



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

LDS workshop in Malta

Dhafer LAHBIB
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

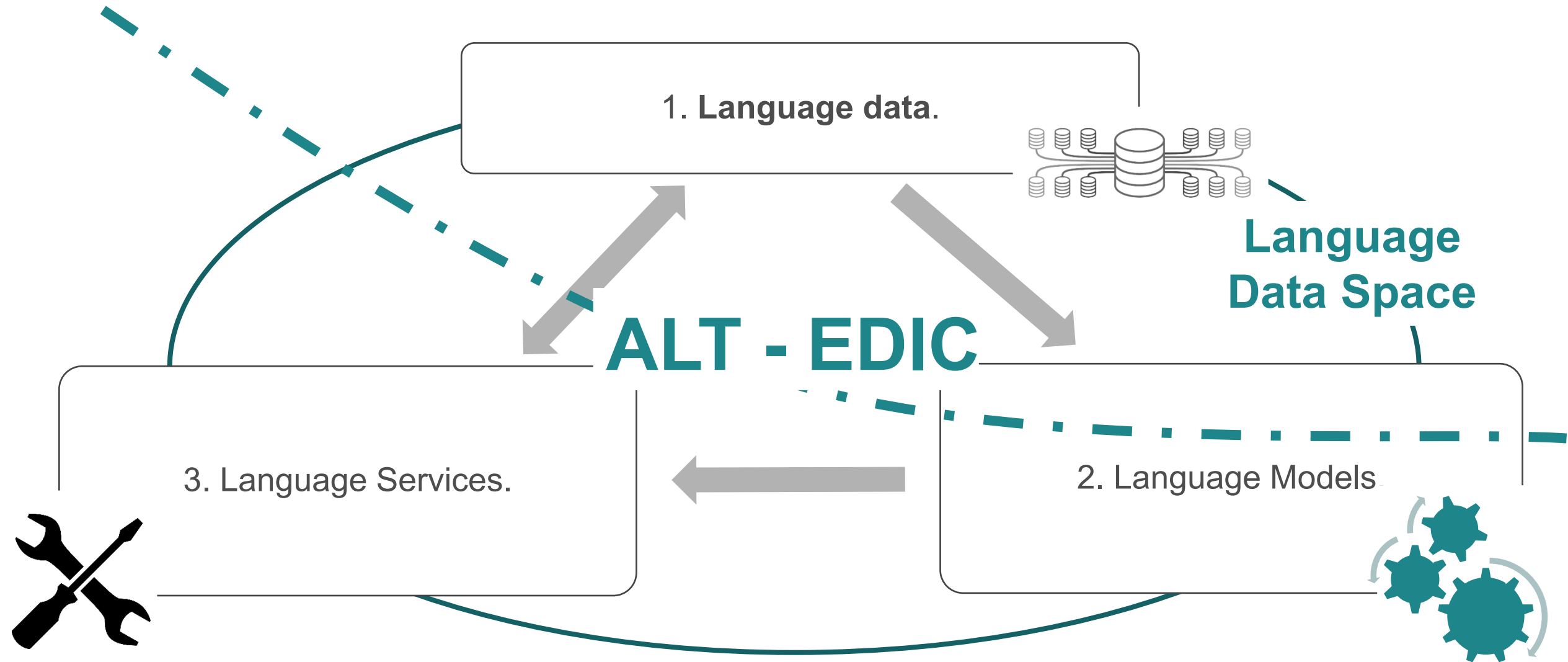
AI Factories

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 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

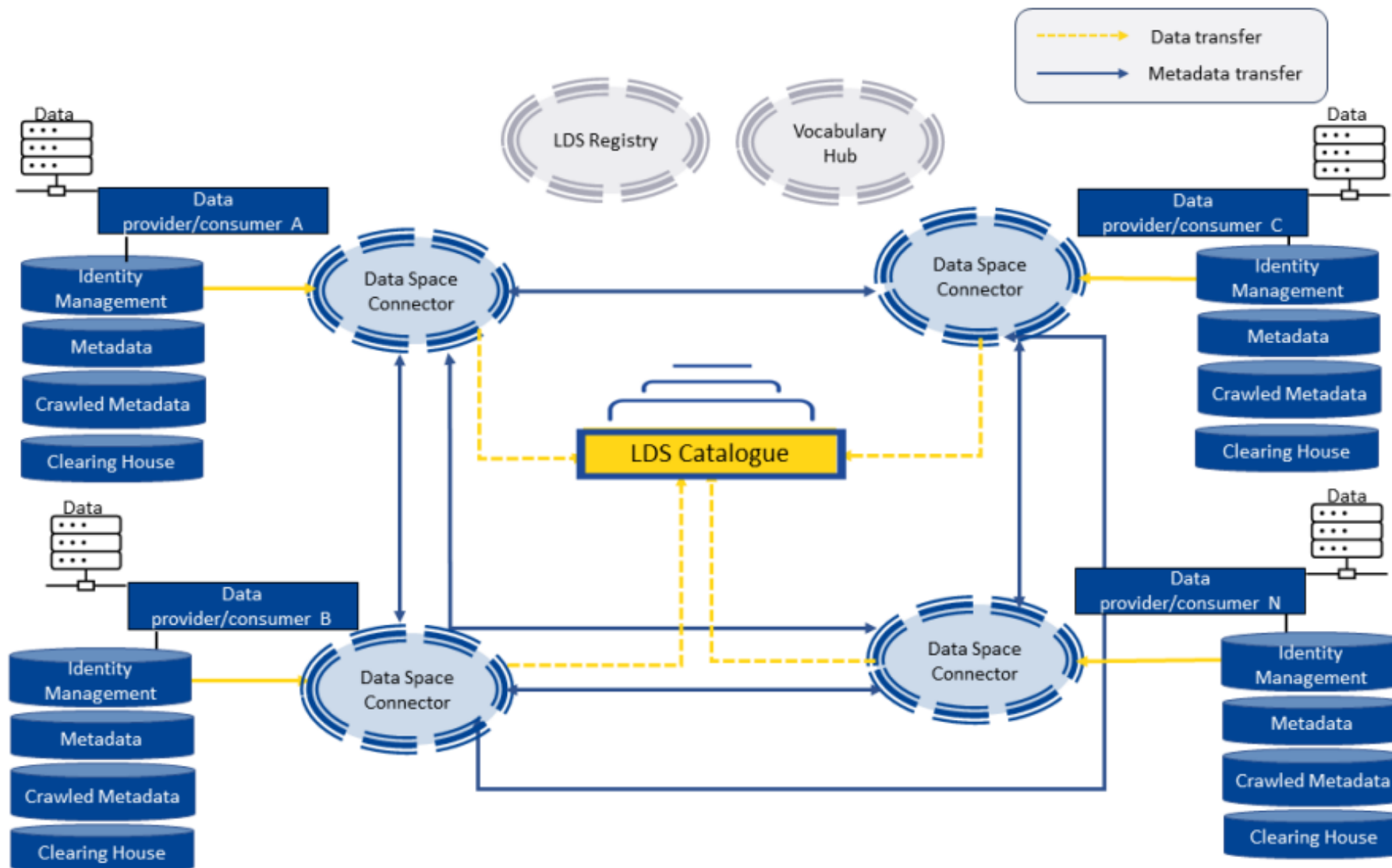
Started in **January 2023** for Three years – until end 2025 (Possible contract renewal for another year).

Beta version in **Q1 2025**

Language Data Space – Architecture

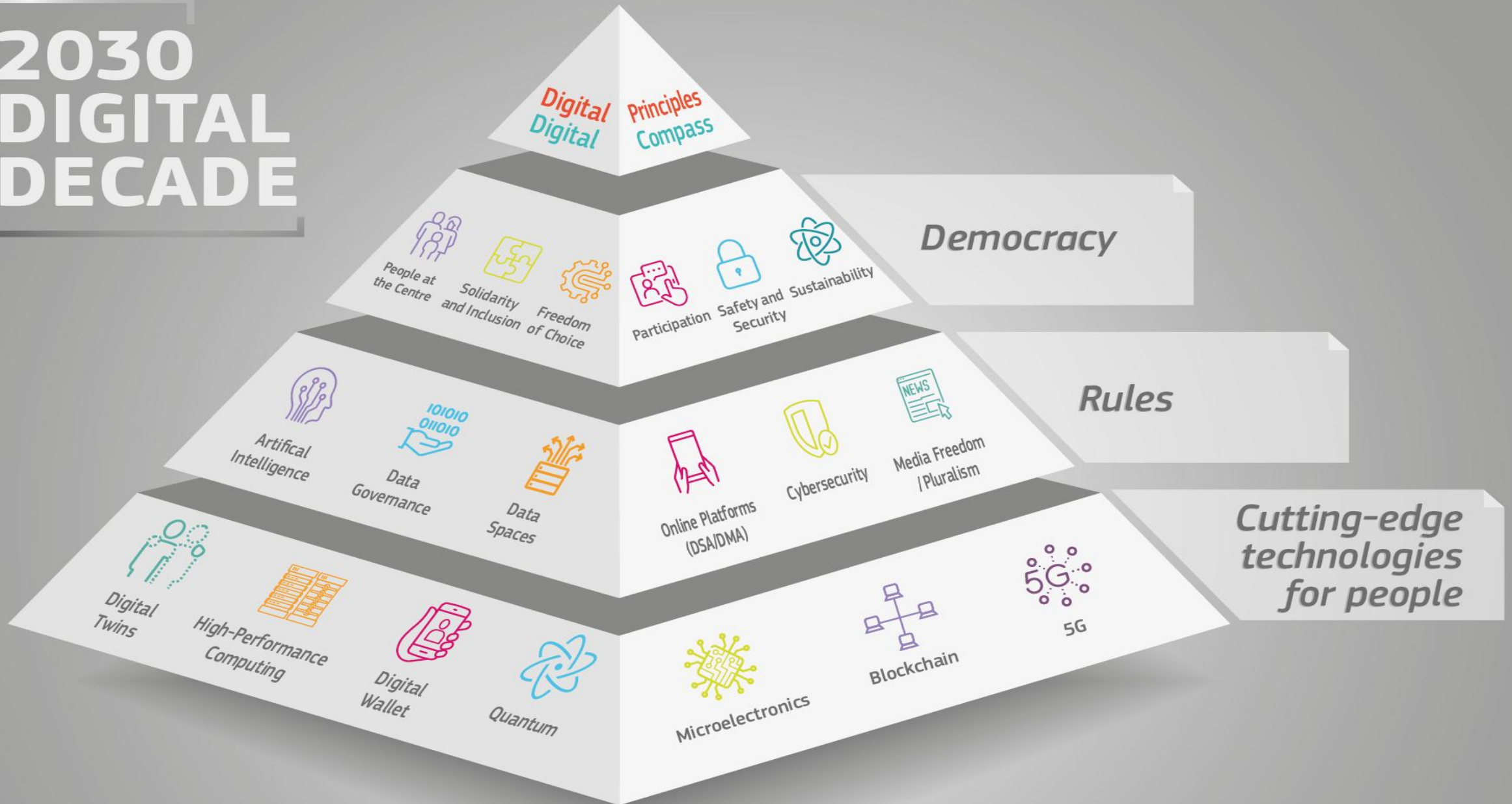


EUROPEAN
LANGUAGE
DATA SPACE



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, SK

Interested [2]

DE, SE

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

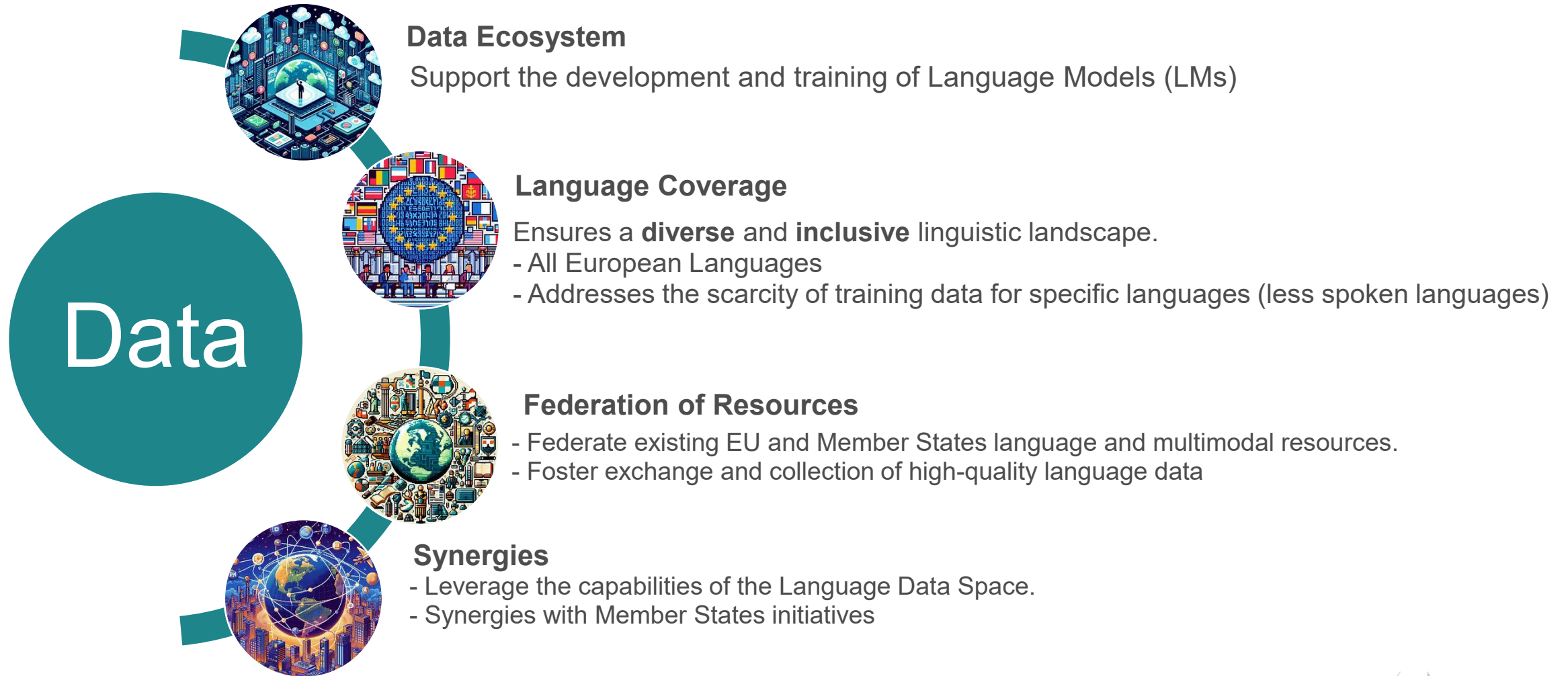
Fine-tuning existing
language models

Ease HPC access

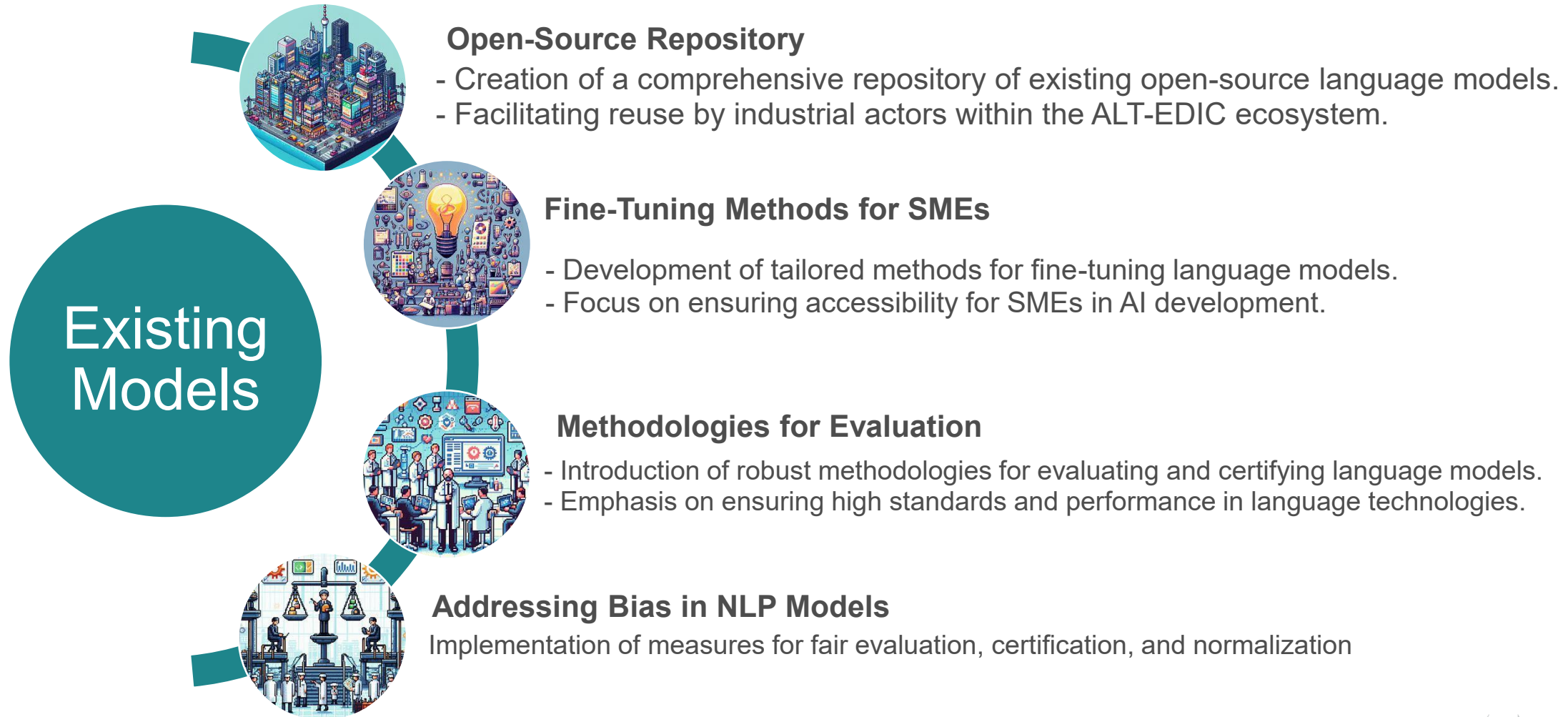
Support SMEs take up

European
Ecosystem

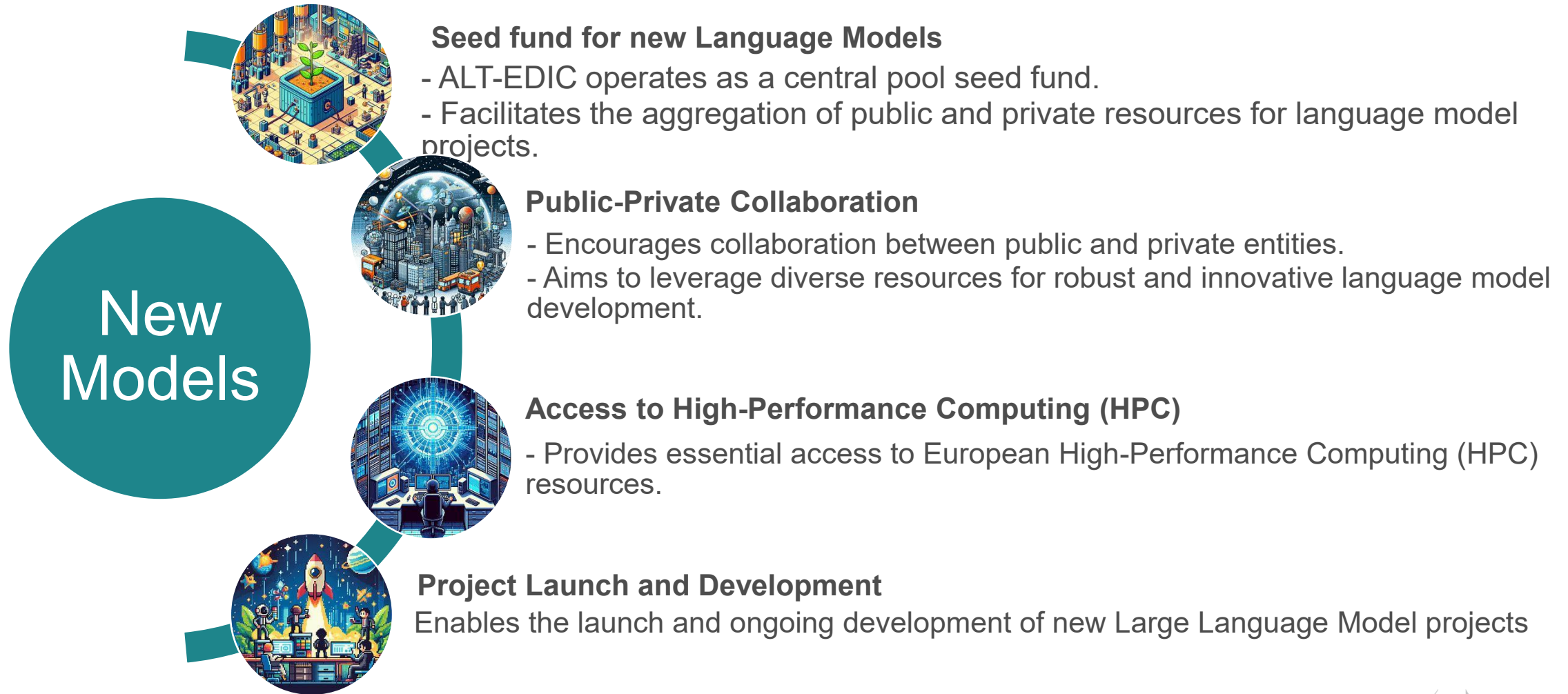
Action Plan - Data



Action Plan – Existing Models



Action Plan – New Models



Action Plan – 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

One Last Thought: Why it matters?

- Join forces with 18+ Member States to co-develop models and services
- Support Local Economy in adopting multilingual AI solutions
- Ensure Maltese is represented in European AI and language models
- It's not too late—Malta can still join ALT-EDIC as a full member
- Benefit from FSTP (cascade funding) opportunities in ongoing Digital Europe projects

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.

Slide **xx: element concerned**, source: [e.g. Fotolia.com](#); Slide **xx: element concerned**, source: [e.g. iStock.com](#)