# National Language Technology Platform (NLTP): Overall View

**Artūrs Vasiļevskis**
Tilde,
Riga, Latvia
arturs.vasilevskis@tilde.com

**Jānis Ziediņš**
Culture Information Systems Centre,
Riga, Latvia
janis.ziedins@kis.gov.lv

**Marko Tadić**
University of Zagreb, Faculty of Humanities
and Social Sciences, Zagreb, Croatia
marko.tadic@ffzg.hr

**Željka Motika**
Central State Office for the Development of
Digital Society, Zagreb, Croatia
Zeljka.Motika@rdd.hr

**Mark Fishel**
University of Tartu,
Tartu, Estonia
fishel@ut.ee

**Hrafn Loftsson, Jón Guðnason**
Reykjavik University, School of Technology,
Reykjavik, Iceland
{hrafn, jg}@ru.is

**Claudia Borg**
University of Malta,
Valletta, Malta
claudia.borg@um.edu.mt

**Keith Cortis, Judie Attard**
Malta Information Technology Agency,
Blata l-Bajda, Malta
{keith.cortis, judie.attard}@gov.mt

**Donatienne Spiteri**
Office of the State Advocate,
Valletta, Malta
donatienne.spiteri@stateadvocate.mt

## Abstract

The work in progress on the CEF action National Language Technology Platform is presented. The action aims at combining the most advanced language technology tools and solutions in a new state-of-the-art, artificial-intelligence-driven, national platform for language technology oriented primarily towards users from public administrations of partner states.

## 1 Introduction

The paper presents the work in progress on the CEF action National Language Technology Platform (NLTP, INEA/CEF/ICT/A2020/ 2278398, duration 2021-04-01–2023-03-31). The general aim of the action is to combine the most advanced Language Technology (LT) tools and solutions in a new state-of-the-art, artificial-intelligence-driven, web-based national platform for LT. Currently, the action is approaching the implementation phase of the prototype systems. The system architecture plan has been created and will be followed by multiple implement-ations. In parallel, data collection and preparations for machine translation (MT) system training is gradually approaching its late stage. The details are described in section 2.

## 2 Development

### 2.1 Related work

The developed solution builds on the already existing `hugo.lv` platform and the results of the *EU Council Presidency Translator* (INEA/CEF/ ICT/A2018/1762093) action, but it will be substantially extended into NLTP in order to provide public administrations and the general public with secure access to high quality MT and

integration with computer aided translation (CAT) tools, e-mail and web plug-ins etc., for translation of texts and documents. This set of services is considered as initial, but the modular design of the platform allows it to be enriched with additional LT services beyond the initial set.

## 2.2 Users

In its final form NLTP will be adapted, localised, and sustainably deployed by the public administration bodies in partner states (Latvia, Croatia, Estonia, Iceland, and Malta), while its development is supported at the same time by local research institutions as complementary partners. In the case of Iceland and Estonia, the research partners were given the role of public authorities as well. Additionally, the NLTP will be customisable to the specific needs of public administrations and will be further linked to eTranslation[*] services, thus enabling translations into and from the 24 official EU languages and other languages of the Digital Single Market.

The NLTP will facilitate the use of a professional translation environment with integrated terminology databases, CAT and MT tools, all wrapped up in a simple-to-use HTML front end and coupled with a number of other technological solutions, such as a translation widget, browser plugin, commercial CAT tool plugins, etc.

## 2.3 Deployment

The NLTP will increase the efficiency of translation, the reuse of translation memories and make use of the existing high-quality MT technologies. Additionally, the action will also integrate speech technologies for selected languages with automatic speech recognition and/or text-to-speech services.

The platform will be developed according to this common overall concept, but for each partner state a deployable version will be adapted to the needs of public administrations at each level (local, regional, national). After the needs were modelled by overall general needs, the specific requirements have been collected through a survey about LT needs and expectations, that has been run in all partner states. The analysis of the survey in Croatia is presented in Motika et al. (2022), while the preliminary results for other languages will be available for the EAMT2022 conference poster.

## 2.4 Additional datasets

Additionally, a number of domain specific parallel data is being collected and will be made available through the ELRC-SHARE[†] repository in the Translation Memory eXchange (TMX) or similar compatible format. Since the sources of data are predominantly expected to come from the public domain, the data will be made accessible under permissive licences.

## 3 Sustainability and Future Directions

The public administration partner institutions will be responsible for the sustainability of each national NLTP after the action ends by securing its inclusion into the national infrastructures for eGovernment as cloud services. This will enable multilingual access to and by public administrations, while, at the same time, the integration with public digital services offered in languages of EU and EEA will be fostered.

For future research and development directions, similar platforms could be developed and deployed for other EU member states, and in this respect this action can be regarded as the proof-of-concept.

## Acknowledgements

## References

Motika, Željka; Didak Prekpalaj, Tanja; Horvat Klemen, Tamara; Košćec Perić, Mirjana. in press. Predstavljanje projekta "Nacionalna platforma za jezične tehnologije" In: *Proceedings of the μPro2022 Conference*, Opatija, May 2022.

---

[*] https://ec.europa.eu/digital-building-blocks/wikis/display/CEFDIGITAL/eTranslation

[†] https://elrc-share.eu