

Empowering Europe with Language Data for European AI

Christos Ellinides

Directorate-General for Translation

(providing language and AI-based services)

20 March 2026

Legal basis

- **Treaty on the Functioning of the EU**
Citizens have a right to address the official EU bodies in any of the EU's official languages and to receive a reply in that language.
- **Council Regulation No 1/58**
Regulations and other documents of general application shall be drafted in the official languages.

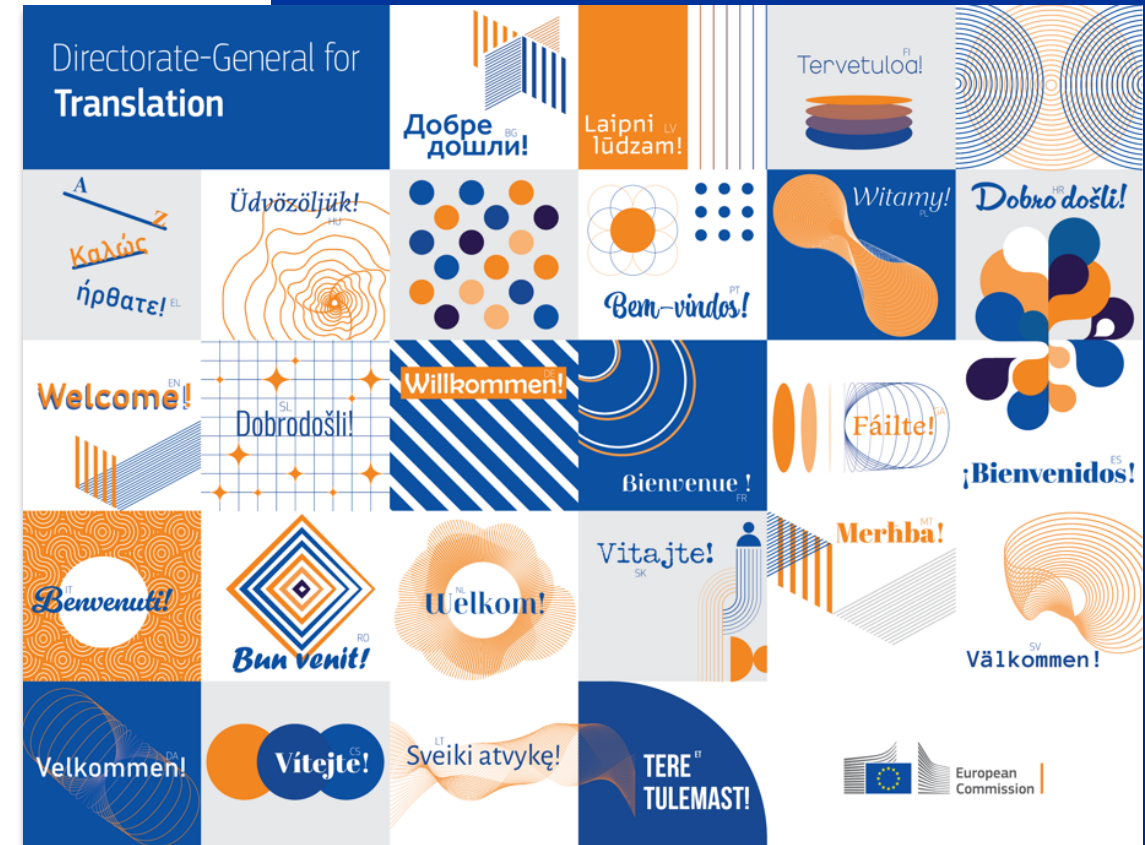


Introduction to DGT

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



EURAMIS

24

languages

100B

tokens

The EU interinstitutional linguistic database

- ✓ High-quality institutional linguistic data
- ✓ Database containing data from all EU institutions since 1996 (3 decades!)
- ✓ Covers all EU official languages and all aspects of EU legislation, policy and other activities.
- ✓ 2 billion sentences
- ✓ 400.000 sentences added every day



From Translations to AI-based services

DGT's Role in AI & Language Data

Machine translation and AI based multilingual services



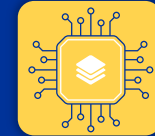
European AI development

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





Expanding into language technologies



Machine translation

MT@EC -> eTranslation (2017)



ELRC Data Collection (2016-2020)

→ special support for low-resource languages



Broadening reach

(expansion into non-translation applications: AI 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.



WebText

Use AI to optimize your text for a website.



Accessible Text

Rewrite your English, French or German text to make it easier to read for people with cognitive difficulties.



eReporting

Have AI help you identify reporting obligations and generate activity reports.



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.



Anonymisation

Replace or redact names, places and more from your documents.

Access and registration

Access to these tools requires registration
EU staff is pre-registered

- Who can access these services
- Registration page
- API registration page

Contact point:

DGT-AI-Language-Services-Advisory@ec.europa.eu

Large Language Models

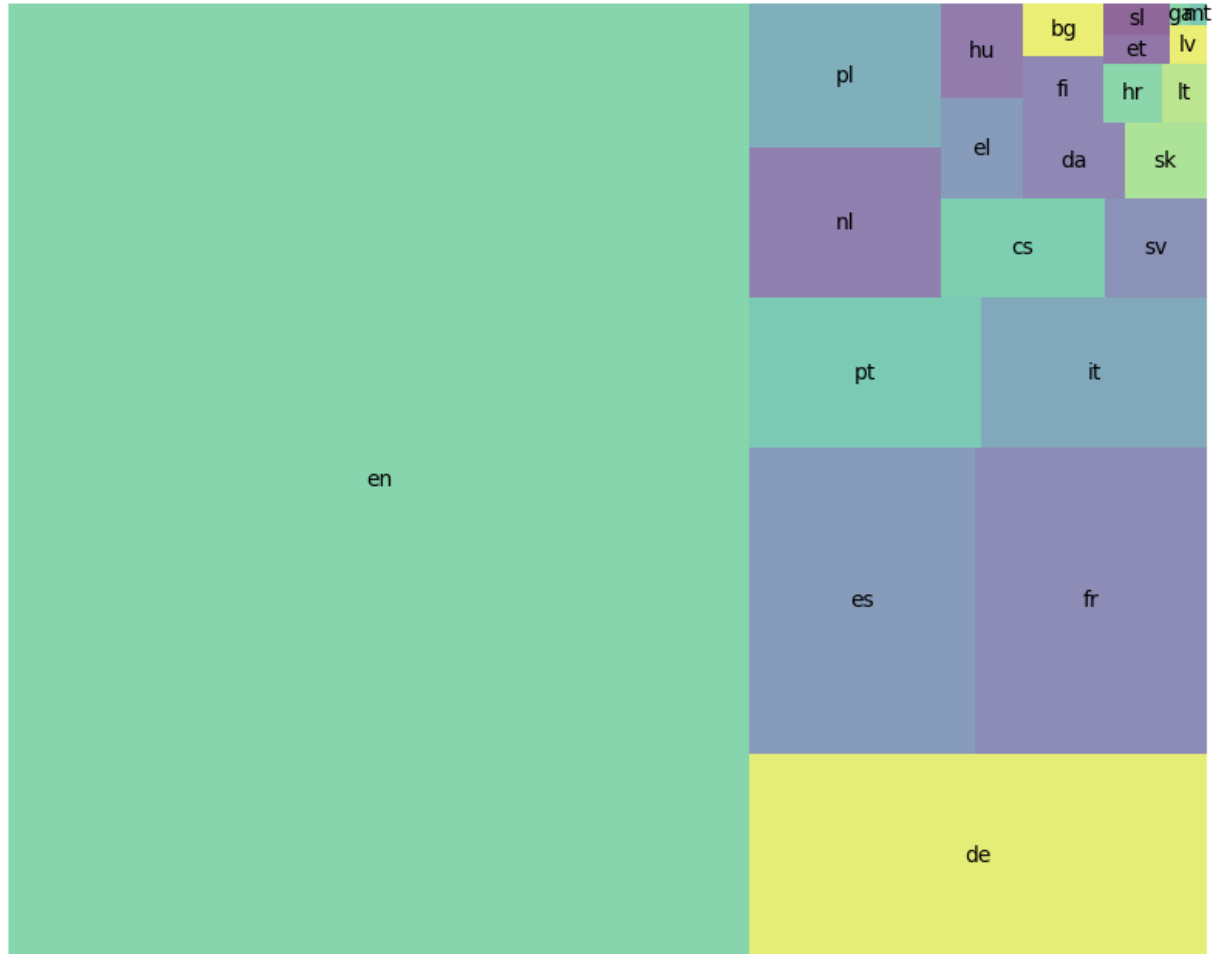
- ✓ Rise of large language models (LLMs)
- ✓ Lack of transparency on training data
- ✓ Biases and influences
- ✓ Lack of control
- ✓ Poor language coverage
- ✓ Own high quality multilingual data
- ✓ Digital Europe programme

LLM

LARGE LANGUAGE MODEL

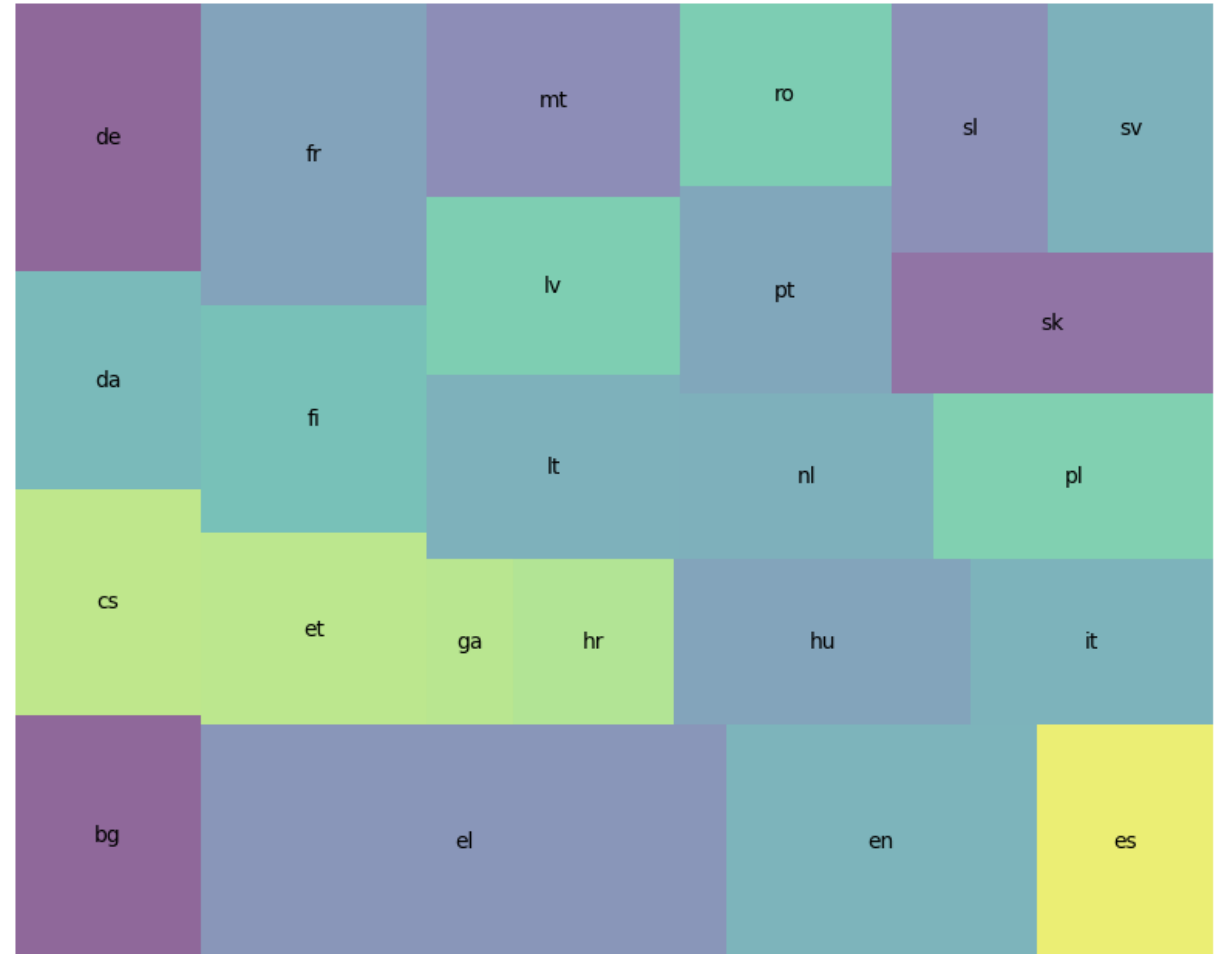
EU language distribution

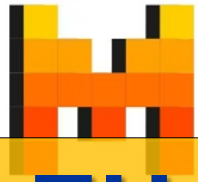
Common Crawl language distribution



EURAMIS

EURAMIS language distribution





**MISTRAL
AI**

EU Institutional LLM

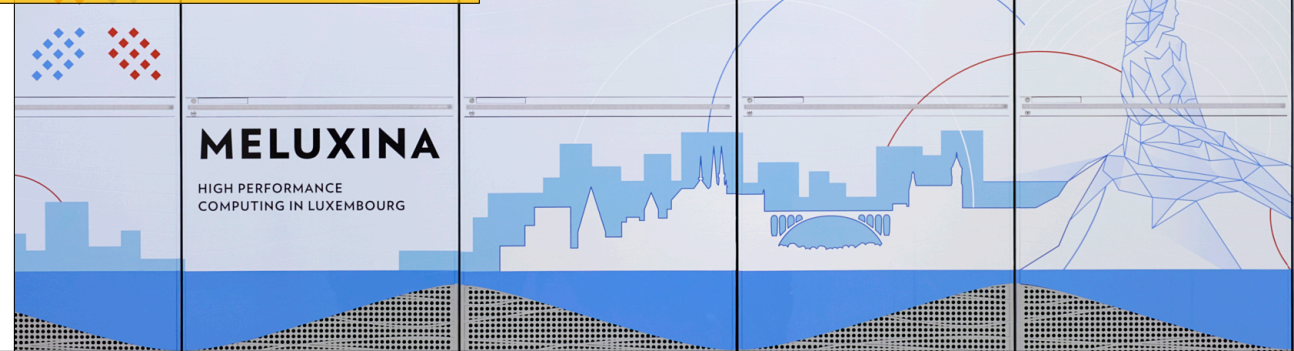
- ✓ institutional data
- ✓ multilingual – EU24
- ✓ enhance underrepresented languages
- ✓ European open source model (Mistral AI, France)
- ✓ EuroHPC supercomputer infrastructure
- ✓ open source: contribution to the EU's AI ecosystem
- ✓ setting an example for the public sector in the EU

EURAMIS

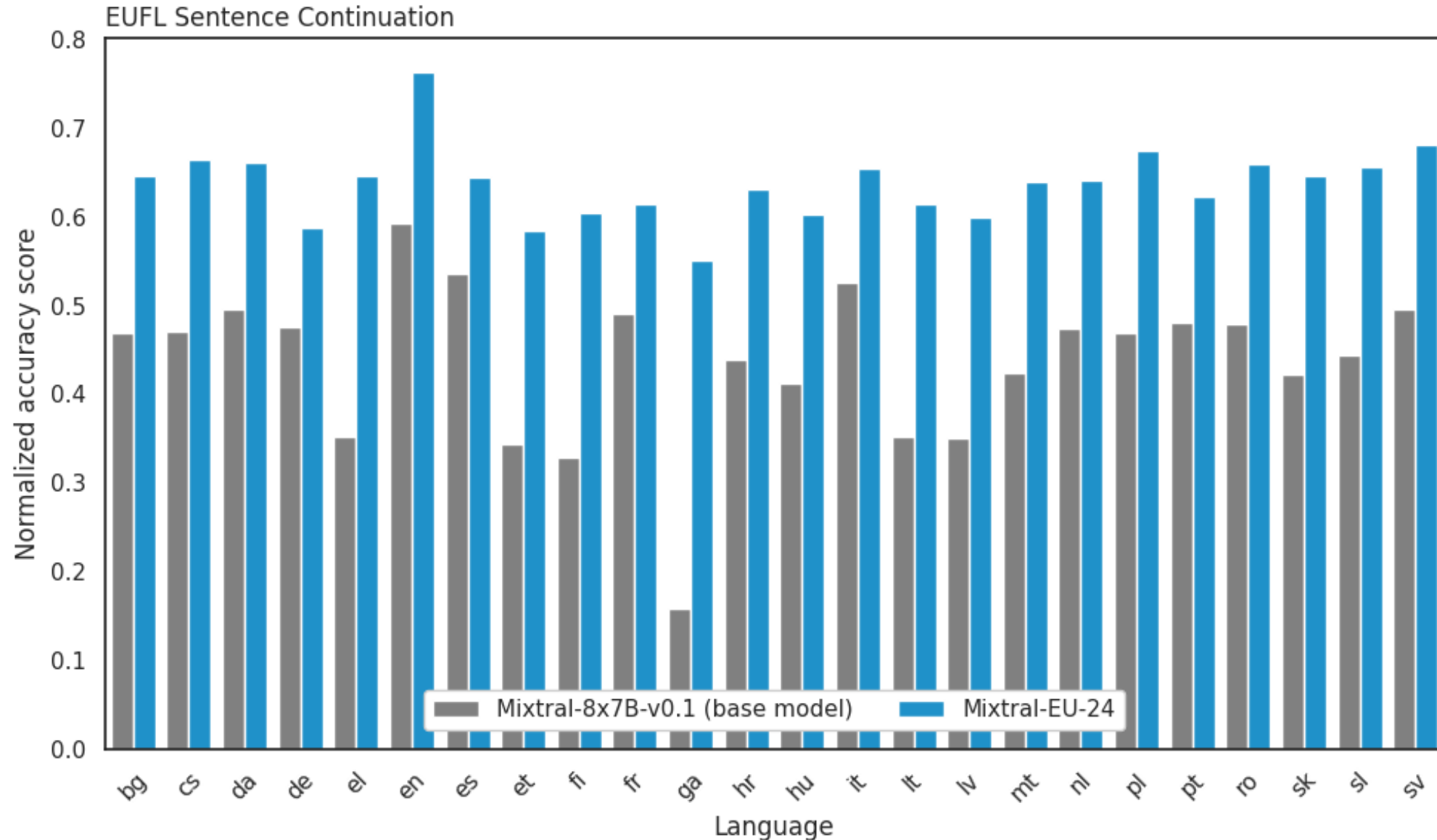
The EU interinstitutional translation database

MN5

MARENÖSTRUM



Evaluation of the EU Institutional LLM





European Language Data Space

The European Marketplace for Language Data.

[Browse the LDS Catalogue](#) 

<https://language-data-space.ec.europa.eu>

Unleash the potential of data – for EU businesses and citizens.

The European Language Data Space aims at building a trustworthy and effective data market for the exchange of language resources in the public and – even more importantly – in the private sector, in line with the [EU Data Strategy](#).



Share

You will be able to share and monetise your language data, language models and other language resources through a single platform, taking European values and compliance with EU regulations fully into account.



Connect

You will be able to connect and exchange with other stakeholders through the European Language Data Space.



Benefit

You will be supported by the European Language Data Space in the development of multilingual and multimodal language technologies and language-centric AI.



Political context

- EU AI Act

- Political Guidelines of the European Commission 2024-2029

Use AI to improve delivery of public services

Exploit untapped data potential and supercomputer infrastructure

- AI Action summit
(Paris, February 2025)

- Apply AI strategy
(October 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.”

AI Action summit | Paris, February 2025



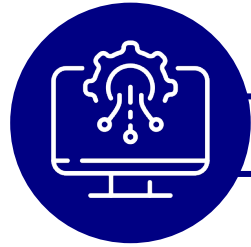
SOMMET
POUR L'ACTION
SUR L'IA

Paris

Mardi 11 février 2025

Towards the creation of an AI continent

(EU digital and technological sovereignty)



Supercomputer infrastructure

Data



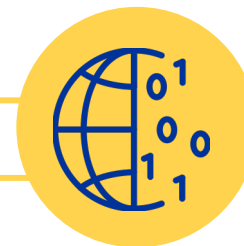
Digital skills

Regulatory simplification



AI Continent

Development of algorithms and adoption



Thank you!

© European Union 2026

Unless otherwise noted the reuse of this presentation is authorised under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.