



European  
Commission

# *The European Language Data Space*

**Philippe Gelin**

**Head of Sector Multilingualism,  
Unit "Accessibility, Multilingualism & Safer Internet"**

**Directorate Data**

**Directorate-General for Communications Networks, Content and Technology  
European Commission**

MY EUROPE, MY LANGUAGE.

...MADE IN EUROPE

# Digital Programme

The [Digital Europe Programme \(DIGITAL\)](#) provides funding for projects in five crucial areas:



Programme VISION  
Digital Decade and digital targets for 2030.

Programme FOCUS  
Building the **strategic digital capacities** of the Union and facilitating the **wide deployment of digital technologies**.

Programme GOAL  
Bridging the **gap between digital technology research and market deployment**, while supporting the **EU twin objectives** of a green transition and digital transformation.

DSs and the LDS are under SO2 'Cloud, Data and AI', whose aim is to drive the *digital transformation* of the *public and private sectors*.

# 'Origins' of the Common European Data Spaces

## Common European Data Spaces

They are meant to harness the **value of data** for the benefit of the European economy and society.

are the centrepiece of the



Objectives



## European Strategy for Data

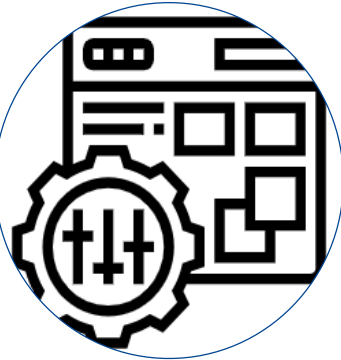
In February 2020, it announced the Commission's vision to empower businesses, public administrations and individuals to make better decisions **through the use of data**.

The creation of EU-wide common, interoperable DSs in strategic sectors is to **overcome existing legal and technical barriers** to data sharing and, as such, **unleash the potential of data-driven innovation**.

→ More data shared in a privacy-compliant, trustworthy, and secure way.

→ **A genuine single market for data.**

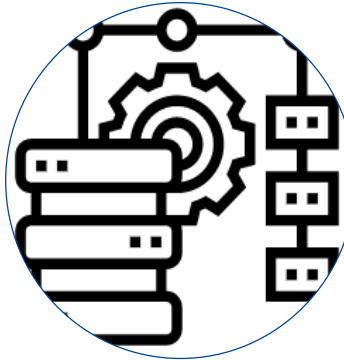
# Design Principles of the DS



## Data control

For data holders:

- Full control over data at all times;
- Option to change data access rights over time.



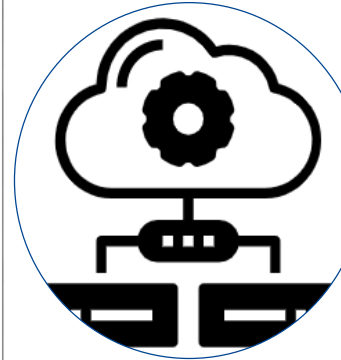
## Governance Framework

- Fair,
- transparent,
- proportionate,
- non-discriminatory access to & sharing/use of data.



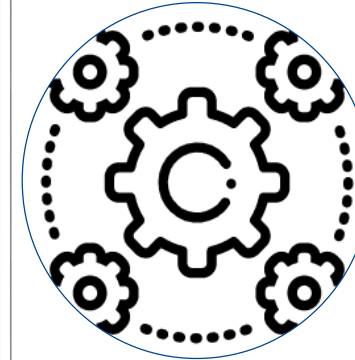
## Respect of EU Rules & Values

- Personal data protection
- Consumer protection legislation
- Competition law



## Technical Data Infrastructure

- Building blocks
- Efficient & coordinated DSs construction



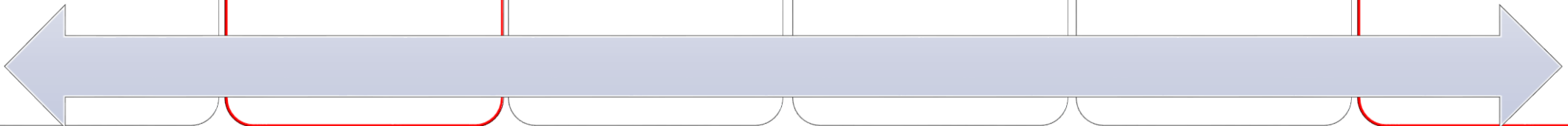
## Interconnection & Interoperability

- Progressive interlinking of DSs
- Aims:*
- Fully integrated ecosystem
  - Cross-exploitation of data

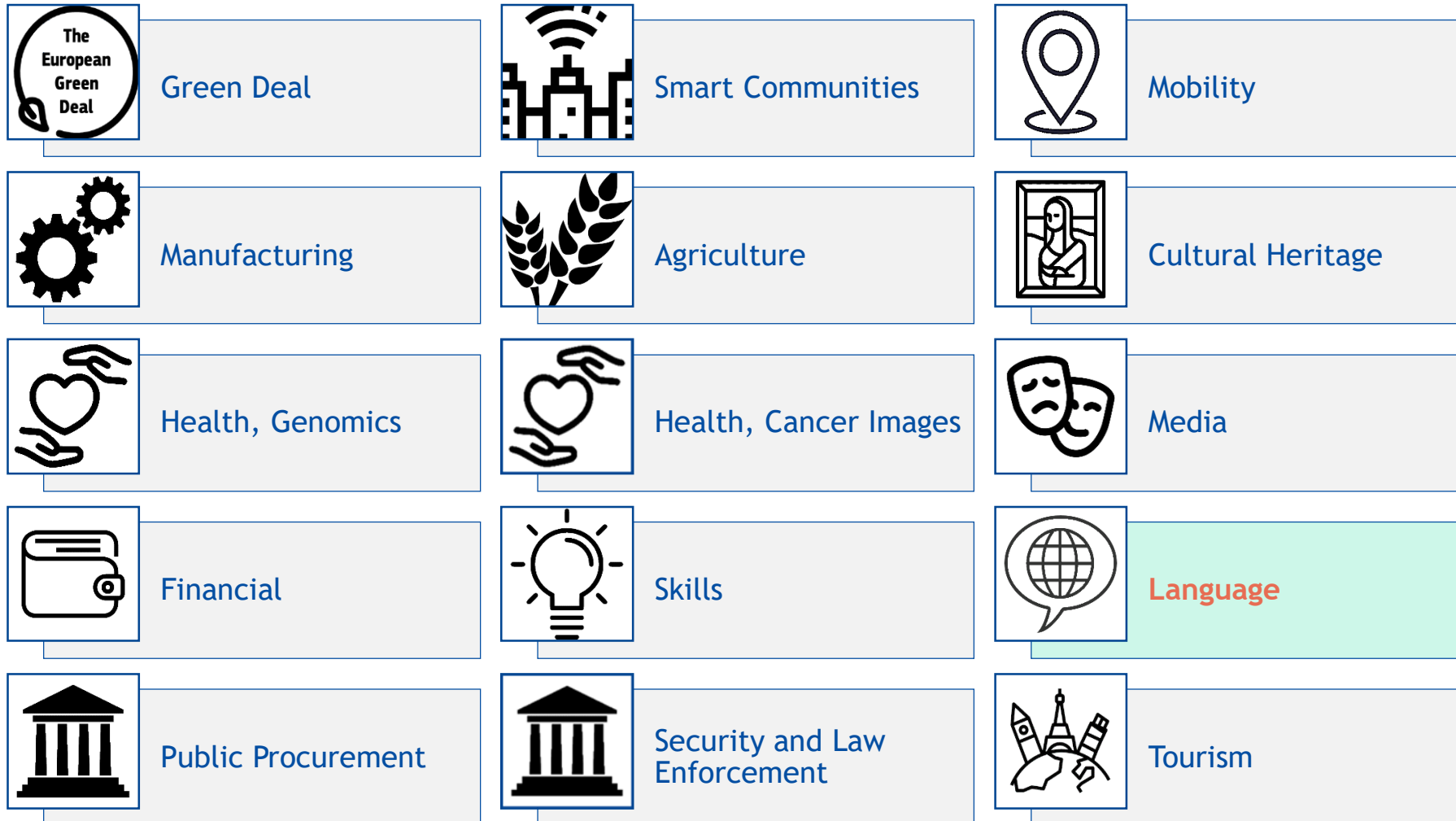


## Openness

- For all actors, organisations or individuals;
- Aims:*
- Fostering competition
  - Avoiding lock-ins



# DIGITAL WP 21-22 Data Spaces at a glance

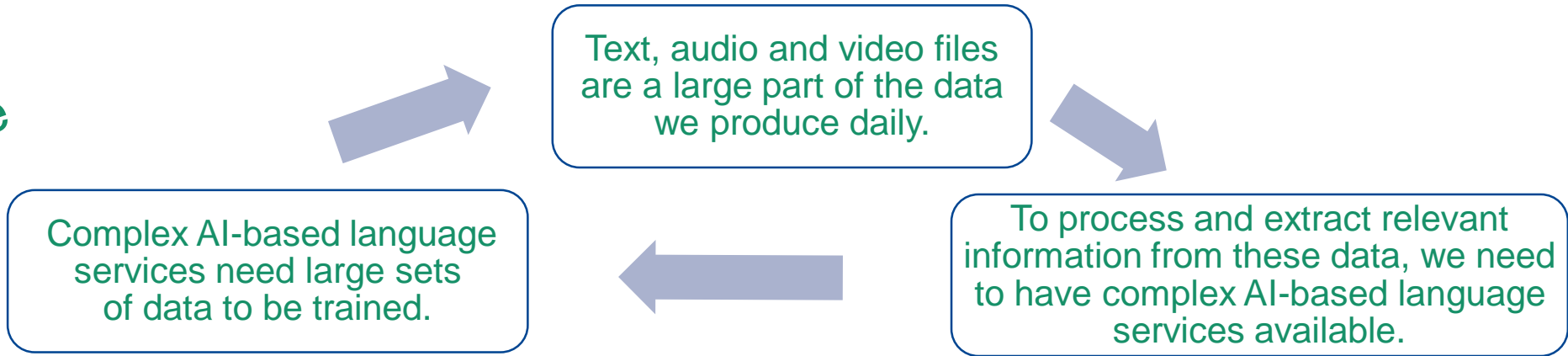


# Language Data Space... What is it?

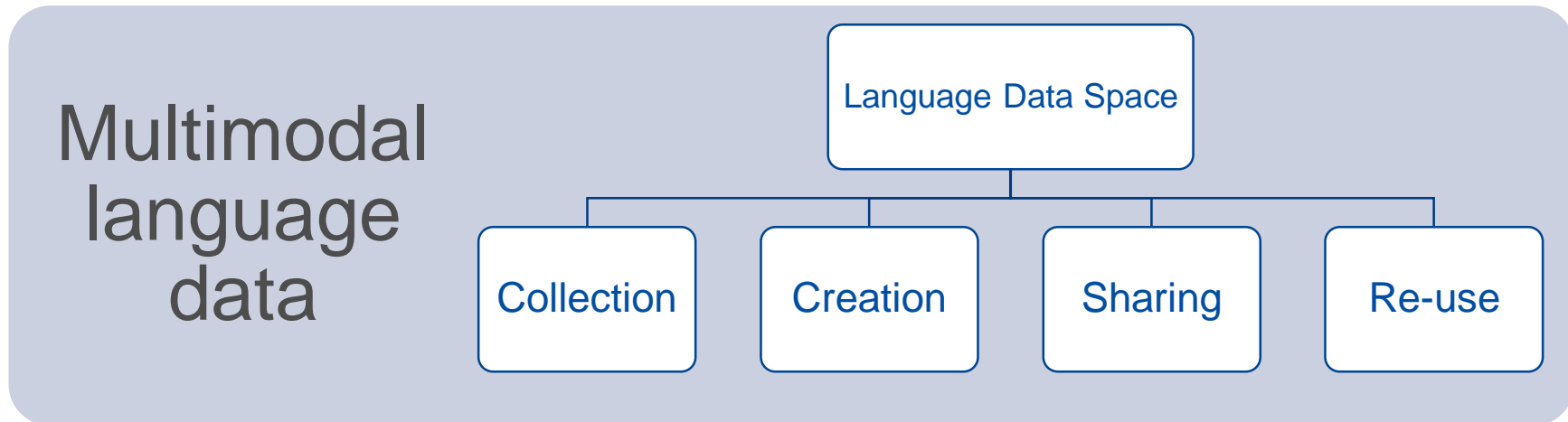
**Goal**

## Increase Data Exchange

*Where we are*



*What we offer*



# European DS Ecosystem – DIGITAL SO2

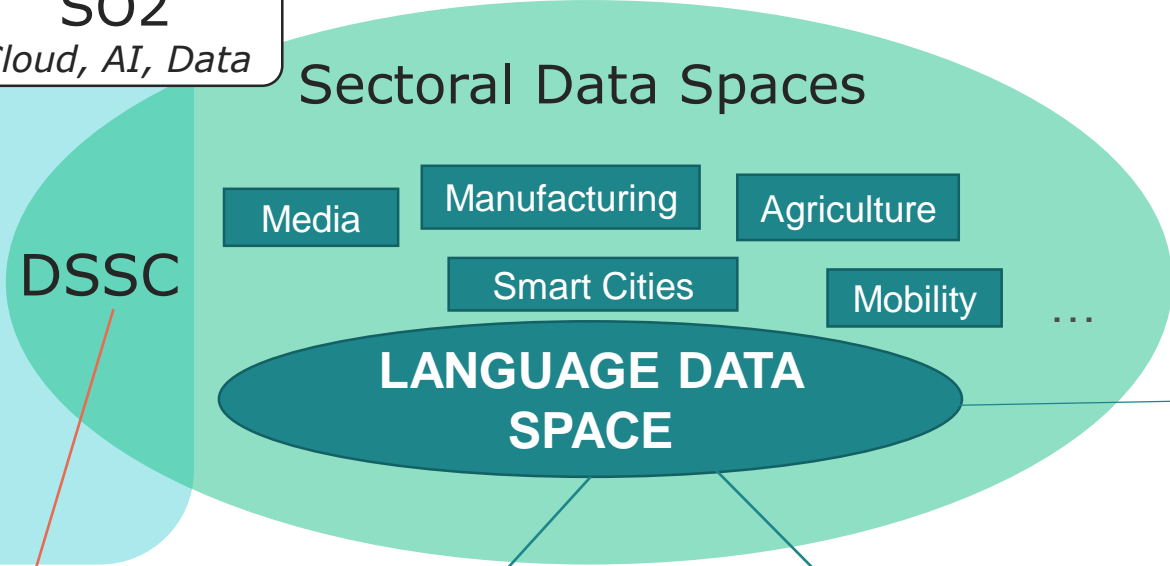
**EDBI - European Data Innovation Board**  
supporting the EC in issuing guidelines on how to facilitate the development of DSs as well as identifying the relevant standards and interoperability requirements for cross-sector data sharing.

**SO1**  
*HPC*

**European Data Spaces Technical Framework**  
consisting of  
Cloud-to-Edge Infrastructures and Services  
including

- Smart Middleware Solutions
- Marketplace

**SO2**  
*Cloud, AI, Data*



**SO4**  
*Advanced Skills*

**SO3**  
*Cyber-security*

**DSSC – Data Spaces Support Centre**  
support all relevant actions on sectoral DSs and their architectures and data infrastructure requirements available for the DSs.

**AI on-demand platform**  
connecting providers and consumers/users of AI resources.

**TEFs – Testing and Experimentation Facilities**  
for experimenting with AI-based solutions in real-world environments.

**SO5**  
*Best use of Tech*  
**Common Services Platform**

# LDS Work Strands

1

**Establishing** an institutional Centre of Excellence for Language Technologies (**CELT**)

*Actions:*

Governance layer

→ to framework the collection, creation, sharing and re-use of **language data & models**, incl. legal aspects.

2

**Deploying** the Language Data Space

Operational layer

→ to implement the **architecture**;  
→ to promote the **exchange of data**;  
→ to support the deployment of **large multimodal language models** and a wide range of **AI language technology services** through the AI on-demand platform.

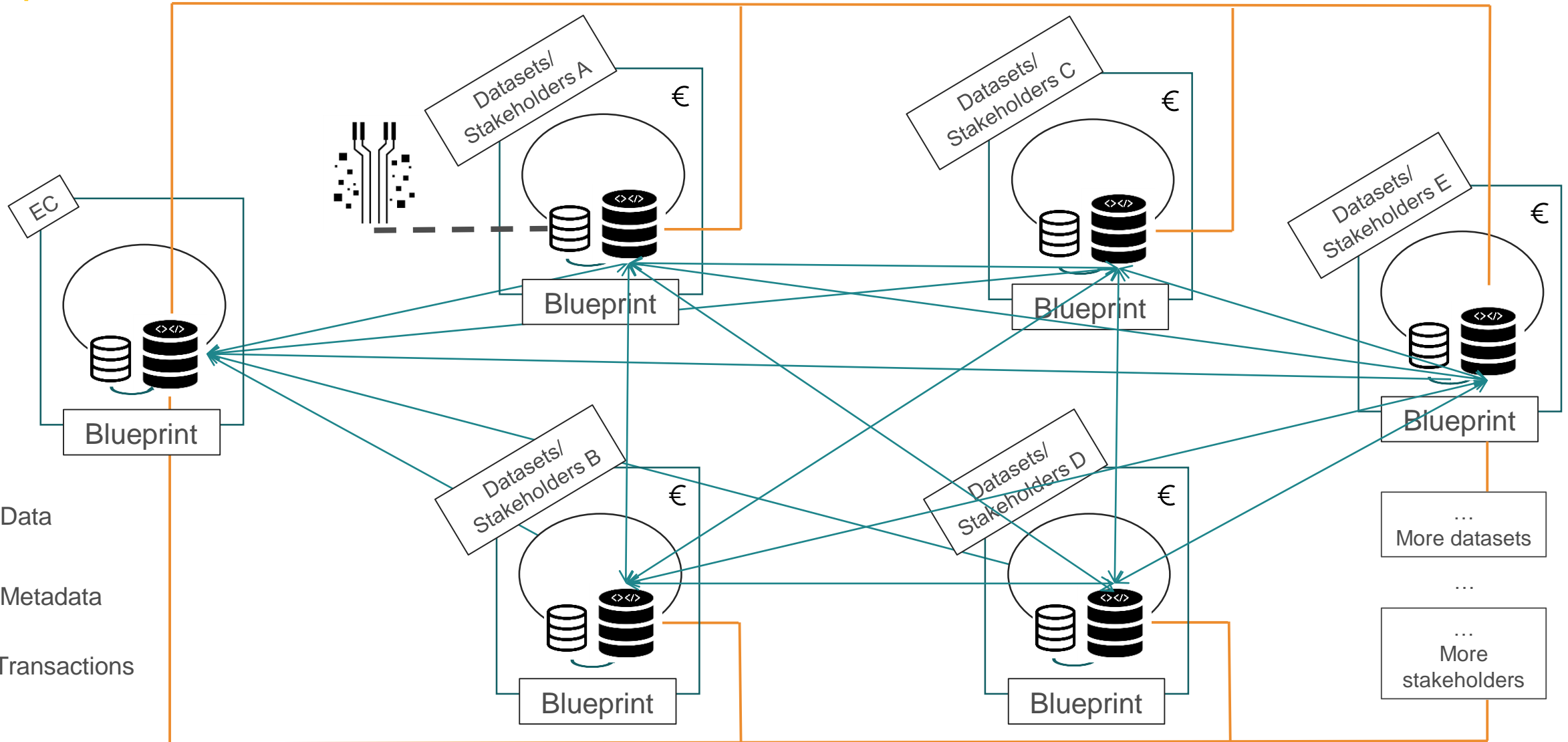


# Establishment of the CELT & the CELT+

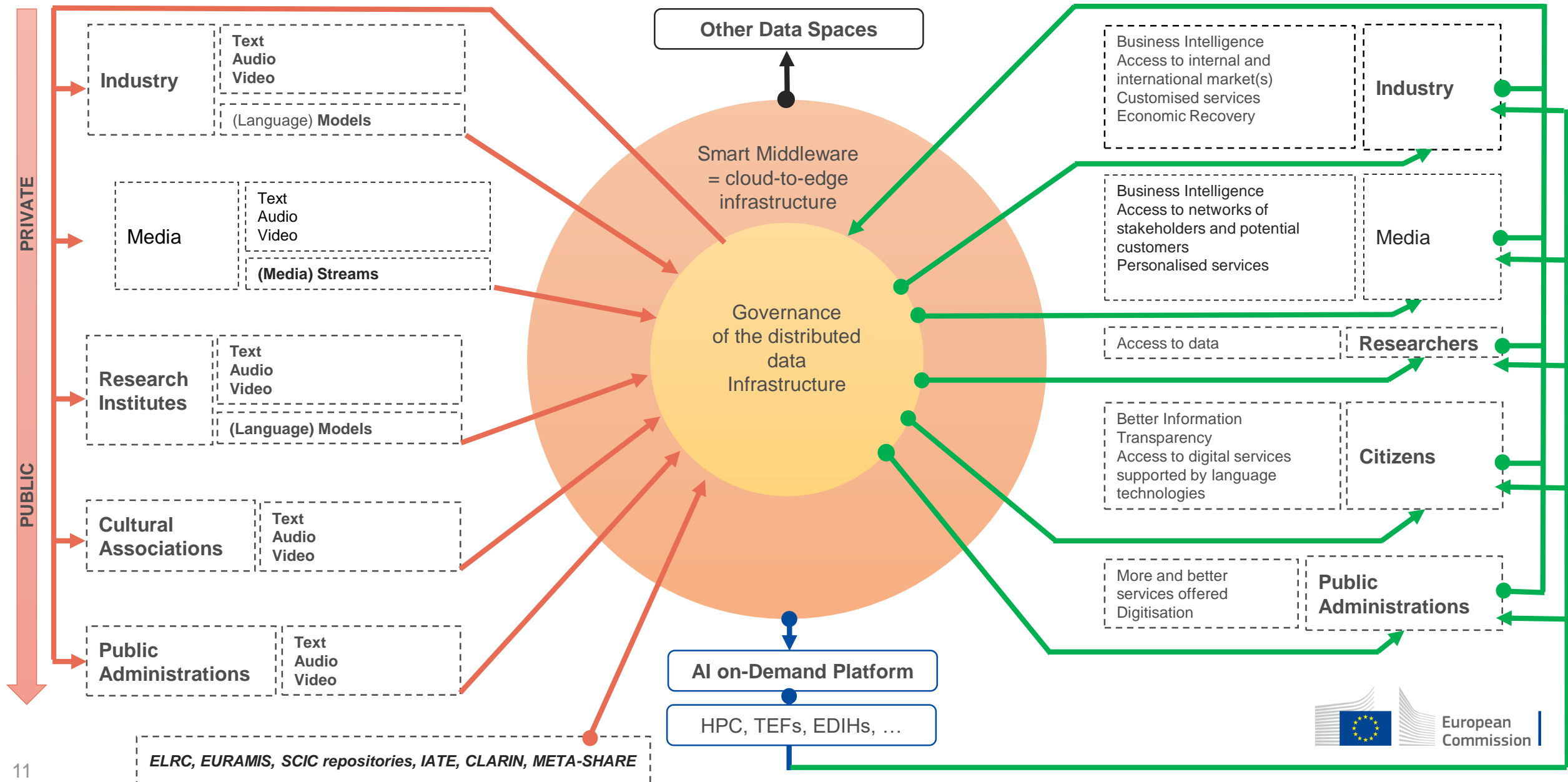
Strategic Level	<p><b>CELT</b> Coordinating across the MSs the creation and collection of multimodal language data and models.</p>	<p>Networking</p> <p>Identifying and bringing together existing European <b>stakeholders/initiatives</b> and <b>language resources</b></p> <ul style="list-style-type: none"> <li>• Datasets and data streams</li> <li>• <u>Large language models</u></li> </ul> <p><i>ELRC, EURAMIS, SCIC repositories, IATE, CLARIN, META-SHARE, etc.</i></p>
		<p>Governance</p> <p>Developing a <b>Multi-Stakeholder Data and Services Governance Scheme</b></p>
Operational level	<p><b>CELT+</b> Driven by Industry</p>	<p>Planning</p> <p>Elaborating a <b>Blueprint for the Language Data Ecosystem</b> based on existing EU/national legislation, incl. a detailed <b>roadmap</b> on how to deploy the language data space</p>

EC Role: Facilitator

# LDS Distributed Architecture

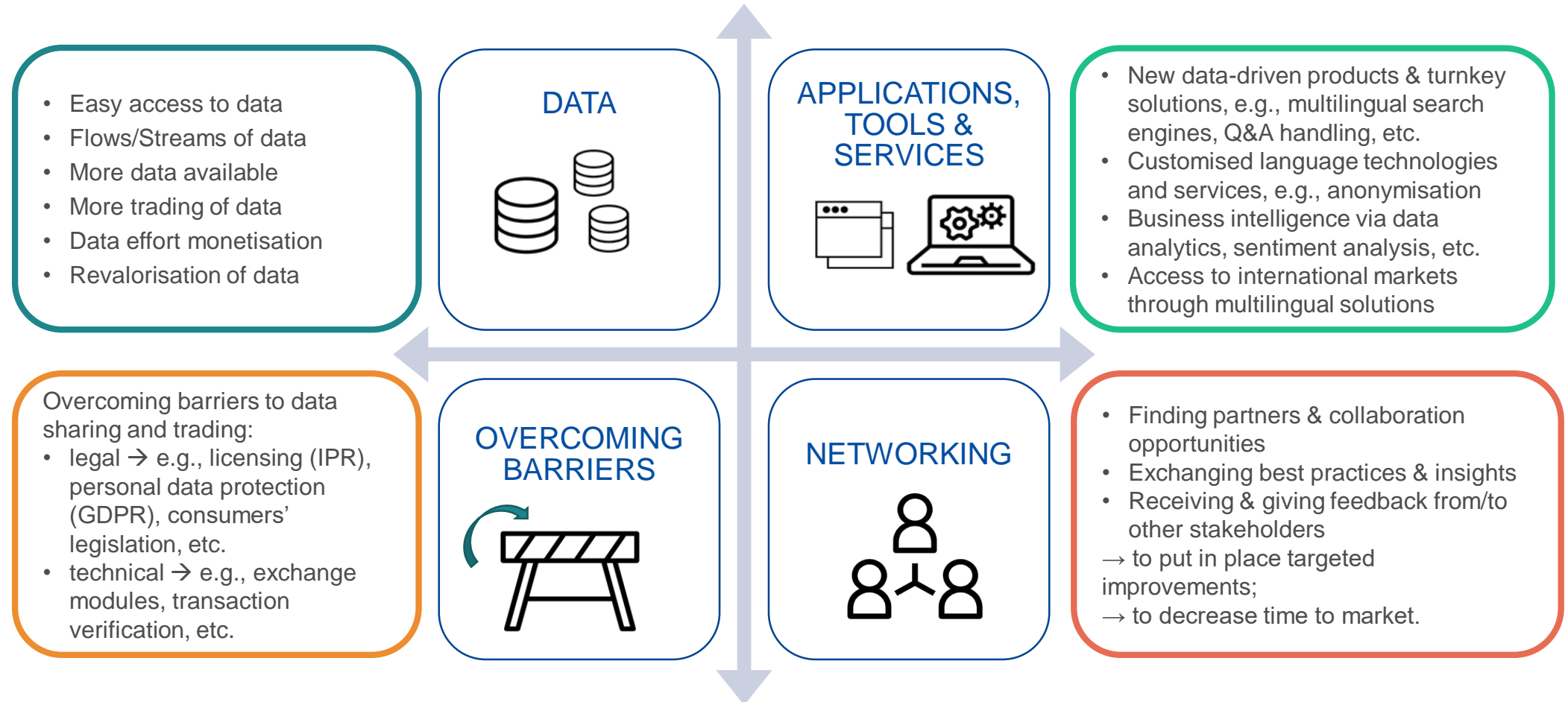


# Stakeholders' Landscape

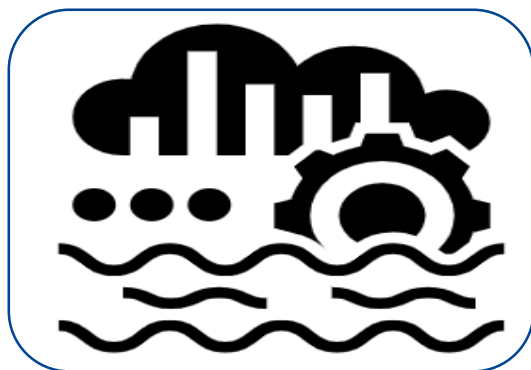


# Benefits of the LDS

The LDS has been thought to be a **platform and a marketplace** *where the request for language resources and services matches the offer.*



# Key messages



**Data sharing  
against  
remuneration**

E.g., multilingual  
and/or multimodal  
data



**Developing,  
selling and  
purchasing  
language-based  
applications,  
tools and models**

E.g., automatic  
speech  
transcription



**Developing,  
selling and  
purchasing  
language-based  
services**

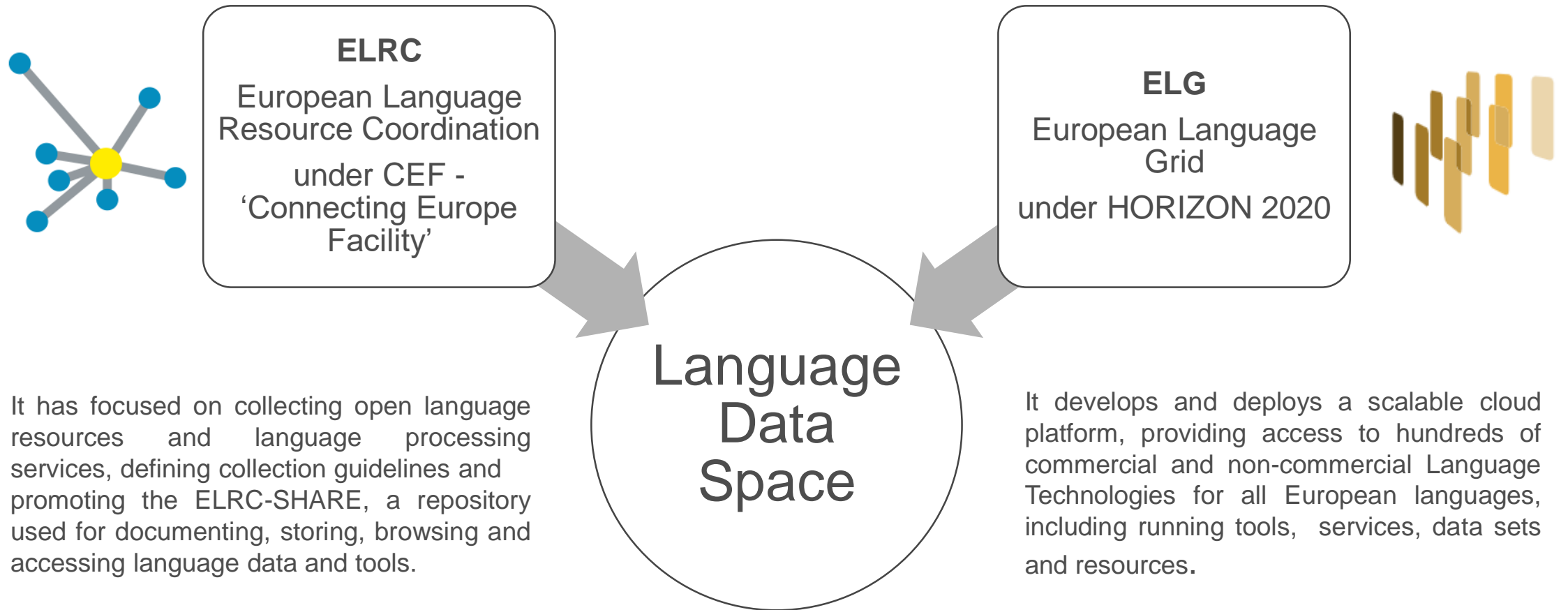
E.g., automatic text  
summarisation



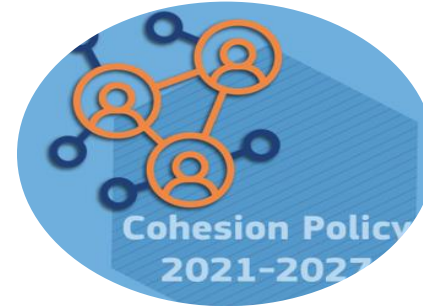
**Performing data  
analytics and  
further exploiting  
the market for  
more reach and  
visibility**

E.g., historical data

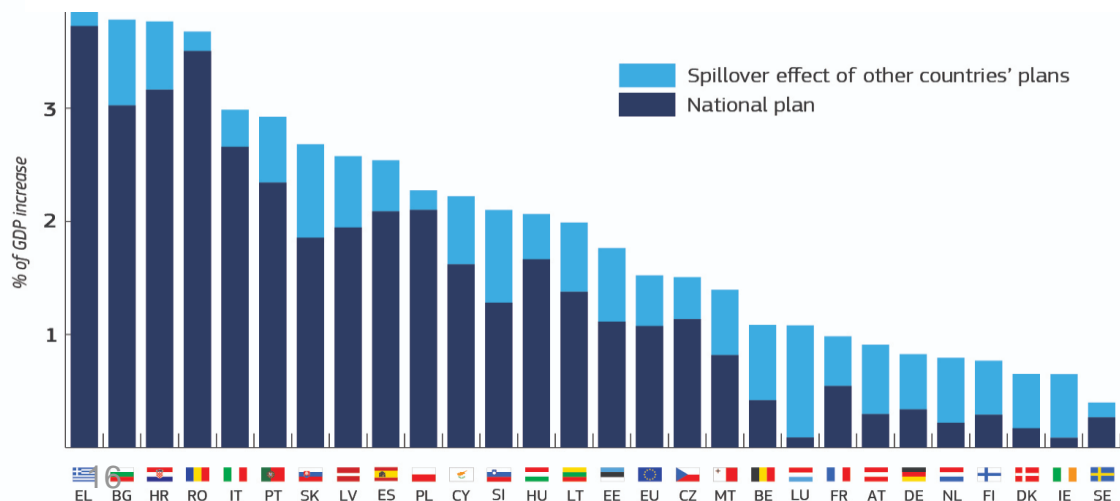
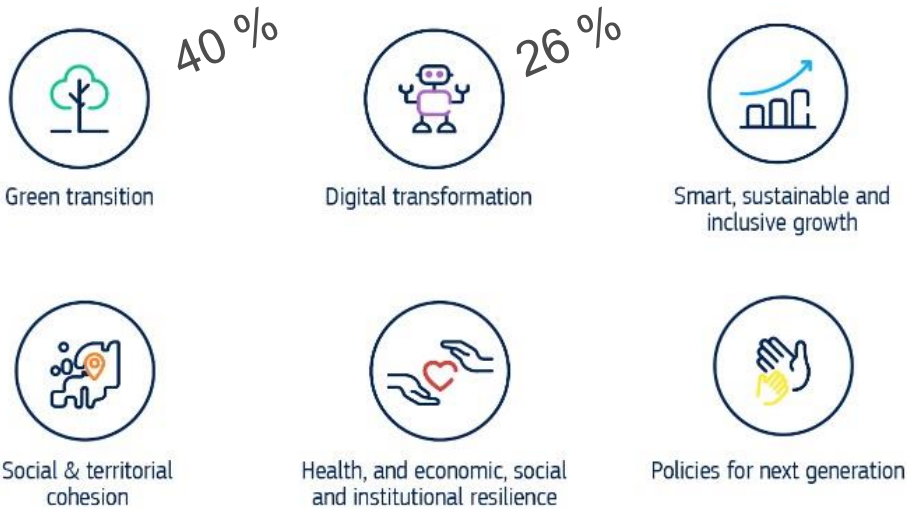
# Previous Projects Relevant for the LDS



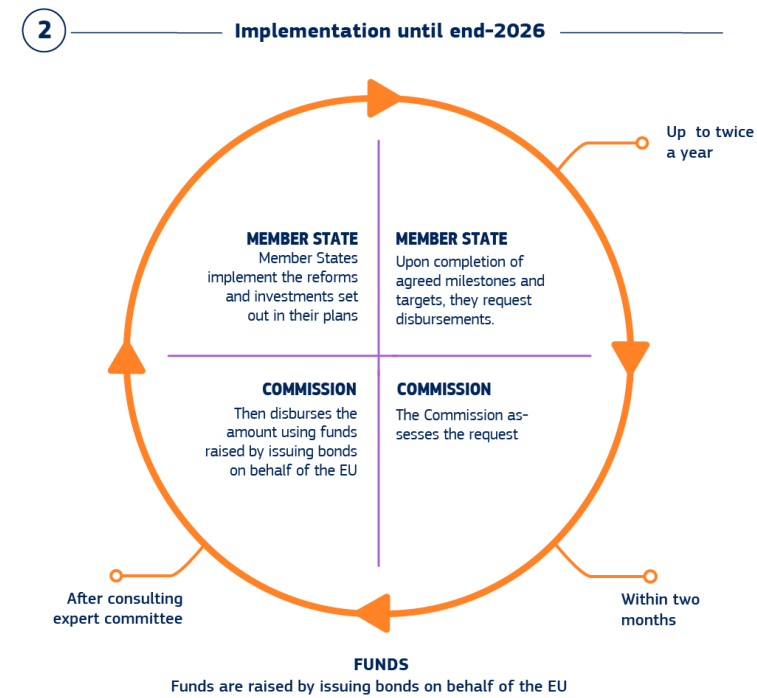
# Our actions: Legislative, Coordination & Funding



# Next Generation EU & Recovery and Resilience Facility



<b>Recovery and Resilience Facility (RRF)</b>	<b>€723.8 billion</b>
<i>of which, loans</i>	<b>€385.8 billion</b>
<i>of which, grants</i>	<b>€338.0 billion</b>





# What is a European Digital Infrastructure Consortium ?

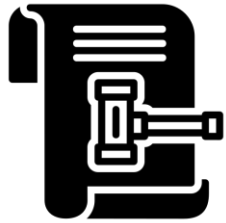
*Pending Legal  
Implementation*

- EDIC is a new legal framework to support Multi Country Project
- They have first been introduced in the Path to the Digital Decade Policy Programme (DDPP) proposal.
- The concept has been inspired by the success of European Research Infrastructure Consortia (ERICs).
- Advantages:
  - ✓ Facilitating coordination of funding between MS;
  - ✓ Speedy establishment and flexible implementation;
  - ✓ Ensuring common standards and interoperability;
  - ✓ Sustainability.

# What are the main features of an EDIC?

Pending Legal Implementation

## LEGAL STATUS



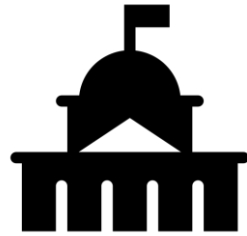
- Legal entity
- Set up by EC decision upon MS request

## PARTICIPATION



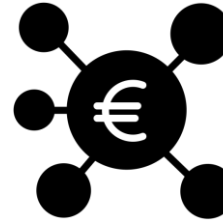
- 3 MS needed for set-up
- Members = voting rights
- Observers = no voting rights
- Open for new MS to join
- Open for entities other than MS

## GOVERNANCE



- Flexible structure, defined in the EDIC Statutes
- Assembly of Members with a Director

## FINANCING



- Commitments of EDIC MS (financial or not)
- Funding from EU Programmes (CEF, DEP, HORIZON Europe, etc.)
- Contributions from private entities

## FINANCIAL ADVANTAGE



- VAT exemption (if EDIC recognised as international body)

## LIABILITY



- MS liability limited to their contribution/ (general)
- No EU liability for EDIC's debts

# What initiatives could a Language EDIC benefit from?

Pending Legal Implementation

## From EC Side

Ongoing **European Language Resource Coordination (ELRC)**, finishing end of 2022

Upcoming **LDS** procurement, starting in early 2023

**EU research initiatives** (e.g., H2020 European Language Grid or ELG, ending in June 2022)  
Future HORIZON calls

## From MS Side (*non limitative*)

Germany: OpenGPT-X project

France: BigScience project and 'Jean Zay' supercomputer

Spain: PERTE – The New Language Economy

Sweden: Autonomous Systems and Software Programme

Slovenia: Development of Slovene in Digital Environment (DSDE)

Latvia: Hugo.lv

...

# What are the next steps (1)?

Pending Legal  
Implementation

## EC side

- **EDIC** significant stages and next steps

- Political agreement on the DDPP Decision reached in *July 2022*;
- Informal Call for Expression of Interest in EDICs: *November/December 2022*;
- Expected DDPP entry into force: *January 2023* (tbc).

= Initiation of a Language EDIC as early as 2023, but sync with LDS necessary.

- EC guidance on implementing best-suited mechanism for the MCP in question.



European  
Commission

# *The European Language Data Space*

**Philippe Gelin**

**Head of Sector Multilingualism,  
Unit "Accessibility, Multilingualism & Safer Internet"**

**Directorate Data**

**Directorate-General for Communications Networks, Content and Technology  
European Commission**

MY EUROPE, MY LANGUAGE.

...MADE IN EUROPE