European Commission actions in Multilingualism & Language Technologies:

The Digital Europe Programme; Common European Language Data Space; ALT-EDIC

LDS workshop in Romania

Philippe Gelin
Head of Sector Multilingualism – DG CONNECT – G3

Presentation outline

- **The Language Technology Landscape**
- **Relevant initiatives from the European Union**

European Commission

Breaking down language barriers in Europe has the potential to boost intra-European trade by up to

EUR 360 billion

Recent study finding, Market study on language technologies

Language Technologies Landscape

Competitiveness, linguistic diversity & culture

European Commission

Advantages of Language Technologies in *Economic* Terms

Think about Chat(GPT)
What is pro?
What is con?

Increasing global reach and expand market share →

- creating new market opportunities while enhancing the competitiveness and visibility of European businesses.

Automating language-based tasks and improving accuracy →

- reducing execution time, costs and errors.

Safeguarding Europe's digital and technological sovereignty →

- creating a B2C and B2B secure and privacy-compliant technology, while incentivising Europe's LT industry.

European Commission

Advantages of Language Technologies In *Cultural* Terms

Think about Chat(GPT)
What is pro?
What is con?

Improving access to information and knowledge →

- promoting inclusion and extending/supporting participation in the democratic discourse.

Maintaining low-resource languages & diversity →

- revitalising such languages while promoting cultural diversity and sensitivity.

Facilitating communication and collaboration across backgrounds →

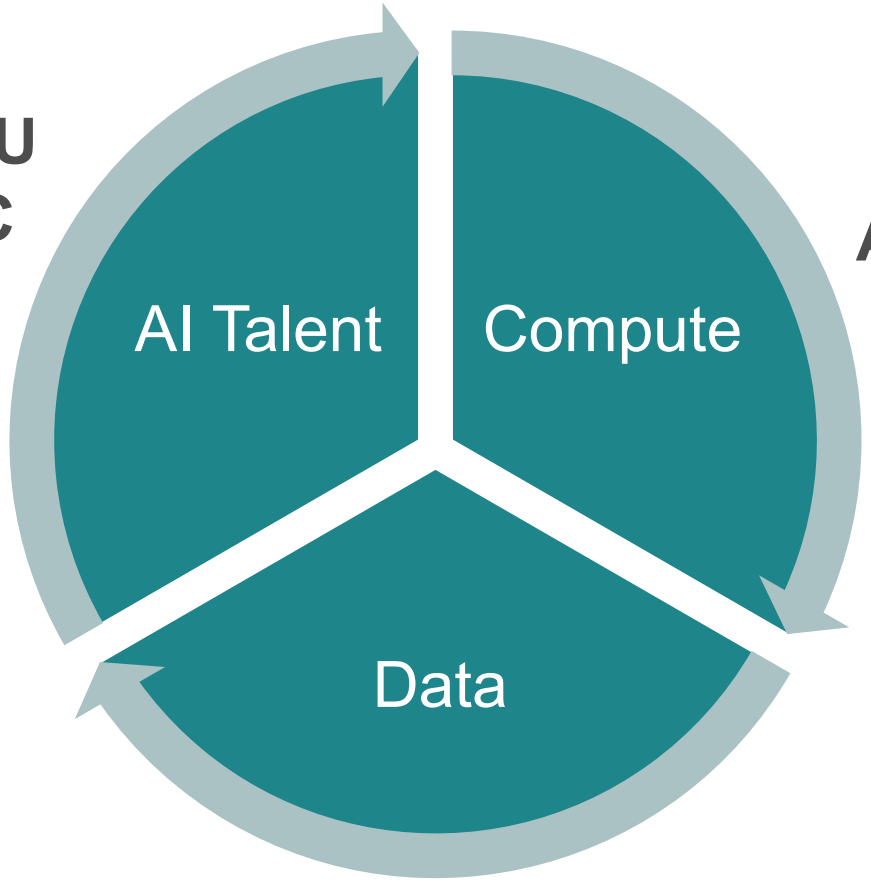
- breaking down language barriers and enabling cross-cultural dialogue and exchange.
- achieving efficiency in the multilingual single market

→ *Turning diversity into opportunities*

European Commission

European Commission's Actions

GenAI4EU
ALT-EDIC

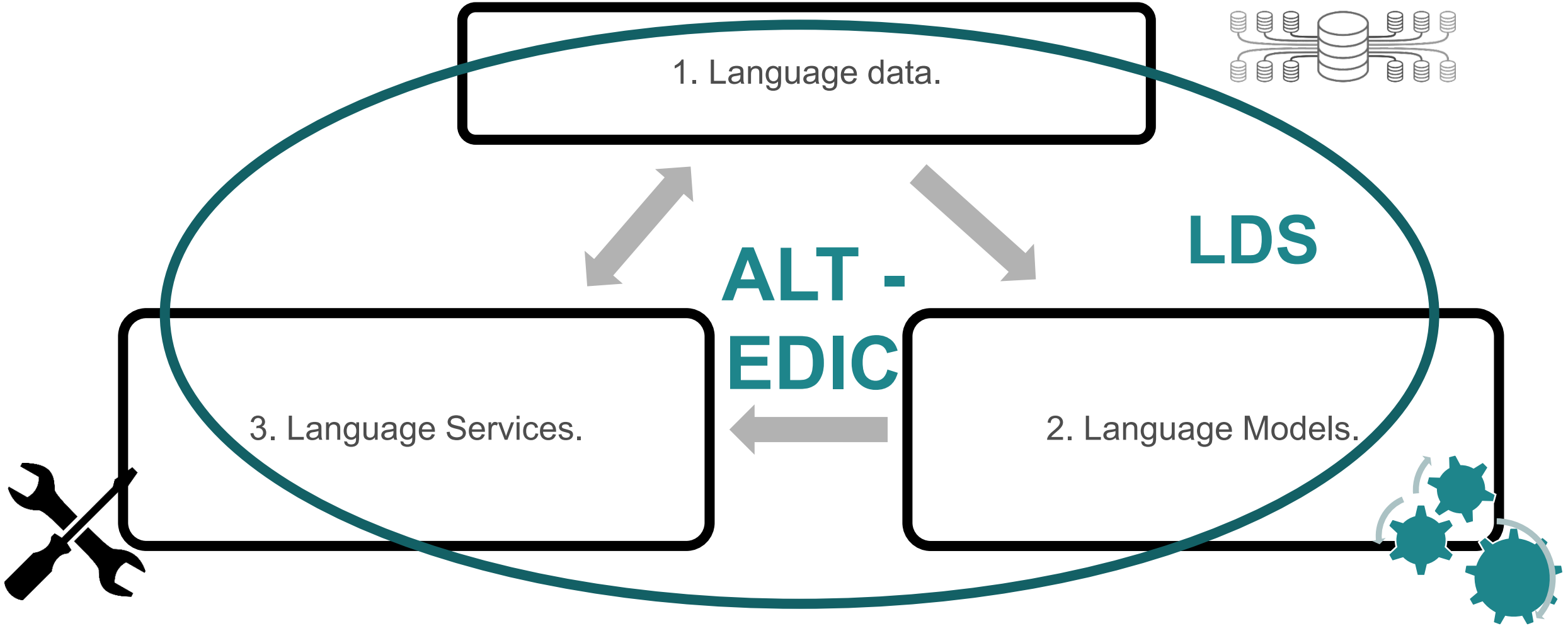


AI Factories

LDS
ALT-EDIC

European Commission

European Language Technologies Landscape



Relevant initiatives from the European Union perspective

Policy

AI-based Multilingual Services

Language Data Space

ALT-EDIC

European Commission

Recent policy documents

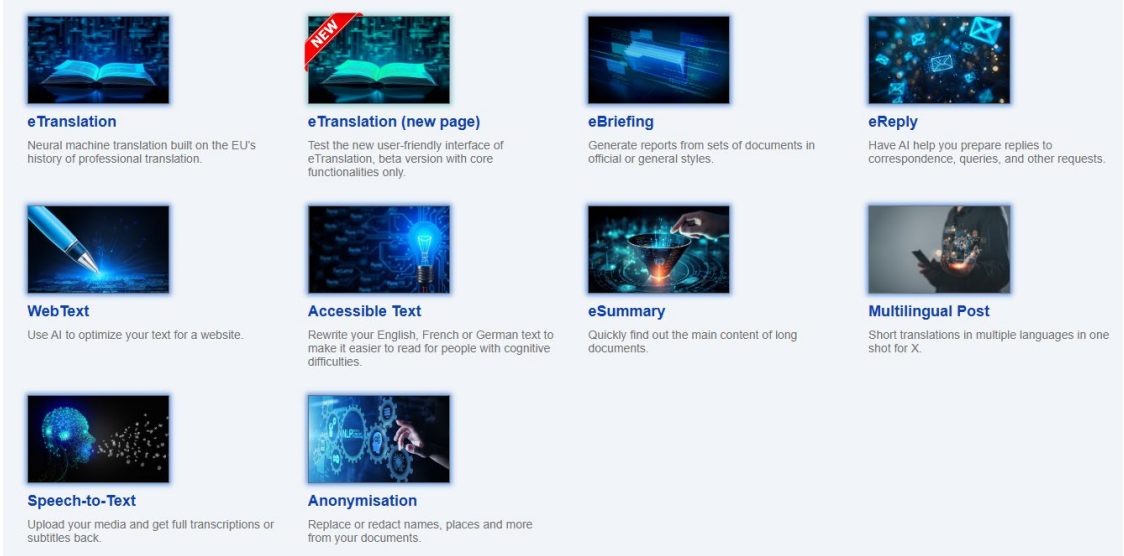
& Digital Omnibus Regulation Proposal for simplifying digital legislation

European strategy for data	Digital Decade Policy Programme	AI Continent Action Plan, Apply AI Strategy	Data Union Strategy
Data spaces should foster an ecosystem creating new products and services based on more accessible data	*Guides digital transformation in Europe, including federating co-operation between countries*	*Crucial link with data resources, data spaces, and data labs* *Data for supporting AI innovation*	*Updated data strategy for EU, published 2025*

European Commission

AI-based Multilingual Services

- Tools that have been developed by the Commission
- They are available in EU for free for the following types of entities: EU institutions, public administrations, academia, SMEs, NGOs, Digital Europe Programme projects and EPSO candidates
- Both web pages and APIs for machine-to-machine access



European Commission

The Common European Language Data Space (LDS)

- Background
- Overview
- Architecture

European Commission

Data spaces → European Strategy for Data

The aim of the strategy: To create a Single European Market for Data

- **The EU aims to enable data to flow freely across borders and sectors.**
- **The objective is to unlock high-quality data for innovation, AI, economic growth, and public value creation.**

While ensuring European Values

- **Data spaces are designed to operate under EU rules and values, including:**
- **IP rights and data protection,**
- **fairness &transparency,**
- **security,**
- **consumer protection,**
- **and competition law.**

Data spaces → the Data Union Strategy

Background

- **Building on the European Strategy for Data**
- **A scarcity of data for AI development**
- **Data increasingly seen as a strategic asset for technological and industrial advantage.**

Actions

- **Made proposals to further expand data spaces**
- **Scaling up access to data for AI**
- **Streamlining data rules**
- **Strengthening the EU's global position on international data flows**

The Language Data Space

- **A sector-specific implementation of the strategies for secure data sharing.**
- **LDS and ALT-EDIC** will **crowdsource** domain-specific datasets including smaller languages

Legislative instruments

The Data Governance Act

•**Applicable since September 2023.**

•**Created mechanisms for trustworthy data sharing** (mostly on voluntary basis).

•**Data altruism:** Concept and framework for voluntary sharing of data by companies for purposes of general interest.

•**Data intermediaries: Neutral third parties and trusted entities that connect data holders with data users.**

The Data Act

Applicable since September 2025.

Ensures fairness in the allocation of the value of data amongst the actors in the data economy.
Obligations to share data B2C / B2B
(/ B2G)
Especially with regard to IoT data
Ground rules for conditions (compensation, contractual terms, etc.)

Provides legal clarity on *who* can use *what* data and under *which* conditions.

Language Data Space – Overview

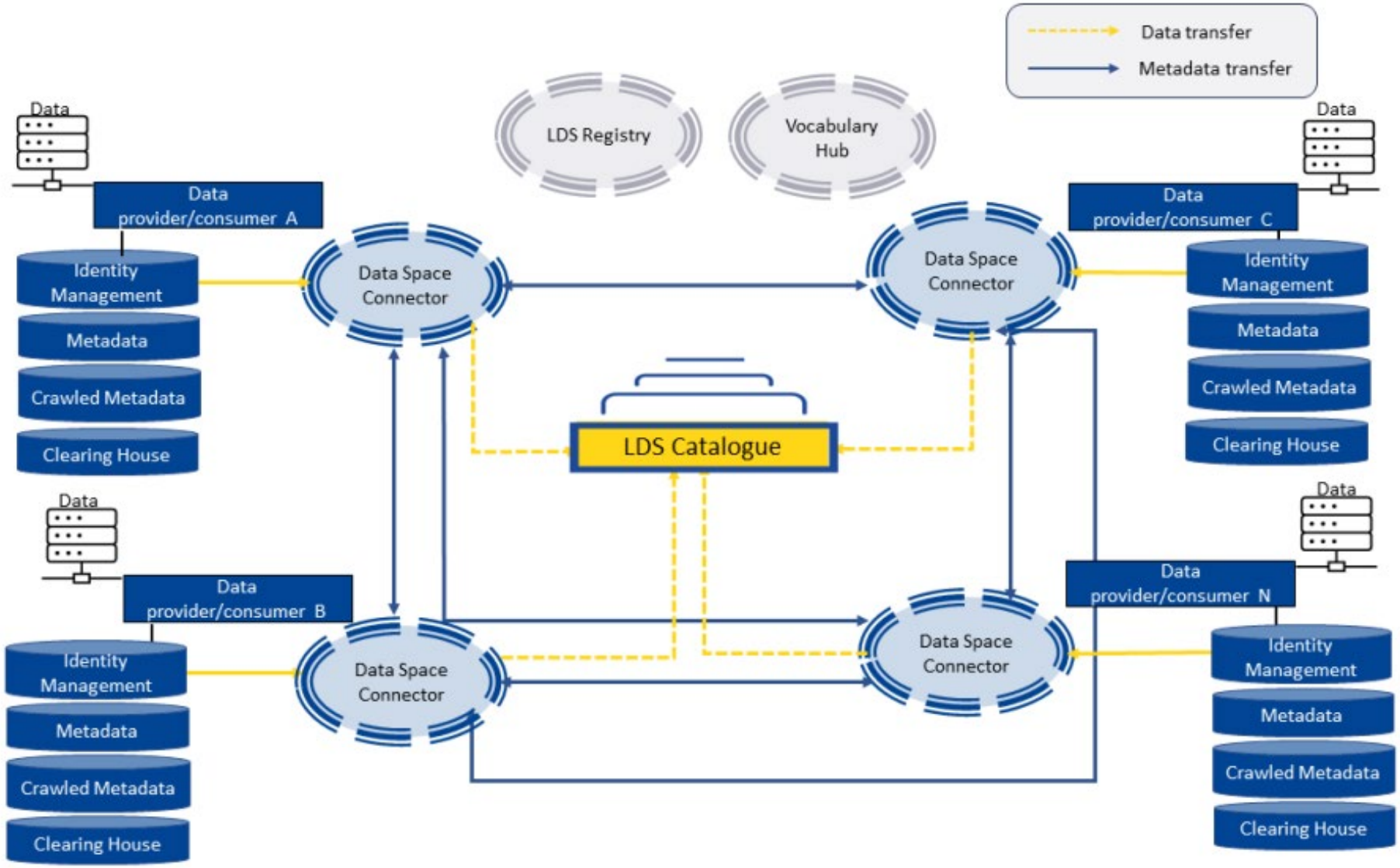
EUROPEAN LANGUAGE DATA SPACE

What	**Platform** and **marketplace** for the collection, creation, sharing and re-use of multilingual and multimodal language data	
	8 Mio € **Procurement** Contract from the **DIGITAL** Work Programme 2021-2022	
Why	Increase **availability** of clean, high-quality, compliant language data	
	Support the development of state-of-the-art European language technologies	
Who	**Consortium**: DFKI, ELDA, TILDE, Athena Institute	
	Stakeholders: Industry, Public Administration, Research, etc.	
How	- Technical Architecture and Infrastructure	- Openness
	- Governance Framework	- Respect of EU Rules and Values
	- Promotion	- Overcoming Legal Barriers (IPR, GDPR, etc.)
When	Started in **January 2023** for Three years – until end of 2026.	
	Full version in **operation**	

European Commission

Language Data Space – Architecture

EUROPEAN LANGUAGE DATA SPACE

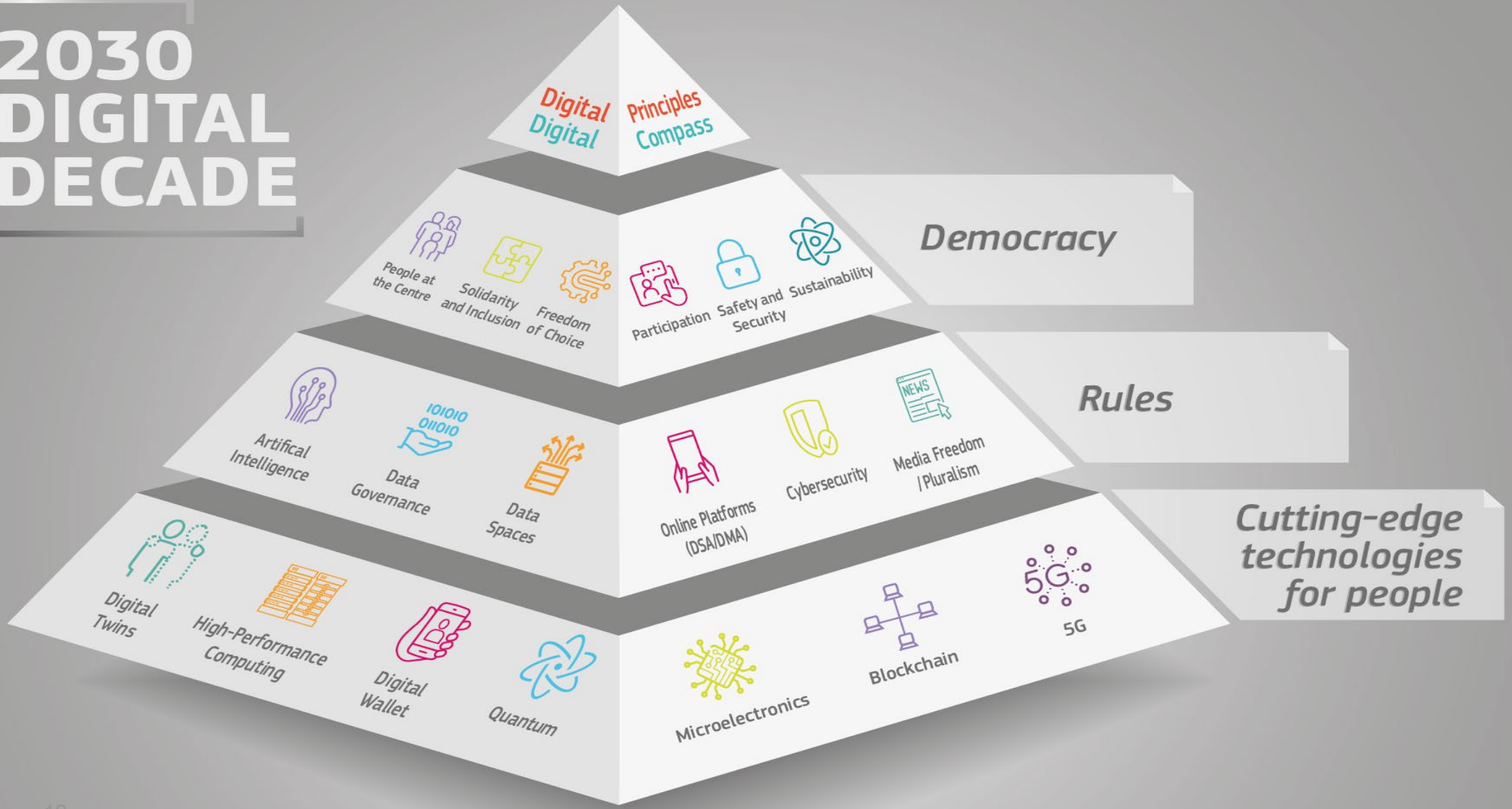


European Commission

Alliance for Language Technology EDIC (ALT-EDIC)

European Commission

2030 DIGITAL DECADE



#DigitalEU

ALT-EDIC

European **D**igital **I**nfrastructure **C**onsortium

which supports the **language technology** ecosystem

by **federating** and strengthening

the production of European language **resources**

and the **evaluation** of language technologies

European Commission

ALT-EDIC

Alliance for Language Technologies
European Digital Infrastructure Consortium

Objectives

- Preserve linguistic and cultural diversity in Europe
- Technological leadership and strategic autonomy
- Respect European rules and values
- Cooperation
- Raising awareness

European Commission

ALT-EDIC

Alliance for Language Technologies
European Digital Infrastructure Consortium

Objectives

- Preserve linguistic and cultural diversity in Europe
- Technological leadership and strategic autonomy
- Respect European rules and values
- Cooperation
- Raising awareness

Participation

Members [17]

FR, BG, CZ, DK, ES, FI, GR, HR, HU, IE, IT, LT, LV, LU, NL, PL, SI

Observers [8]

AT, BE, CY, EE, MT, PT, RO, SK, SE

Interested [2]

DE, UA

BUDGET

~80M€

European Commission

ALT-EDIC

Alliance for Language Technologies
European Digital Infrastructure Consortium

Objectives

- Preserve linguistic and cultural diversity in Europe
- Technological leadership and strategic autonomy
- Respect European rules and values
- Cooperation
- Raising awareness

Participation

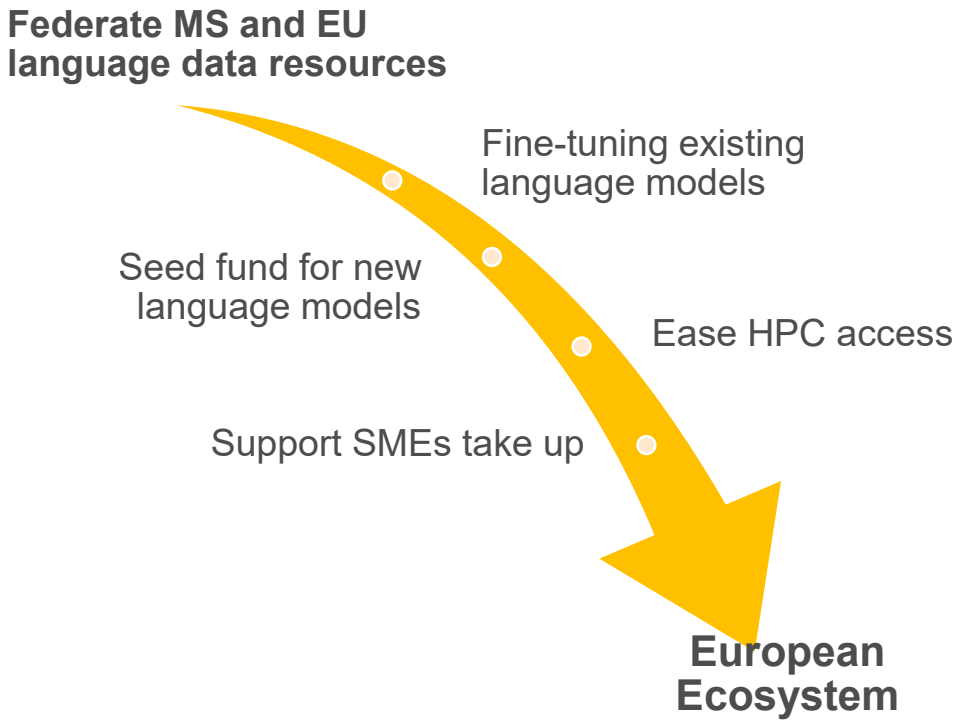
Members [17]
FR, BG, CZ, DK, ES, FI, GR, HR, HU, IE, IT, LT, LV, LU, NL, PL, SI

Observers [8]
AT, BE, CY, EE, MT, PT, RO, SK, SE

Interested [2]
DE, UA

BUDGET
~80M€

Actions Plan



European Commission

ALT-EDIC activities & action plan

Data:

- Supports development and training of language models
- Leverages the capabilities of Language Data Space (LDS)
- Federates resources and existing language data from EU & MS
- Ensures a diverse and inclusive linguistic landscape & addresses data scarcity

Existing models:

- Creates a repository of open-source LLMs
- Develops specific methods for fine-tuning

New Models:

- Acts as pool seed fund
- Brings together public and private resources
- Facilitates access to HPC
- Launches and coordinates new LLM projects

Evaluation, certification, normalisation:

- Introduces robust evaluation methodologies for evaluating and certifying language models
- Discrimination bias
- Supports institutions for investing in LTs

Ecosystem:

- Advisory point & guidance for public administration
- Cultural programme around AI for languages, reaching the public (discussion & awareness)
- Enabling LT end-users to participate in the Language Technology landscape

DIGITAL Work Programme 2024 (4 projects)

ALT-EDIC4EU

€ 4 millions
Coordination and Support Action (100%)

Support the **Coordination** of a European Language Technologies Ecosystem

LLMs4EU

€ 20 millions
Simple Grant (50%)

Support the Collection of Language Data

Support Fine-tuning of Large Language Models

OPEN Euro LLM

€ 20 millions
SME Grant (75%)

Develop and make available one open-source large language foundation model

LLM-BRIDGE

€ 6.8 millions
Simple Grant (50%)

LLM Booster for Regional Innovation and Digital Growth in Europe. Startup accelerator

European Commission

One Last Thought: Why it matters?

- Joint forces with the other 18+ Member States to:
  - Collect data;
  - Co-develop models and services;
  - Contribute to and network within the Ecosystem
- Being a member of ALT-EDIC allows to
  - Support Local Economy in adopting multilingual AI solutions and exporting them
  - Ensure your language is represented in European AI and language models
  - Benefit from FSTP (cascade funding) opportunities in ongoing Digital Europe projects

European Commission

Thank you

© European Union 2023

Unless otherwise noted the reuse of this presentation is authorised under the CC 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.

European Commission