



*Quo vadis?*  
**The European  
Language Data Space**

**Philippe Gelin**

**6<sup>th</sup> ELRC Conference**

**31 March 2022, 10:40 – 11:00**

**Head of Sector Multilingualism**  
Data Directorate  
Directorate-General for Communications Networks,  
Content and Technology  
European Commission

# Content

## 1. The Digital European Programme (DIGITAL)

- Vision, Focus and Goal
- Indicative Budget Breakdown

## 2. The Common European Data Spaces (DSs)

- Concept and Design Principles
- The European Data Spaces Ecosystem and the Sectoral Data Spaces

## 3. The Language Data Space (LDS)

- First Work Strand
- Second Work Strand
- Visualisation
- European Digital Infrastructure Consortium (EDIC)

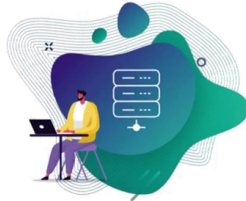
# DIGITAL: *Vision, Focus and Goal*

S03



CYBERSECURITY

S01



HIGH PERFORMANCE COMPUTING

S04



ADVANCED DIGITAL SKILLS



DIGITAL  
EUROPE  
PROGRAMME

S05



BEST USE OF DIGITAL TECHNOLOGIES

**S02**



CLOUD, DATA AND AI

## **VISION**

Digital Decade and digital targets for 2030

## **FOCUS**

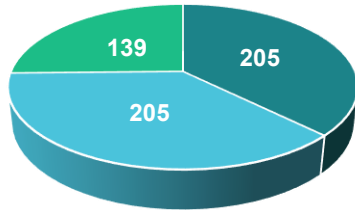
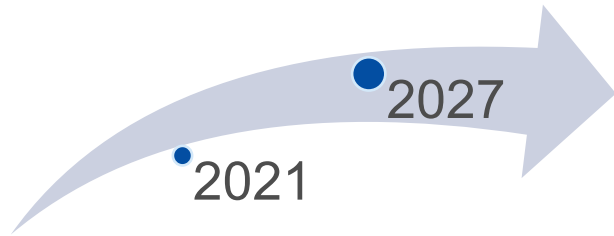
Building the strategic digital capacities of the Union and facilitating the wide deployment of digital technologies

## **GOAL**

Bridging the gap between digital technology research and market deployment

**Digital Europe Programme (DIGITAL)**

# DIGITAL: *Indicative Budget Breakdown*



- Cloud-to-edge
- Data Space
- TEF & AI on Demand Platform

## DIGITAL in MFF 2021-2027

Total financing <b>€7.6 Bio</b>	Total financing for <i>Cloud, Data and AI</i> : <b>€2.1 Bio</b>
------------------------------------	--

[Funding and Tenders](#)

## DIGITAL WP 2021-2022

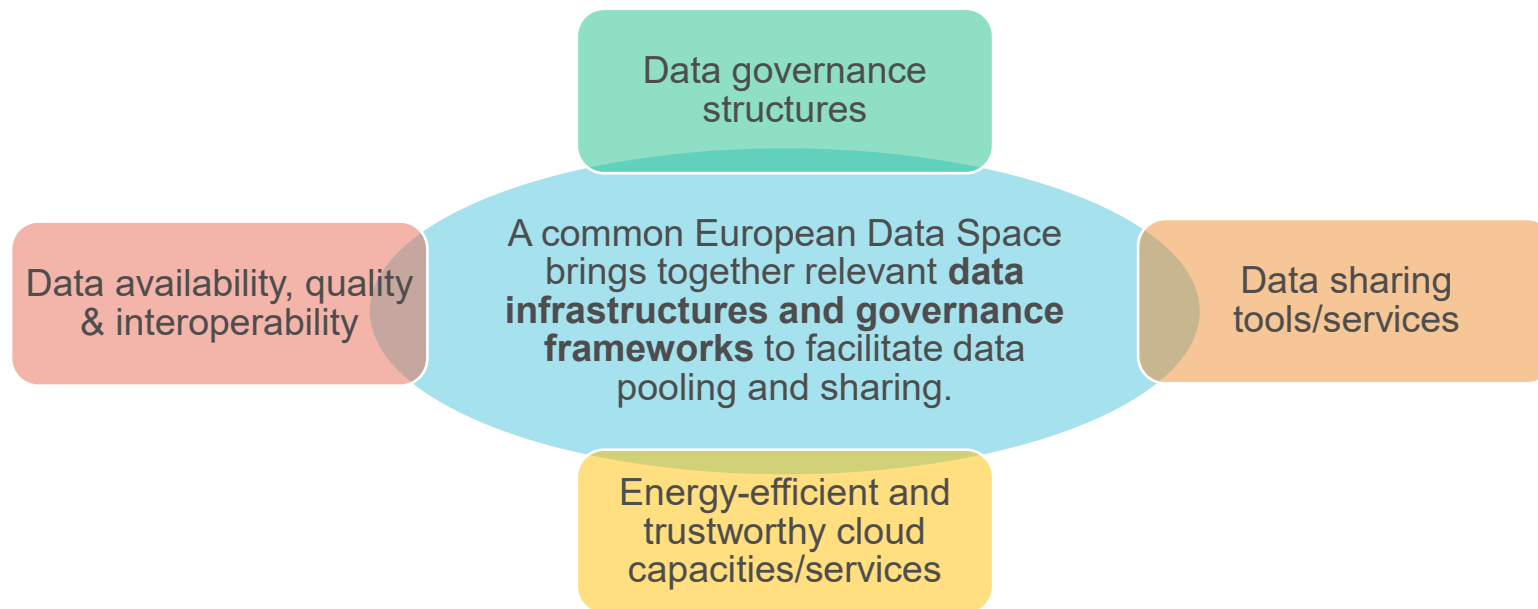
Total financing <b>€1.4 Bio</b>	Cloud-to-Edge: <b>€205 Mio</b> <b>Data Spaces: €205 Mio</b> TEFs & AI on-demand Platform: <b>€139 Mio</b>  Total financing for <i>Cloud, Data and AI</i> : <b>€549 Mio</b>
------------------------------------	---

[DIGITAL WP 2021-2022](#)

*Acronyms*  
MFF: Multiannual Financial Framework ([link](#))  
WP: Work Programme ([link above](#))  
TEF: Testing and Experimentation Facility ([link](#))



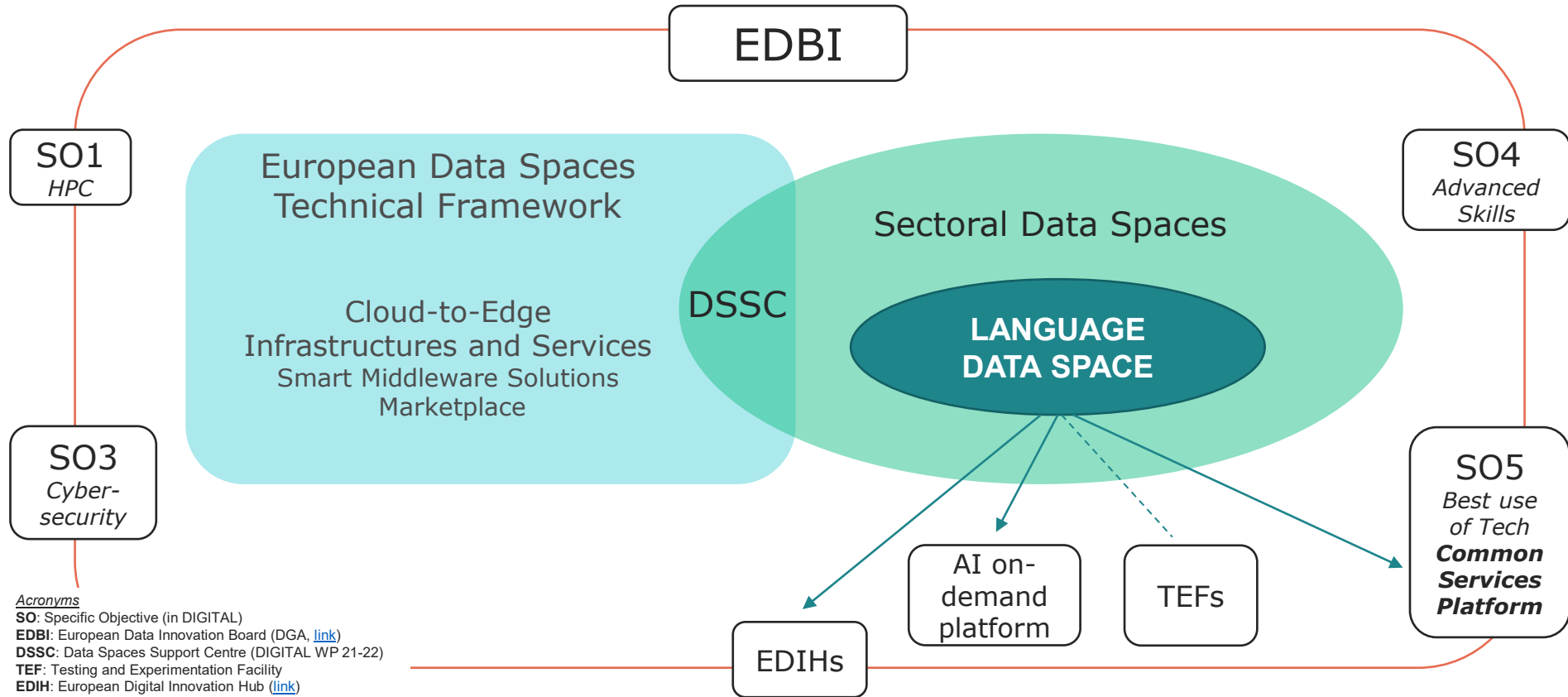
# Common European Data Spaces: *Concept*



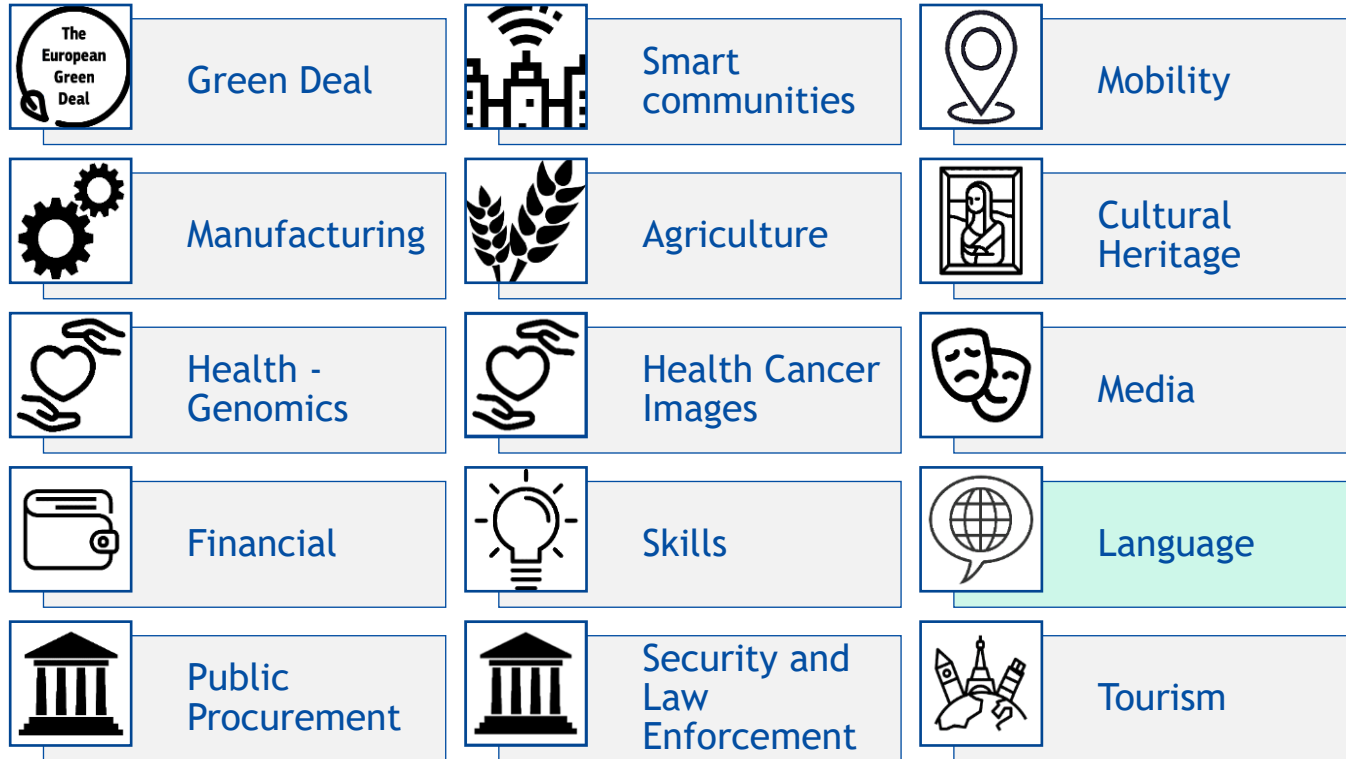
# Common European Data Spaces: *Design Principles*



# European Data Spaces Ecosystem



# Sectoral Data Spaces in DIGITAL WP 21-22





# Language Data Space in DIGITAL WP 21-22

<b>Type of Action</b>	PROCUREMENT
<b>Indicative Budget</b>	EUR 6 million
<b>Indicative Time of the Call Opening</b>	2022
<b>Implementation</b>	European Commission

# Language Data Space: *Work Strands*

## Two Work Strands

1

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

Collection, creation, sharing and reuse of **language data**.

2

Deploying the Language Data Space

This will support the deployment of **large multimodal language models** and a wide range of **AI language technologies services** to be offered through the AI platform.

# Language Data Space: 1<sup>st</sup> Work Strand

## Establishment of the CELT

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

# Language Data Space: 1<sup>st</sup> Work Strand

## Establishment of the CELT

*Strategic Level*

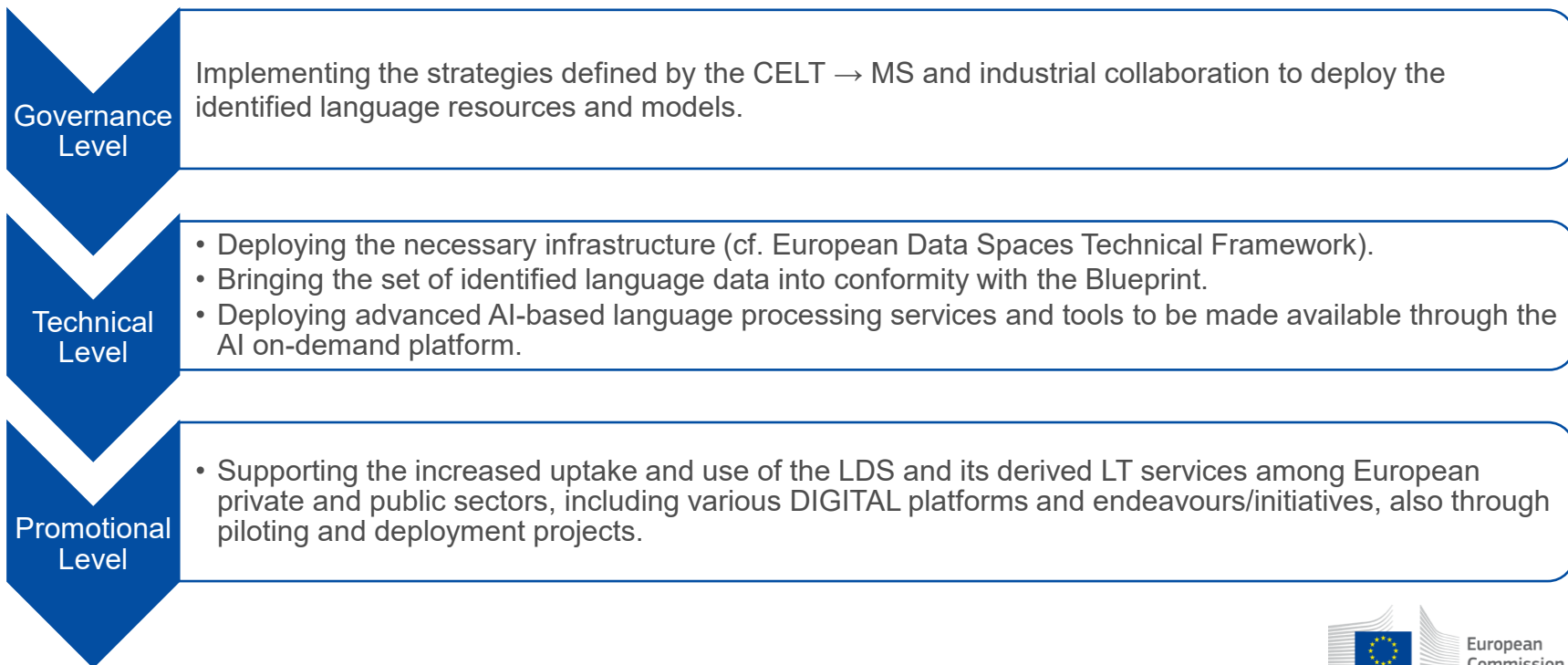
### CELT

Coordinating across the MSs the creation and collection of multimodal language data and models.

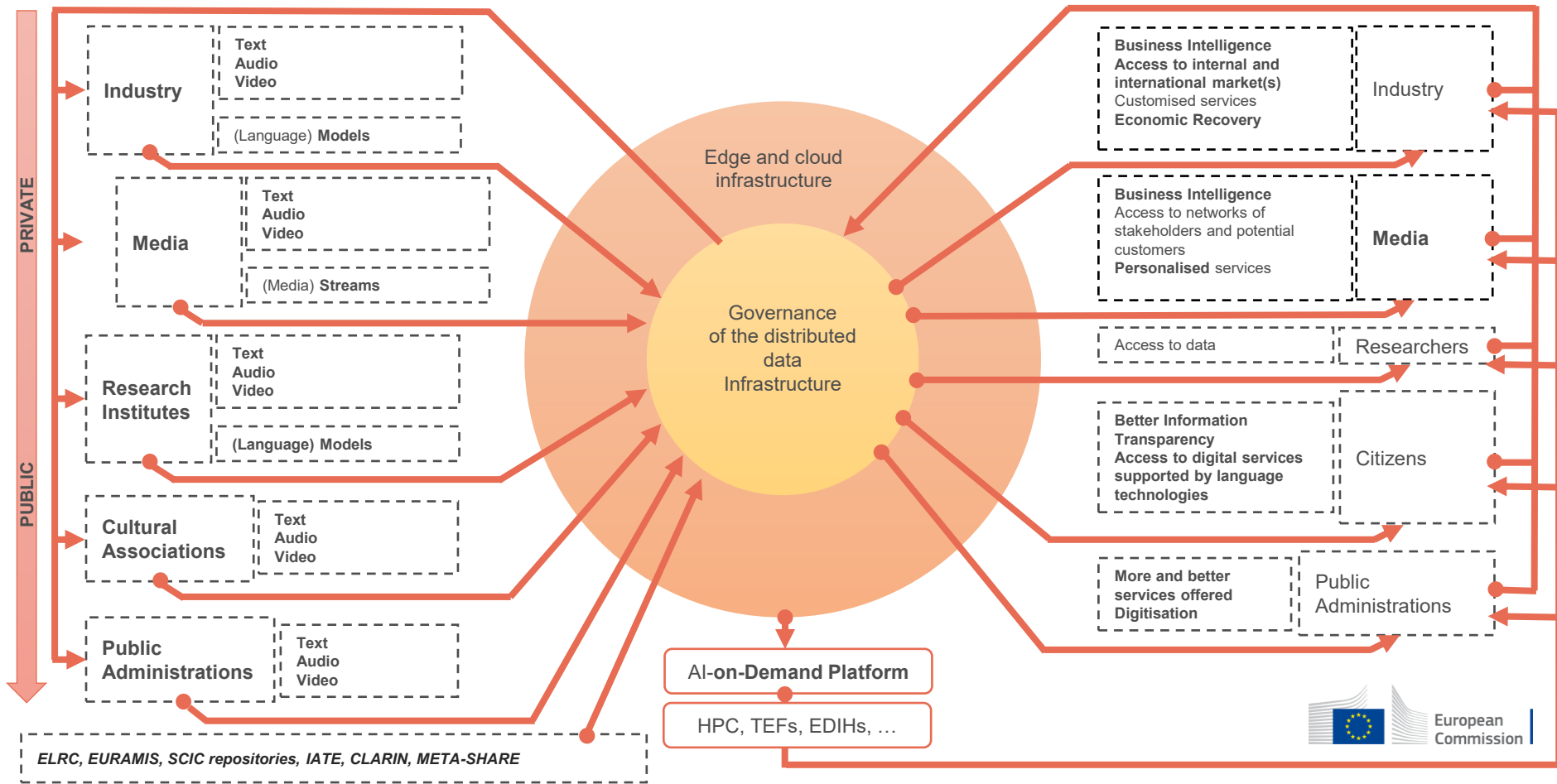
<i>Networking</i>	<i>Governance</i>	<i>Planning</i>
<p>Identifying and bringing together existing European <b>stakeholders/initiatives</b> and <b>language resources</b></p> <ul style="list-style-type: none"><li>• Datasets and data streams</li><li>• <u>Large language models</u></li></ul> <p><i>ELRC, EURAMIS, SCIC repositories, IATE, CLARIN, META-SHARE, etc.</i></p>	<p>Developing a <b>Multi-Stakeholder Data and Services Governance Scheme</b></p>	<p>Elaborating a <b>Blueprint for the Language Data Ecosystem</b> based on existing EU/national legislation</p> <p>Detailed <b>roadmap</b> on how to deploy the language data space</p>

# Language Data Space: 2<sup>nd</sup> Work Strand

## LDS Deployment



# Language Data Space: At a glance



# European Digital Infrastructure Consortium (EDIC)

---

<i>Concept</i>	Influenced by the <b>European Research Infrastructure Consortia</b> (ERICs).
<i>Purpose</i>	Establish and operate a multi-country project (MCP).
<i>Composition</i>	At least three Member States as members.
<i>Financing</i>	Commitments of EDIC members and grants from directly managed EU instruments.

---

Cf. European Research Infrastructure Consortium (ERIC) ([link](#))

Cf. Multi-Country Projects ([link](#)) and DECISION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL establishing the 2030 Policy Programme "Path to the Digital Decade" ([link](#))

# Language Data Space & EDIC

In the ANNEX to the Proposal for a Decision of the European Parliament and of the Council establishing the 2030 Policy Programme “Path to the Digital Decade”, the first area of activity for an EDIC is

*‘European common data infrastructure and services’.*



The LDS would be implemented through an EDIC:

- bringing together European and MS investments;
- coordinating the actual deployment of the LDS.



# Links & contacts



## **PHILIPPE GELIN**

European Commission

DG CONNECT - Communications Networks, Content and  
Technology

Dir G: Data

Unit G3: Accessibility, **Multilingualism** and Safer Internet

[Philippe.Gelin@ec.europa.eu](mailto:Philippe.Gelin@ec.europa.eu)