



The Alliance for Language Technology EDIC and the Common European Language Data Space

LDS workshop in Romania

Philippe Gelin
Head of Sector Multilingualism & Language Technologies - DG CONNECT – G3

Language Technology Landscape

LT Crucial for Preserving Identity and Competitiveness

Maintaining languages, cultures, and diversity

Achieving efficiency in a multilingual Europe

Turning diversity into opportunities

Language Technology Landscape

LT Crucial for Preserving Identity and Competitiveness

Maintaining languages, cultures, and diversity

Achieving efficiency in a multilingual Europe

Turning diversity into opportunities

European Commission's Actions

GenAI4EU

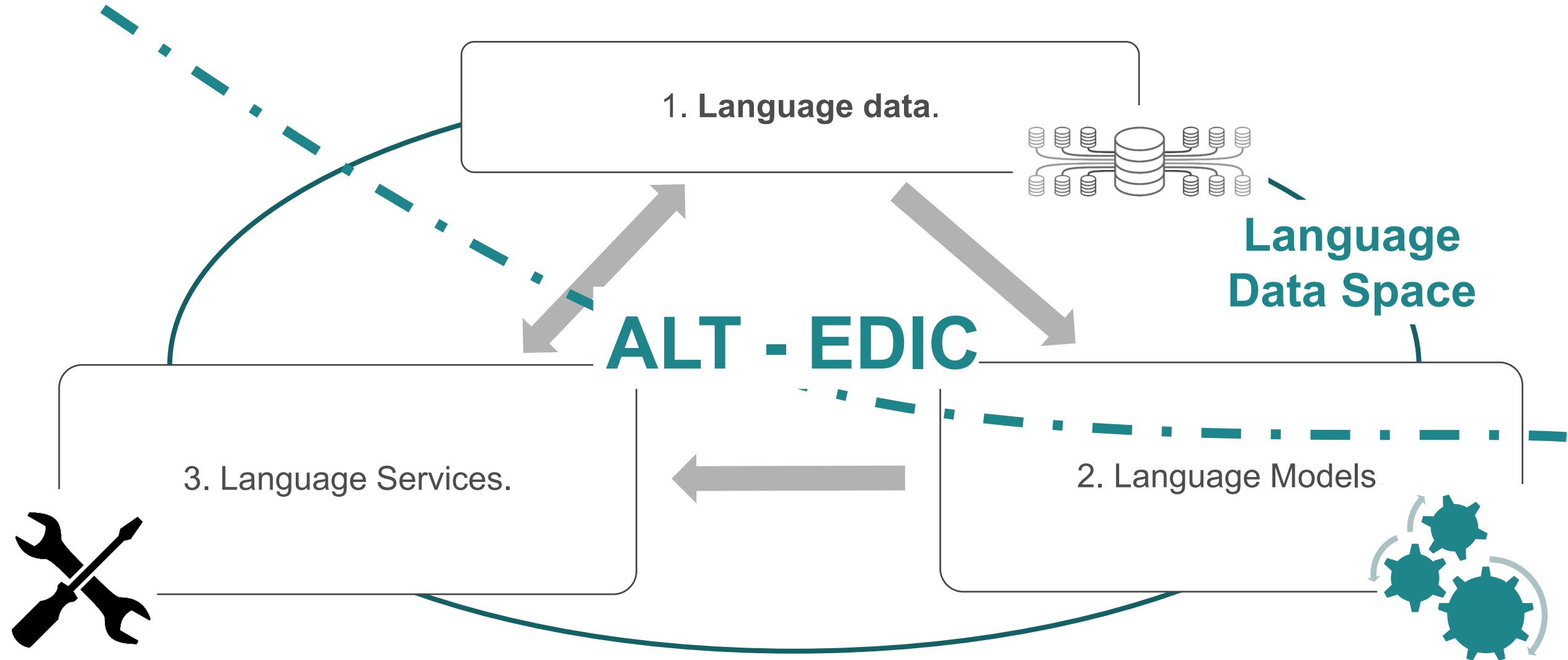
AI Factories

AI Talent

Compute

Data

European Language Technologies Landscape



The Common European Language Data Space (LDS)

- Overview
- Architecture

Language Data Space – Overview



What

Platform and **marketplace** for the collection, creation, sharing and re-use of multilingual and multimodal language data

6+2 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

Governance: Member States representatives and Industry User group

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

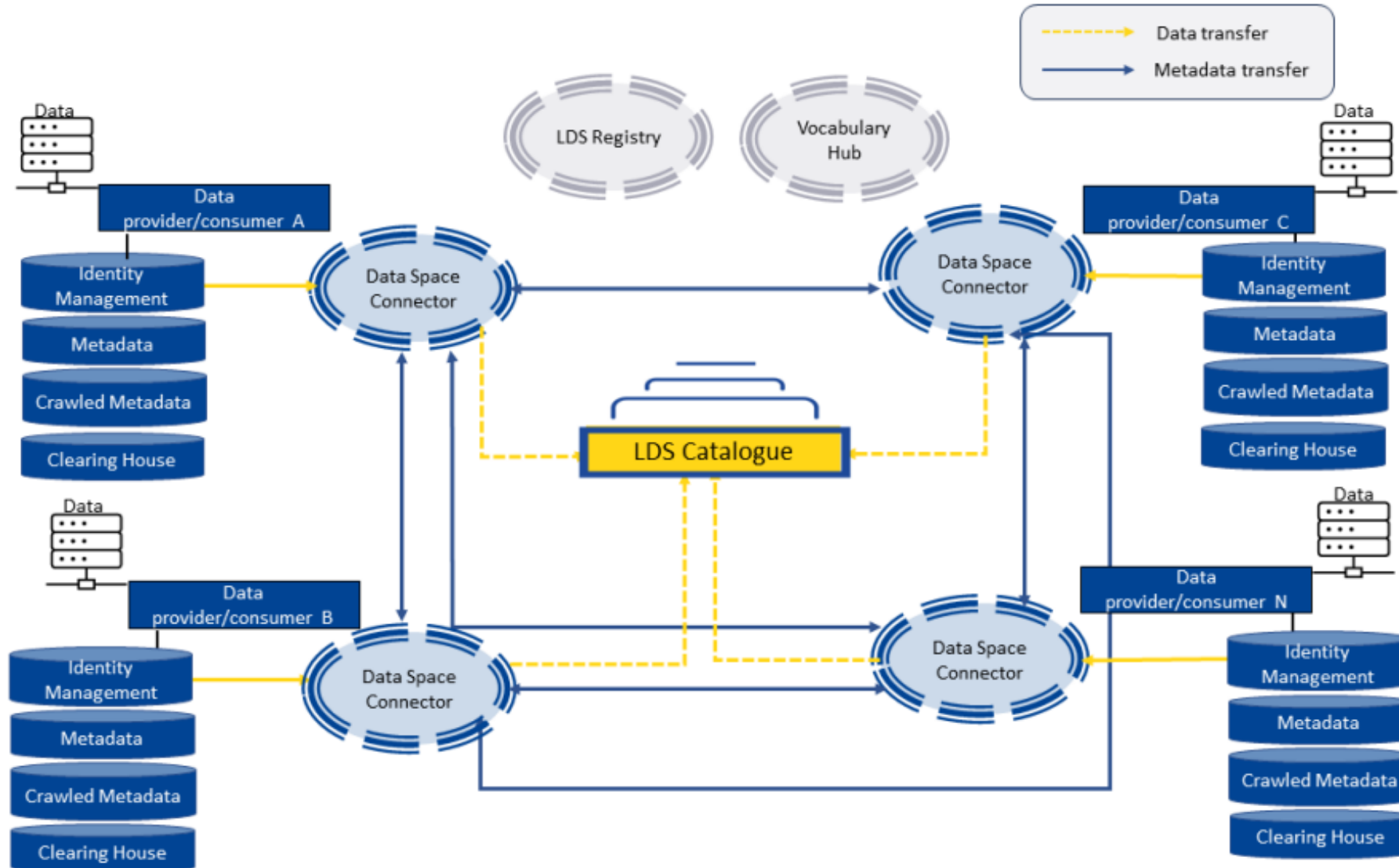
Started in **January 2023** for 3+1 years – until end 2026

Beta version in **Q1 2025** - **Fully operational in January 2026**

Language Data Space – Architecture

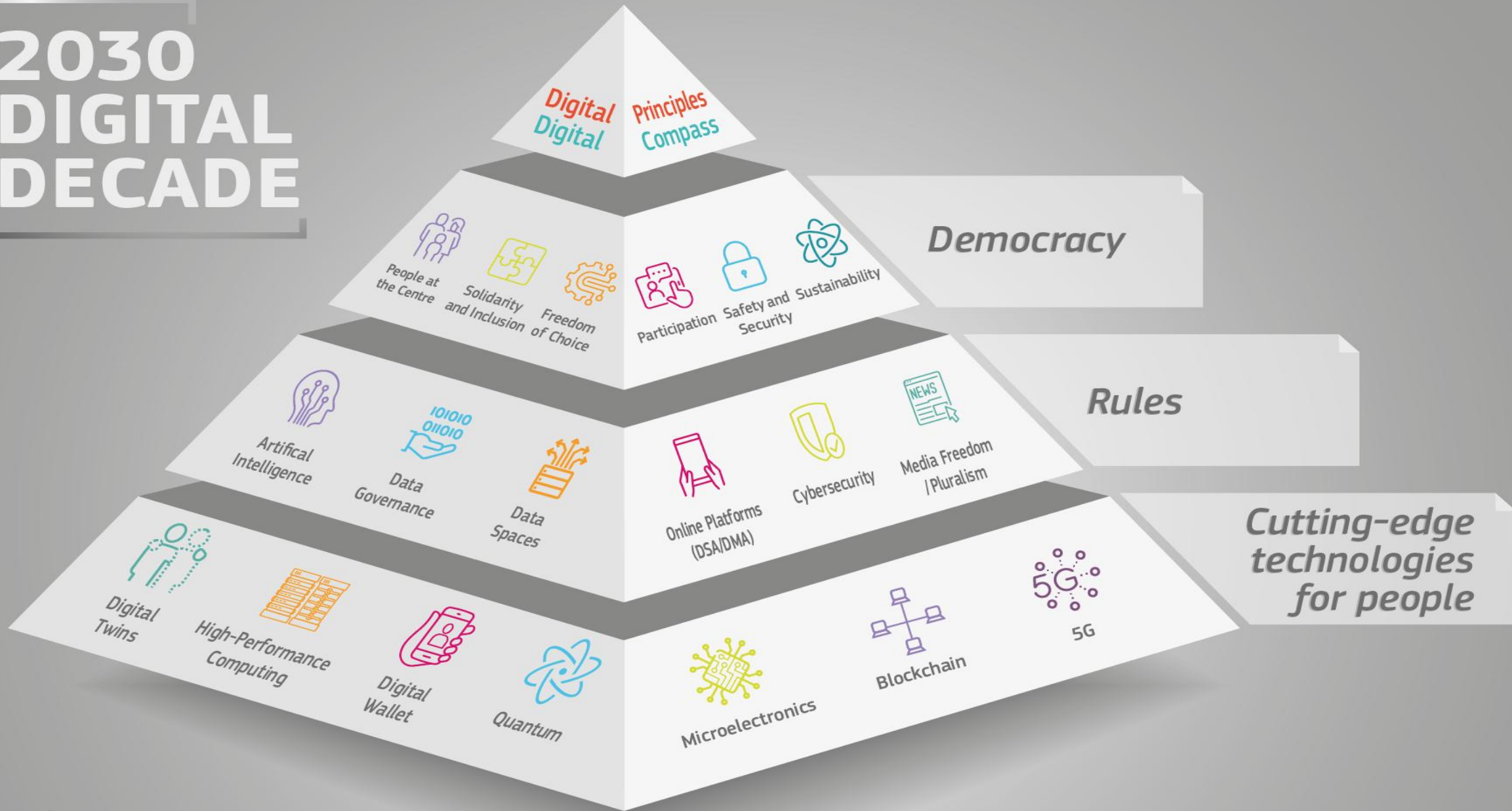


EUROPEAN
LANGUAGE
DATA SPACE



Alliance for Language Technology EDIC (ALT-EDIC)

2030 DIGITAL DECADE



Objectives

Preserve linguistic and cultural diversity in Europe

Technological leadership and strategic autonomy

Respect European rules and values

Cooperation

Raising awareness

Participation

Members [18]

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

Observers [8]

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

Interested [2]

DE

Budget

Projects (Costs)

- LLMs4EU	€ 40 Mio
- ALT-EDIC4EU	€ 4 Mio
- OpenEuroLLM	€ 40 Mio
- LLM-BRIDGE	€ 6,8 Mio

Actions Plan

Federate MS and EU language data resources

Seed fund for new language models

Fine-tuning existing language models

Ease HPC access

Support SMEs take up

European Ecosystem

DIGITAL Work Programme 2024 (3 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

OPENEuroLLM

€ 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

One Last Thought: Why Joining matters?

- Join forces with 18+ Member States to co-develop models and services
- Support Local Economy in adopting multilingual AI solutions
- Ensure your language is represented in European AI and language models
- Benefit from FSTP (cascade funding) opportunities in ongoing Digital Europe projects
- Have your voice (voting rights)
- Less than 400.000 RON/year, for the whole country.
- Observing is good, but participating is better!

Thank you



© European Union 2023

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.