



The Digital Europe Programme and the Common European Language Data Space

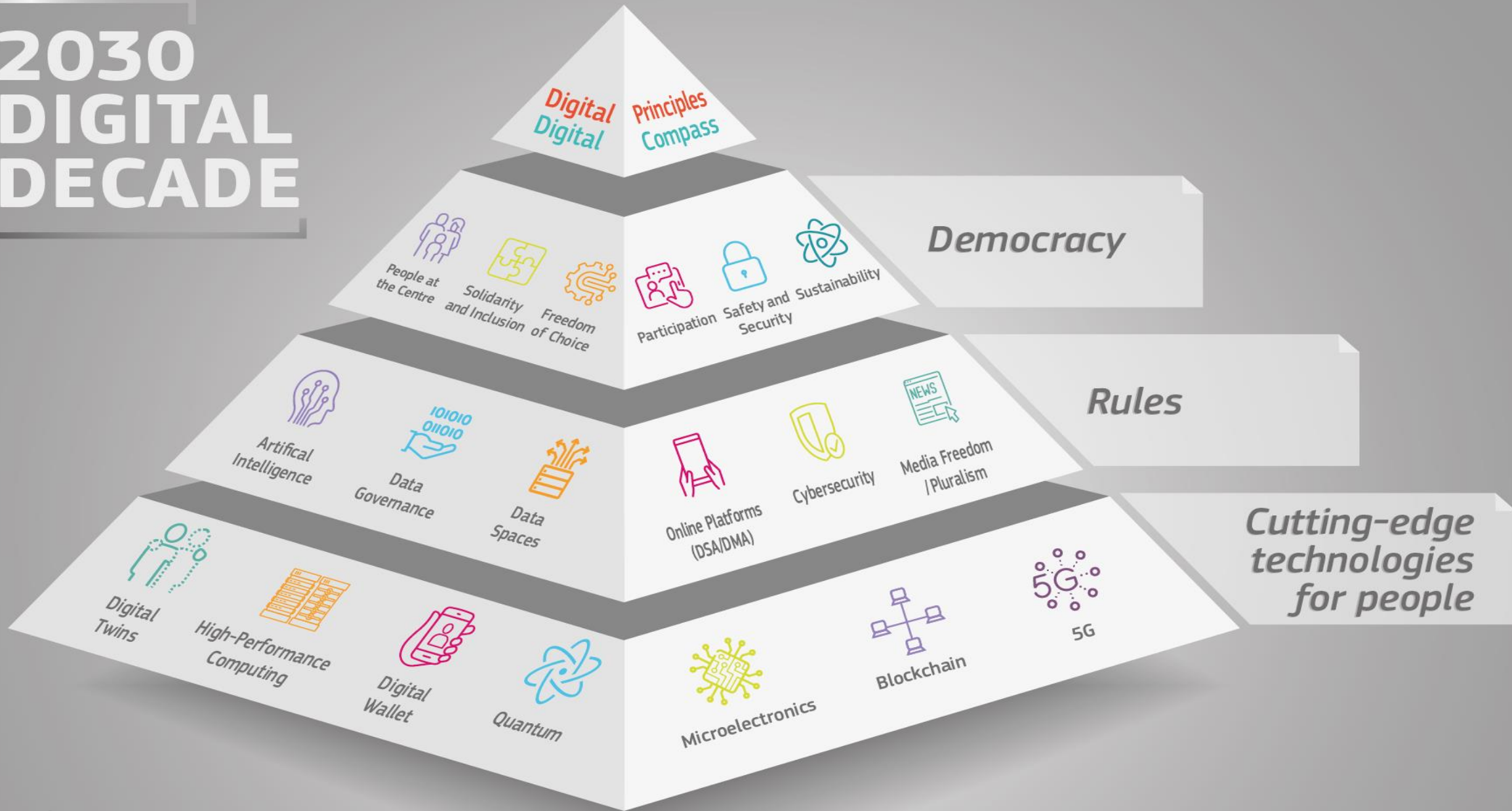
The European Language Data Space (LDS)
Workshop in Luxembourg

*Philippe GELIN & Dhafer LAHBIB
European Commission – DG CONNECT – G3*



Revolution !

2030 DIGITAL DECADE



Landscape

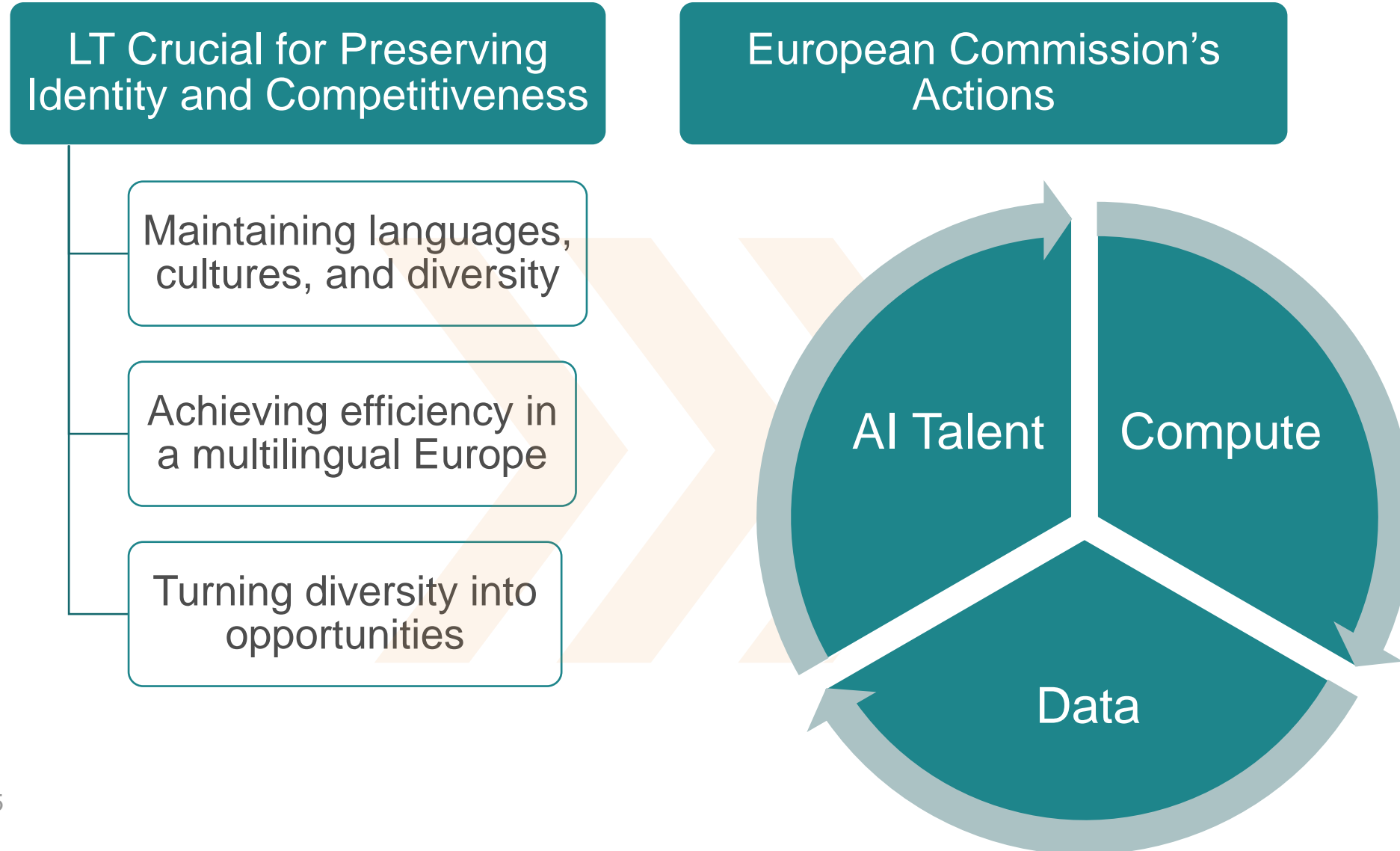
LT Crucial for Preserving Identity and Competitiveness

Maintaining languages, cultures, and diversity

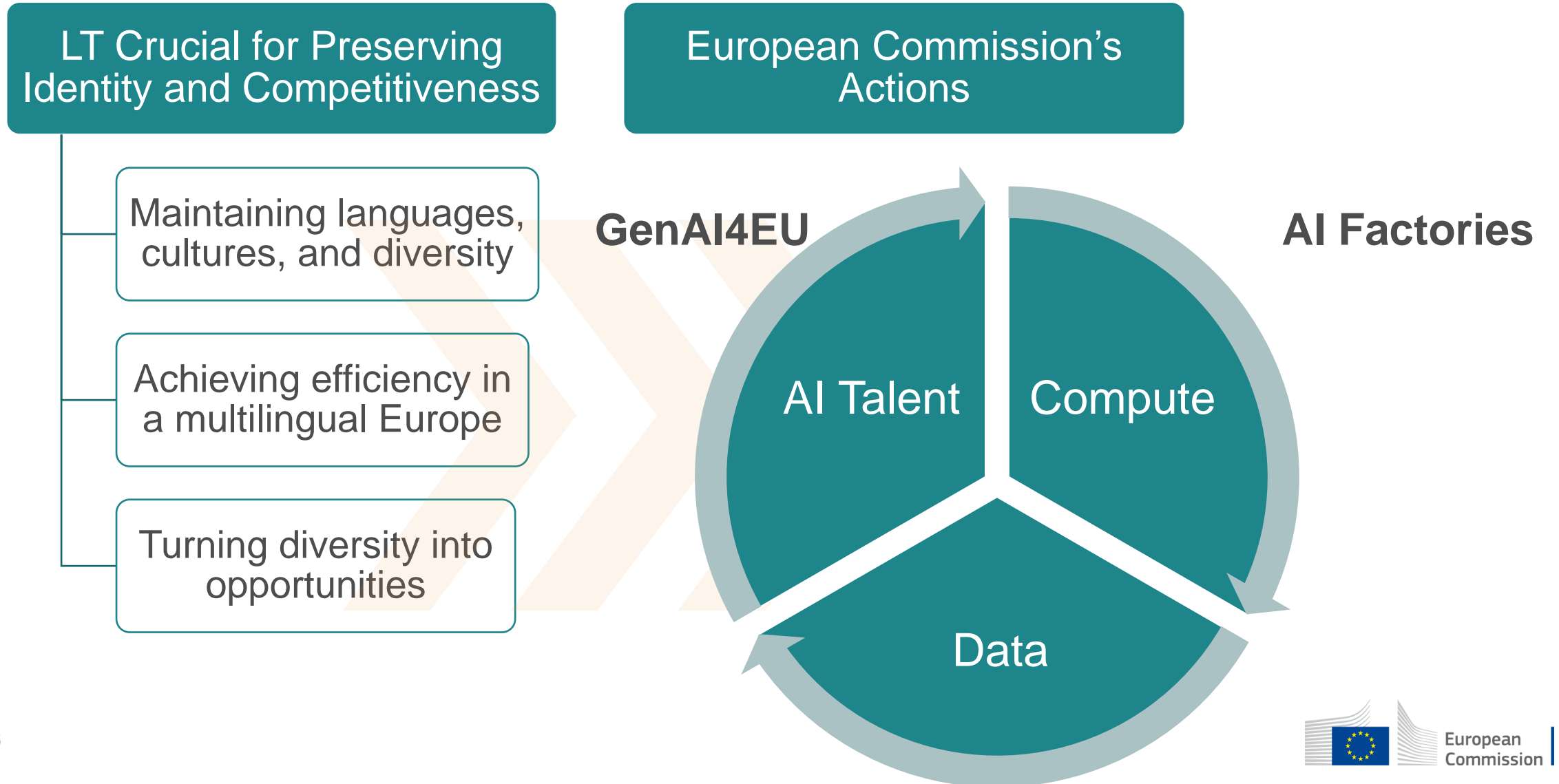
Achieving efficiency in a multilingual Europe

Turning diversity into opportunities

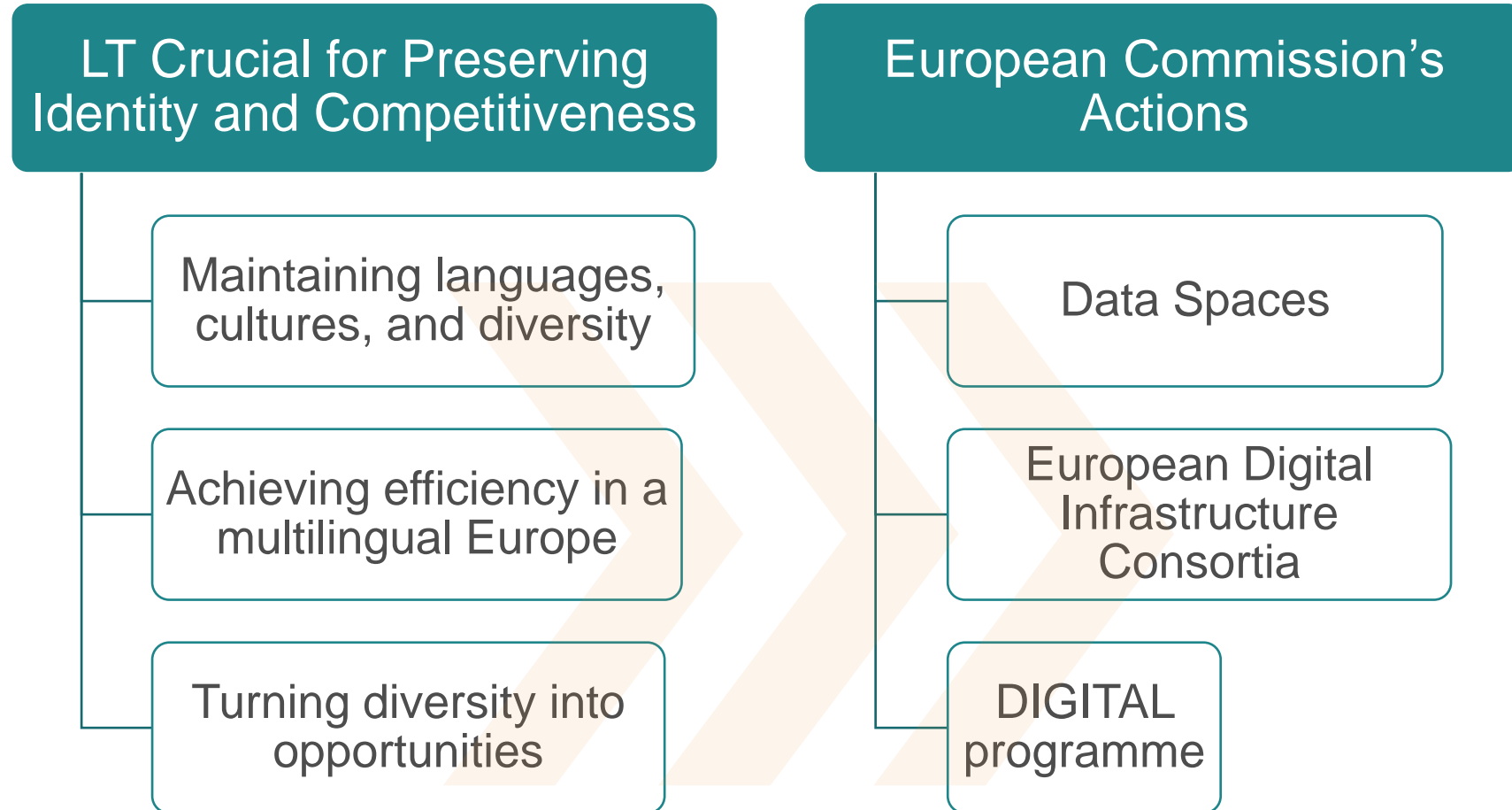
Landscape



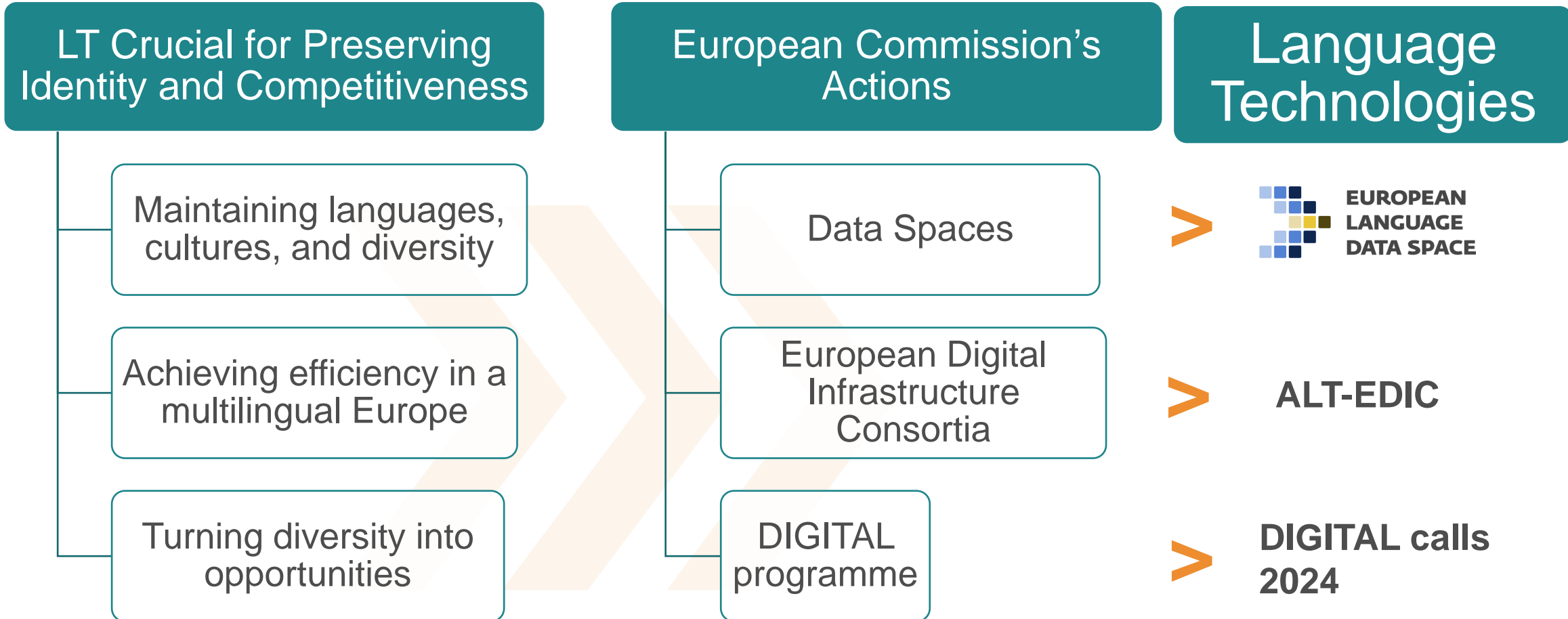
Landscape



Landscape



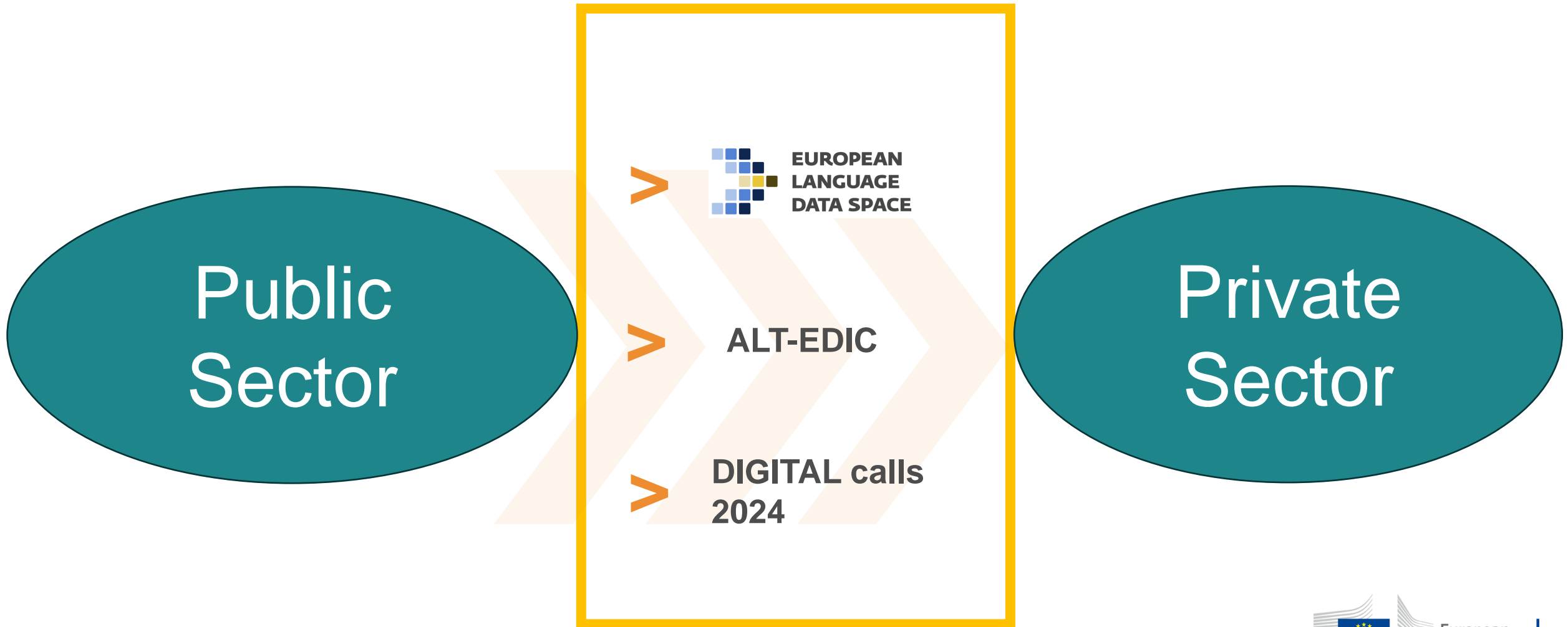
Landscape



Landscape



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

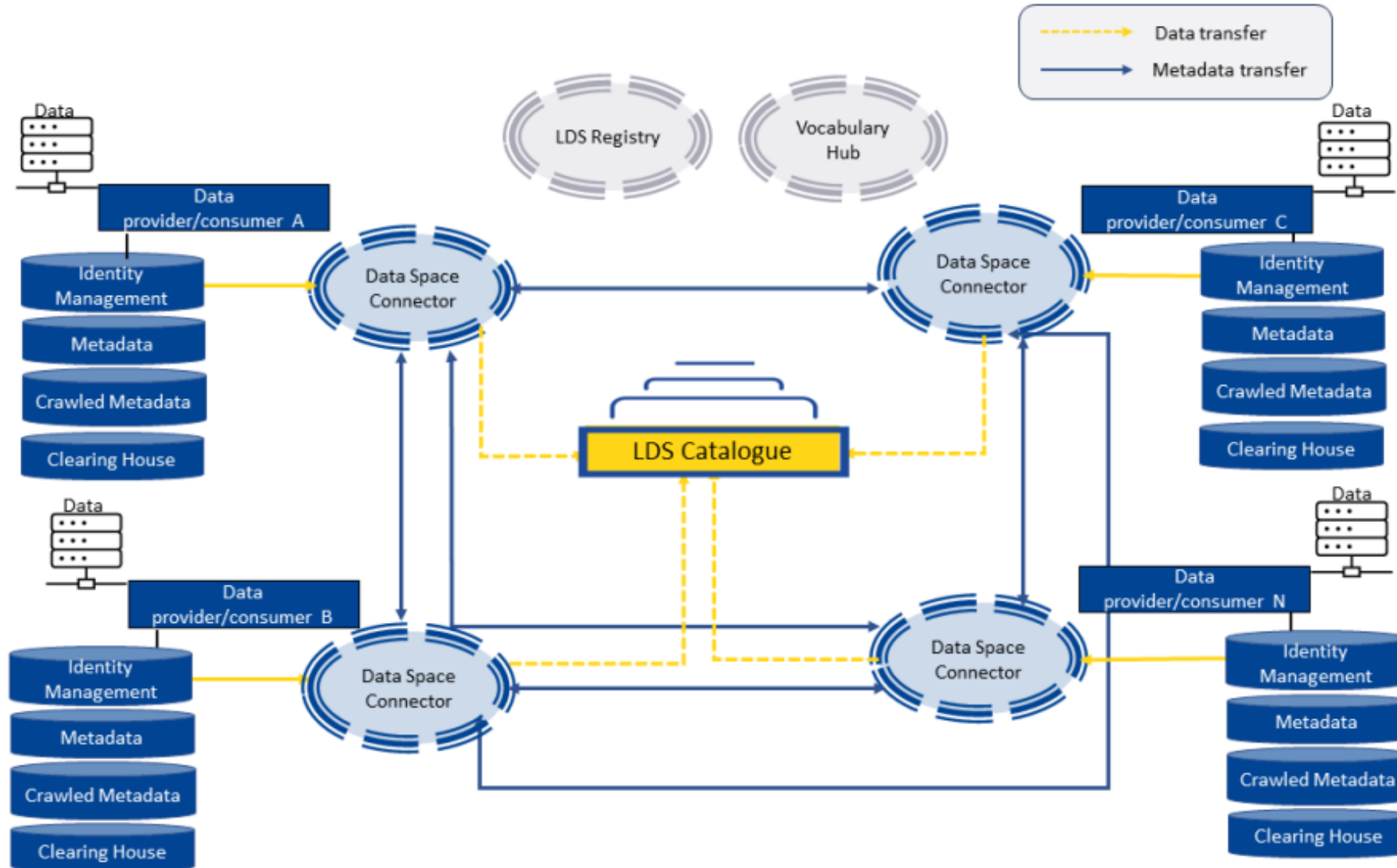
- Overcoming Legal Barriers (IPR, GDPR, etc.)

When

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

First version expected in **Q4 2024**

Language Data Space – Architecture



Alliance for Language Technology EDIC (ALT-EDIC)

ALT-EDIC – Alliance for Language Technologies EDIC

What	DDPP Multi-Country Project in the area of “European Common Data Infrastructure and Services” Build a European ecosystem around Language Data and Technologies
Why	Federate data and resources from EU and Member States to boost Data and AI competitiveness Promote technology and service deployment for European leadership and strategic autonomy
Who	Coordinated by: France (Ministry of Finance and Digitalization & Ministry of Culture) Members: BG, CZ, DK, ES, FR, GR, HR, HU, IE, IT, LT, LU, LV, NL, PL, SI – In process: FI Observers: AT, BE, EE, MT, PT, RO, SK – Interested: CY, DE, SE
How	~25 Mio € Member States contribution ~25 Mio € EU funding (DIGITAL Programme)
When	7 February 2024: Set up by Commission Implementing Decision (EU) 2024/458 Q3 2024: Fully operational

ALT-EDIC and Luxembourg

- Luxembourg is an active member in ALT-EDIC
- Luxembourg's unique multilingual landscape
- Relevant Initiatives
 - MeluXina Supercomputer
 - Digital Luxembourg Initiative, the Luxembourgish Open data platform
 - Relevant interdisciplinary projects at LIST and the University of Luxembourg

Digital Europe Programme

Work Programme 2024

DIGITAL Work Programme 2024 (3 calls - **Closed**)

DIGITAL-2024-AI-06-LANGUAGE-01

Alliance for Language Technologies:

Data and Finetuning

€ 20 millions (Simple Grant)

- Support the Collection of Language Data
- Support Fine-tuning of Large Language Models

DIGITAL-2024-AI-06-LANGUAGE-02

Alliance for Language Technologies:

Ecosystem

€ 4 millions (CSA)

Support the Coordination of a European Language Technologies Ecosystem.

DIGITAL-2024-AI-06-FINETUNE

Open-source European foundation model

€ 25 millions (SME Grant)

Develop and make available one open-source large language foundation model as an infrastructure designed to be largely used by public or private users.

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.

Slide xx: element concerned, source: e.g. Fotolia.com; Slide xx: element concerned, source: e.g. iStock.com