



EUROPEAN LANGUAGE DATA SPACE



LDS Value Demonstration Project for Culture Information Systems Centre

Andrejs Vasiljevs (Tilde, Latvia)
andrejs@tilde.com

Jānis Ziediņš (CISC, Latvia)
janis.ziedins@kis.gov.lv

19-03-2025 LDS Launch Conference
<https://language-data-space.ec.europa.eu>

LDS Value Demonstration Project for CISC

- A subproject within the scope of the LDS project to demonstrate value creation enabled through participation in the Language Data Space.
- CISC is ideally suited to serve for a proof-of-concept:
 - CISC is the key institution in Latvia managing cultural heritage information and digital access to it.
 - CISC operates Latvia's national language technology and AI platform HUGO.LV.
- The project will pilot full value chain from on-boarding in LDS, sharing and acquiring data on LDS, to fine-tuning Large Language Model (LLM) to fit the needs of Latvian cultural institutions and public administration.

Key Activities

- On-boarding CISC in LDS, setting up and configuring LDS Connector.
- Defining data assets and creating data offers by curating HUGO.LV data resources for distribution on LDS.
- Defining requirements for LLM fine-tuned to the needs of Latvia public administration.
- Identifying language resources relevant for LLM fine-tuning – data in HUGO.LV repository and data from other LDS participants.
- Fine-tuning selected foundational LLM on EuroHPC infrastructure.
- Assessing suitability of the resulting LLM for HUGO.LV services.
- Documenting and sharing key takeaways and best practice guidelines.



v1.0.0-beta

MANAGEMENT

Home

History

Storage solutions

OPERATIONS

Assets

Policies

Offers

ACCOUNT (Andrejs)

Logout



Culture Information
Systems Centre
Republic of Latvia

Learn

Andrejs

Connector Management Panel

2
Assets

Create asset Assets

18
Policies

Create policy Policies

2
Offers

Create offers Offers



v1.0.0-beta

MANAGEMENT

- Home
- History >
- Storage solutions >

OPERATIONS

Assets ^

View assets

Create asset

Policies >

Offers ^

Shared offers from others

View offers

Create offer

ACCOUNT (Andrejs)

Logout



Culture Information Systems Centre Republic of Latvia

Learn v

Andrejs v



List of Assets

Here you view or delete assets.

Search your assets...

Filters

Sort By
By name asc v

- FOLD Parallel Corpus on the Creative Industries**

corpus
Created At: 18 March 2025

Portal fold.lv contains human-translated articles on modern design, art, architecture. This dataset contains aligned parallel sentences in the domain of creativ...

[More details](#)
- Parallel corpus from portal fold.lv**

corpus
Created At: 18 March 2025

Portal fold.lv contains human-translated articles on modern design, art, architecture. This dataset contains aligned parallel sentences in that domain.

[More details](#)



v1.0.0-beta

MANAGEMENT

- Home
- History >
- Storage solutions >

OPERATIONS

Assets

View assets

Create asset

Policies >

Offers

Shared offers from others

View offers

Create offer

ACCOUNT (Andrejs)

Logout



Culture Information Systems Centre
Republic of Latvia

Learn

Andrejs

Creative Commons Attribution 4.0 International

View Policy

FOLD Parallel Corpus on the Creative Industries

corpus • version: v.2025.03

Publisher
name
Culture Information Systems Centre

Description
Portal fold.lv contains human-translated articles on modern design, art, architecture. This dataset contains aligned parallel sentences in the domain of creative industries.

language

language
Latvian

language
English

Categories

keyword
design art creative industries culture

modern

media type
text

linguality type
bilingual

Data protection properties

anonymized
no

distribution

byte size
3157028

packaging format
zip

Format
TMX

size
31005 translation unit



Culture Information
Systems Centre
Republic of Latvia



Ministry of Culture
Republic of Latvia

Culture information systems centre

The Centre for Culture Information Systems operates as a direct administrative entity under the Ministry of Culture's authority.

The mission of the Centre is to assist institutions—such as archives, libraries, and museums—in safeguarding and providing access to cultural heritage for future generations by utilizing cutting-edge information technology solutions.

CISC oversees the national language technology platform, [Hugo.lv](https://hugo.lv).

Public administration language tools

[Text translation](#) and [text simplifying](#), [speech recognition](#) and [speech synthesis](#), as well as [various tools](#) to support multilingual e-services.

View all tools →

Hugo.lv translates into 24 languages of the European Union!

Hugo.lv is connected to the European Commission's translation service eTranslation and offers automatic translation into all languages of the European Union member states. Translations for languages marked with "e" may take a little longer.

From: Latvian **English** Russian ▾ General ▾

To: Latvian

Translate

! Only the first 1000 words will be translated. Translation is available without restrictions for public administration users. By authenticating with public administration e-mail, translation is available without restrictions. If you have authenticated with a different e-mail, you can change it in the user profile: User details > New e-mail address. ✕

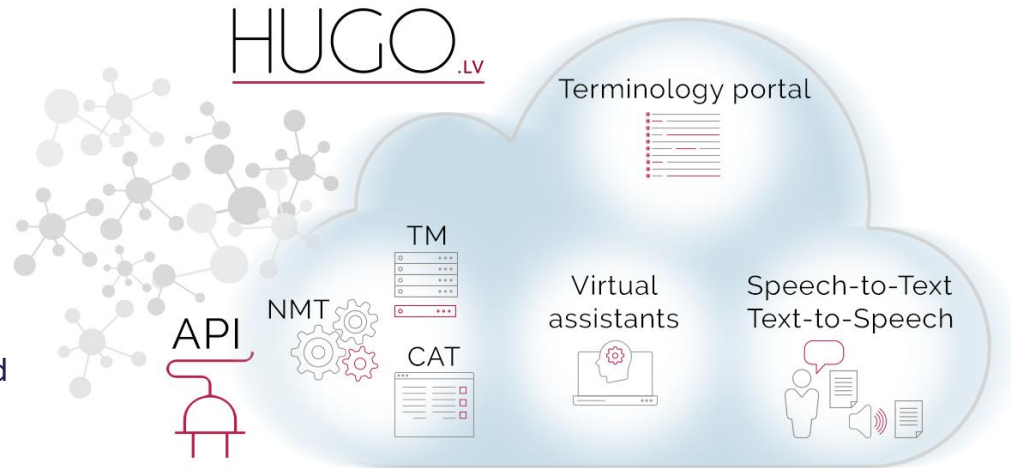
Enter the text or homepage address you would like to translate

Upload or drag a file

Machine translation allows to understand the meaning of a source text, but cannot substitute a human created translation.

Language technology platform for Latvian eGovernment

- Essential **AI-based** translation, speech processing, multilingual communication tools for public institutions, businesses and citizens
- National language technology platform available to **all citizens** for free
- **Instant and secure:** text, document and website translation, speech recognition and voice synthesis, virtual assistants etc.
- Fosters **information inclusivity** and service **accessibility**
- **500+ million** – number of words yearly translated on Hugo.lv
- **85+** Government institutions use **virtual assistants** on HUGO.LV to serve their clients



LLMs as part of National Language Technology & AI Infrastructure



LARGE LANGUAGE MODELS

ANONYMIZATION

MACHINE TRANSLATION

CHATBOTS

SPEECH PROCESSING

NATIONAL LANGUAGE TECHNOLOGY PLATFORM



Common European Language Data Space

Thank you!



A Common European Language Data Space – funded under contract LC-01936389 with the European Union.