loodame teist peagi kuulda.

Lugupidamisega,

Pekka Myllylä

Tilde Eesti tegevjuht

pekka.myllyla@tilde.com

Tel. +372 533 37108

## 6.2   Localised agenda

| | |
|---|---|
| 09:00 | **Registreerimine ja hommikukohv** |
| 09:30 – 9:40 | **Sissejuhatus päeva**<br><br>*Pekka Myllylä, (Tilde EE)* |
| 09:40 – 10:10 | **Tõlkimine DGT-s**<br><br>*Merit-Ene Ilja, (DG Translation)*<br><br>**Esitlus** |
| 10:10 – 10:35 | **Avalike teenuste piiriülesus, eesmärgid ja senised tulemused (video on inglise keeles)**<br><br>*DGCONNECT* |
| 10:35 – 11:00 | **Ülevaade ELRC tegevustest (ettekanne on inglise keeles)**<br><br>*Andrejs Vasiļjevs (ELRC, Tilde)* |
| 11:00 – 11:30 | ***Coffe Break*** |
| 11:30 – 11:50 | **Mitmekeelsusega seotud väljakutsed Riigi Teataja näitel**<br><br>*Jüri Heinla (Ministry of Justice)*<br><br>**Esitlus** |
| 11:50 – 12:30 | **Eesti Keeletehnoloogia Programm ja selle rakendatavad tulemused**<br><br>*Kadri Vider, Kadri Vare, Keaty Siivelt (Estonian Language Resources Center, Minitry of Education)*<br><br>**Esitlus I  Esitlus II** |
| 12:30 – 12:50 | **Keeleressursside korduvkasutuse õiguslikud aspektid**<br><br>*Aleksei Kelli (University of Tartu)*<br><br>**Esitlus** |

| | |
|---|---|
| 12:50 – 13:50 | *Lunch* |
| 13:50 – 14:15 | **Tõlkehanked avalikus sektoris**<br><br>*Inge Rätsep (ETBL, Interlex)*<br><br>**Esitlus** |
| 14:15 – 14:40 | **eTranslation automaattõlkeplatvormi rakendamine**<br><br>*Ingrid Sibul (EC), Kelly Lilles (EKI)*<br><br>**Esitlus I  Esitlus II** |
| 14:40 – 14:50 | **Kokkuvõte** |

## 6.3   Localised press release

**Euroopa ühendamine üle keelebarjääride:**
**Eesti positsioon Euroopa digitaalteenuste maastikul**

**Teine Euroopa keeleressursside koordineerimise Eesti seminar**

Euroopa Liidu Maja, Rävala pst. 4.

3. oktoober 2018

Korraldajad:

Euroopa Komisjoni kirjaliku tõlke peadirektoraat

Euroopa keeleressursside koordineerimise konsortsium

Tilde Eesti

2014. aastal asutas Euroopa Komisjon Euroopa Ühendamise Rahastu (CEF), mille eesmärk on toetada üleeuroopalise transpordi-, telekommunikatsiooni- ja energeetikavõrgustiku ning -taristu arendamist, mis võimaldaks välja kujundada 28 Euroopa Liidu liikmesriigi, Islandi ja Norra digitaalse ühisturu. CEFi üks peamisi sihte on töötada välja ja pidada käigus mitmesuguseid digitaalteenuste taristuid (DSI), mis muudaksid sujuvamaks piiriülese suhtluse riigiasutuste, ettevõtete ja tavakodanike vahel. Need digitaalteenuste taristud on korduskasutatavad digitaalsed alusteenused, mis loovad vajalikud tingimused Euroopa ühendamiseks muuhulgas tarbijaõiguste, tervishoiu, avalike hangete, sotsiaalkindlustuse ja kultuuri vallas.

CEFi teenuste hulka kuuluv eTranslation-süsteem aitab ületada keelebarjääre, mis on tänapäeval nii Euroopa kui ka kohaliku tasandi avalikus halduses ühed peamised teabevahetust takistavad murekohad. Seetõttu on Euroopa keeleressursside koordineerimise (ELRC) eesmärk tuvastada ja koguda riikliku avaliku halduse ja riiklike teenuste seisukohast olulisi keeleressursse, mille abil saaks eTranslation-süsteemi kohandada riiklike haldusasutuste vajadustega ning muuta avalikud digitaalteenused mitmekeelseks.

Eesti on Euroopa keeleressursside koordineerimise tegevuses osalenud alates 2015. aastast ning toetab algatuse nägemust ja sihte täielikult. Kuid Eesti ja eesti keele positsiooni tugevdamiseks digitaalsel ühisturul nii saab kui ka tuleb veel palju ära teha.

Just seetõttu korraldabki ELRC nüüd – kaks aastat pärast esimest, edukat ja 30 eri riigis aset leidnud seminarivooru – Euroopa Komisjoni nimel Eestis järgmise seminari, mis toimub 3. oktoobril Tallinnas Euroopa Liidu Majas. Seminaril arutatakse Eesti positsiooni ja väljavaateid digitaalselt ühendatud mitmekeelses Euroopas. Lisaks käsitletakse e-riigiga, mitmekeelsusega ja avaandmetega seotud asjakohaseid riiklikke strateegia- ja tegevuskavasid, mille toel saab Eesti tõhustada Euroopa Komisjoni algatuste mõju. Soovime Euroopa Komisjoni valdkonnaspetsialistidega, Eesti avaliku halduse ja valitsuse kõrgetasemeliste esindajatega, keeletehnoloogidega ning keeleteenuste osutajatega pidada lahendustele orienteeritud arutelu, mille käigus jagame üksteisega tänapäevase mitmekeelse avaliku halduse kogemusi ja vajadusi. Samuti selgitame välja asjakohased mitmekeelsete andmete allikad, mis aitavad kohandada CEFi eTranslation-süsteemi Eesti riigi avalike teenustega, ning käsitleme automaattõlke jaoks andmete kasutamisega seostuvaid võimalikke tehnilisi ja õiguslikke murekohti.

Seminari sihtrühma kuuluvad avalike digitaalteenuste osutajad, riigiasutuste juhid, ametnikud ja partnerid, tõlkijad ning mitmekeelsete dokumentide ja andmete haldamisega tegelevad inimesed.

Seminar on osalejatele tasuta, kuid üritusest osavõtmiseks tuleb ennast registreerida veebilehel http://www.lr-coordination.eu/et/l2estonia_registration.

| Lisateave | |
|---|---|
| | |
| ELRC | www.lr-coordination.eu |
| ELRC Eesti seminar | www.lr-coordination.eu/et/l2estonia |
| Registreerimine | http://www.lr-coordination.eu/et/l2estonia_registration |