

**Christos Ellinides**

Director-General – Directorate-General for Translation, European Commission

# Language Data for European AI: Striking on Two Fronts

# Striking on Two Fronts

1. Modern and technologically revamped Public Administrations
2. EU digital and technical sovereignty

# Introduction to DGT

Over 70-year history

DGT translates into and out of 24 official EU languages

Current output: ±2.5 million pages per year



*People are at the heart of translation*

# From Translations to AI-based services

## DGT's Role in AI & Language Data

Machine translation and AI based multilingual services



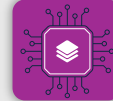
European AI development

DGT's translations throughout the years



**25+ Years of Data**

→ Translation archives



**AI-based services and Machine Translation**

→ Evolution of data usage



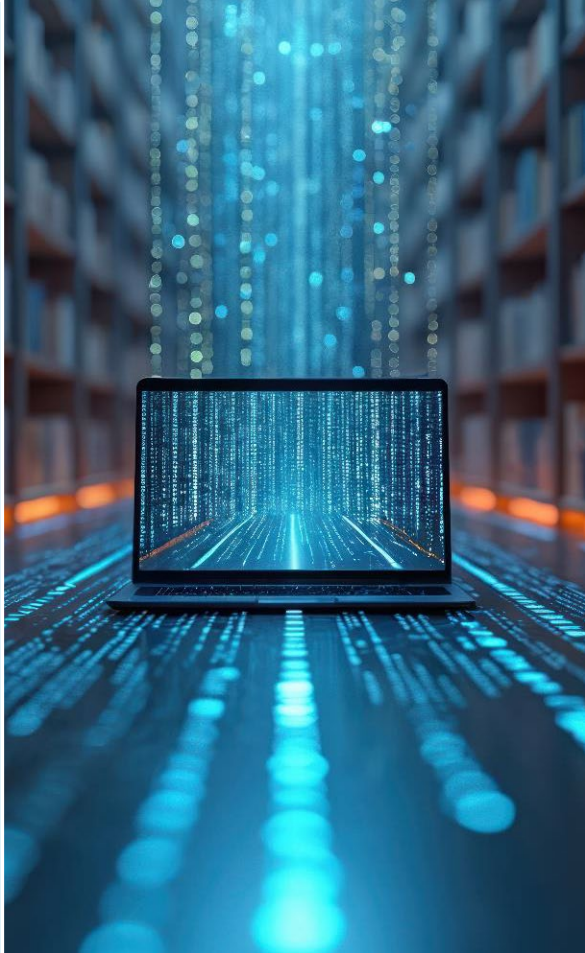
**Multilingual Dataset**

→ Includes low-resource languages



**EU Collaboration on AI**

→ European sovereignty & wealth of languages



# Linguistic data assets: (Euramis and DGT-TM)

## ● Euramis data repository

Launched in 1995

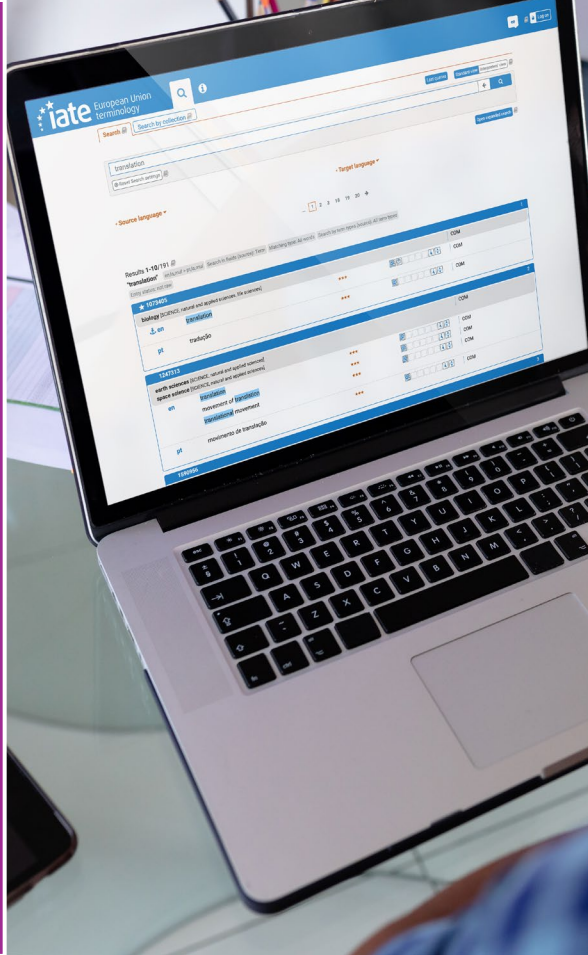
## ● 2 billion sentences

> 300 000 a day

## ● DGT-TM

Launched in 2007

→ used by research, business,  
freelance translators and more



# Expanding into language technologies



## Machine translation

MT@EC -> eTranslation



## ELRC Data Collection (2016-2020)

→ special support for low-resource languages



## Broadening reach

(expansion into non-translation applications: AI services)

# Digital Europe

## AI-Based Multilingual Services

### AI-Based Multilingual Services

These services are made available in the EU under the Digital Europe Programme for use by EU institutions, public administrations, academia, SMEs, NGOs, Digital Europe Programme projects and EPSO candidates.

They offer both web pages and APIs for machine-to-machine access.



#### eTranslation

Neural machine translation built on the EU's history of professional translation.



#### eBriefing

Generate reports from sets of documents in official or general styles.



#### eReply

Have AI help you prepare replies to correspondence, queries, and other requests.



#### eSummary

Quickly find out the main content of long documents.



#### Multilingual Post

Short translations in multiple languages in one shot for X.



#### Speech-to-Text

Upload your media and get full transcriptions or subtitles back.



#### Natural Language Processing Tools (NLP)

Anonymisation, Classification and Named-Entity Recognition.

### Additional useful resources



#### Video guides and help

Simple explanations of our Services



#### Developer's corner

Guidelines and procedures on how to integrate eTranslation in your web pages (only in English)



#### Web translation tool

Free plugin for your website



#### Language Data Space

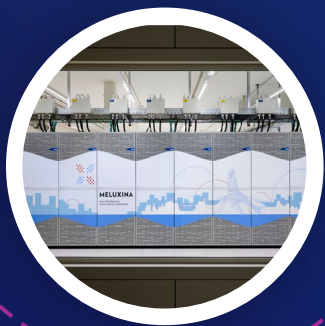
European platform for exchanging languages resources



#### IATE

Interactive Terminology for Europe

# Roadmap for the future



## European HPC Joint Undertaking

→ MeluXina and Leonardo supercomputers

→ EU institutional LLM

- ✓ made in Europe
- ✓ open source
- ✓ institutional, curated data

## European Language Data Space

- European marketplace for language data
- supporting Europe's digital sovereignty





# The political context

- EU AI Act

- Political Guidelines of the European Commission 2024-2029

  - Bolster supercomputing capacity

  - Support industrial uses for AI

  - Step up delivery of AI in public services

- AI Action summit  
(Paris, February 2025)

AI Action summit | Paris, February 2025

*'Action is exactly what we need right now. The time has come for us to formulate a vision of where we want AI to take us, as society and as humanity. And then we need to act and accelerate Europe in getting there.'*



**Ursula von der Leyen**  
President of the European  
Commission



**EU**

**OPEN**

**DATA**

**DAYS**

**19 – 20  
MARCH  
2025**

**Thank you!**