



e-IRG Workshop under Czech EU
Presidency

Prague, Czech Republic; 13 December 2022

Session 3 - "Data Infrastructures and Data Spaces"

The Language Data Space (LDS) *

Franciska de Jong
Executive Director CLARIN ERIC
www.clarin.eu



- Some slides reused from recent EC presentations by Philippe Gelin;
thanks also to CLARIN colleagues who helped preparing the slides, in particular John Picard

Overview

Context and key drivers

- CLARIN ERIC RI in a nutshell
- Digital Europe Program: sectoral Data Spaces

Actions and next steps




- The Language Data Space (LDS)
- CLARIN's role in the LDS

CLARIN ERIC in a nutshell - after 10 years

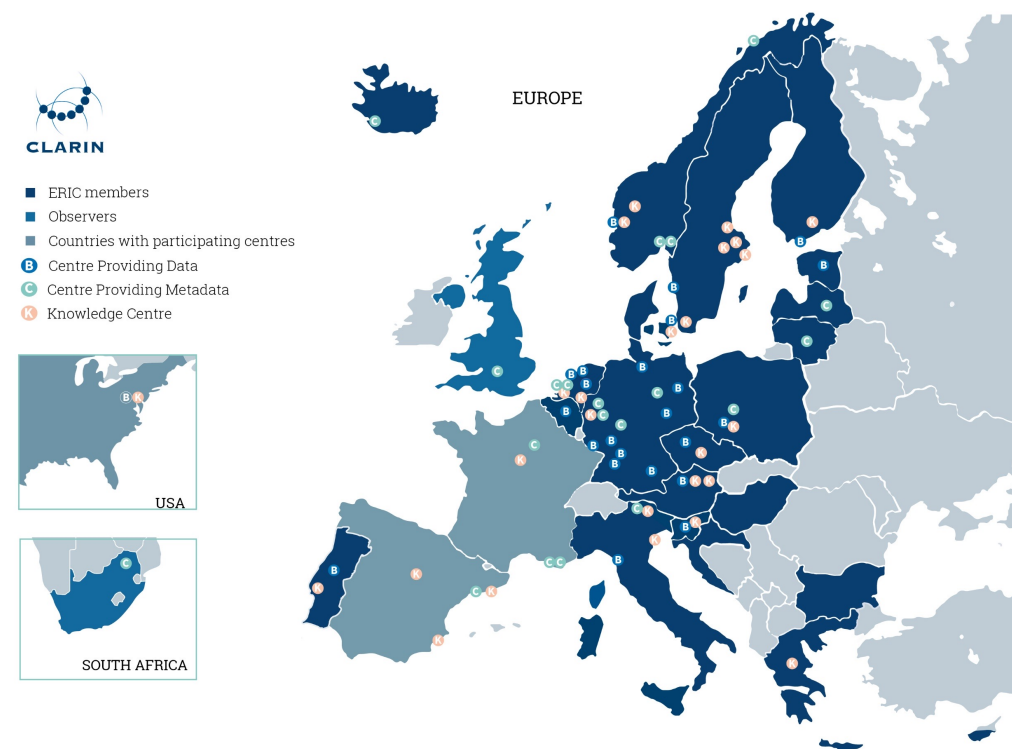
A mature ESFRI Landmark*, a distributed pan-European RI for language resources

- a vibrant, committed community;
- roots in multiple disciplines, such as Digital Humanities, Language Science & Technology, including AI
- 22 member countries and growing
- 2 observers
- 1 linked party

A distributed network of more than 60 data centres

- >20 CTS certified data centres
- strong focus on FAIRness of data
- interoperability of services
 - *federated login* 
 - *central metadata harvesting for easy discovery via the Virtual Language Observatory (VLO)* 
 - *chained services* 
 - *rich catalogue of language resources*

Member of  EOSC



*ESFRI is pushing across the board for strong RI linkages to industry to spur innovation and financial sustainability.

EU 2030 Policy Programme *Path to the Digital Decade*

The “Path to the Digital Decade” aims to ensure that the European Union achieves its objectives and targets towards a digital transformation of society and economy in line with the EU’s values, reinforcing digital leadership and promoting human-centred, inclusive and sustainable digital policies empowering citizens and businesses.

<https://digital-strategy.ec.europa.eu/en/library/proposal-decision-establishing-2030-policy-programme-path-digital-decade>



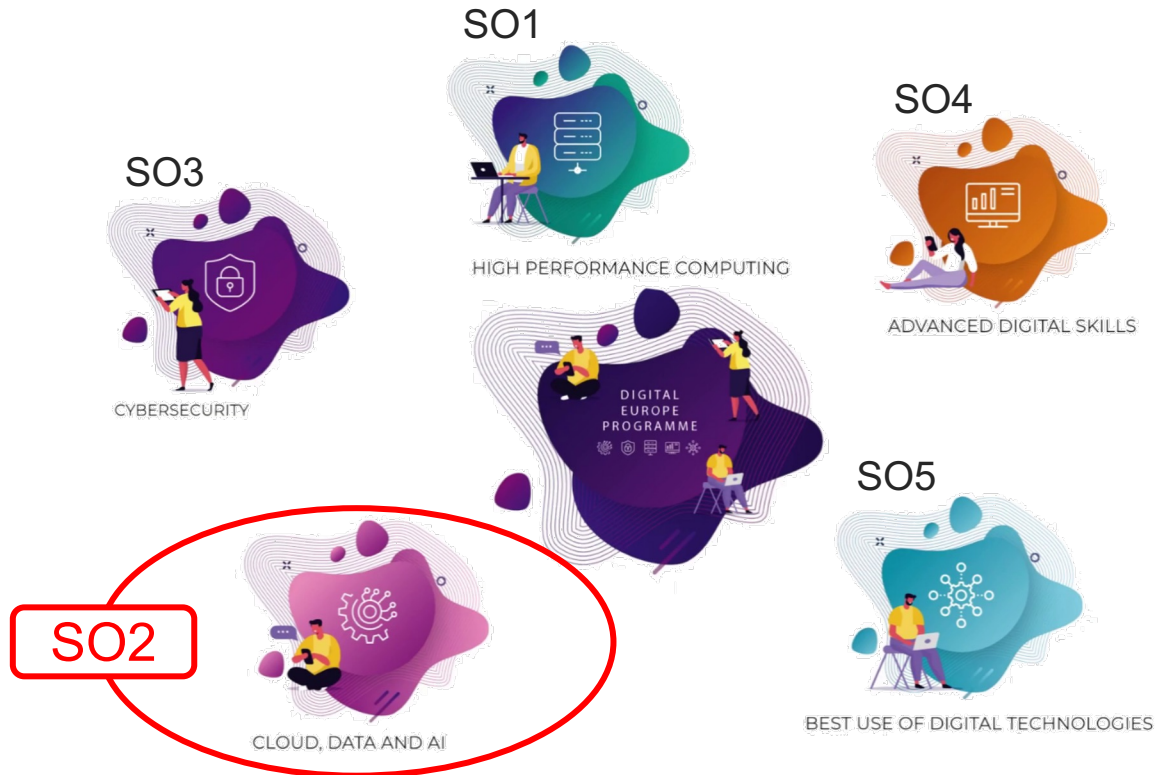
Digital Europe Programme

Sectoral Data Spaces



Position of Language Data Space in DIGITAL

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*.

The European Language Data Space (LDS) - *Why*

The European Commission is preparing to launch a ‘European Language Data Space’, a platform and a marketplace where the request for language resources and services will match the offer. The indicative launch date of the LDS is early 2023.

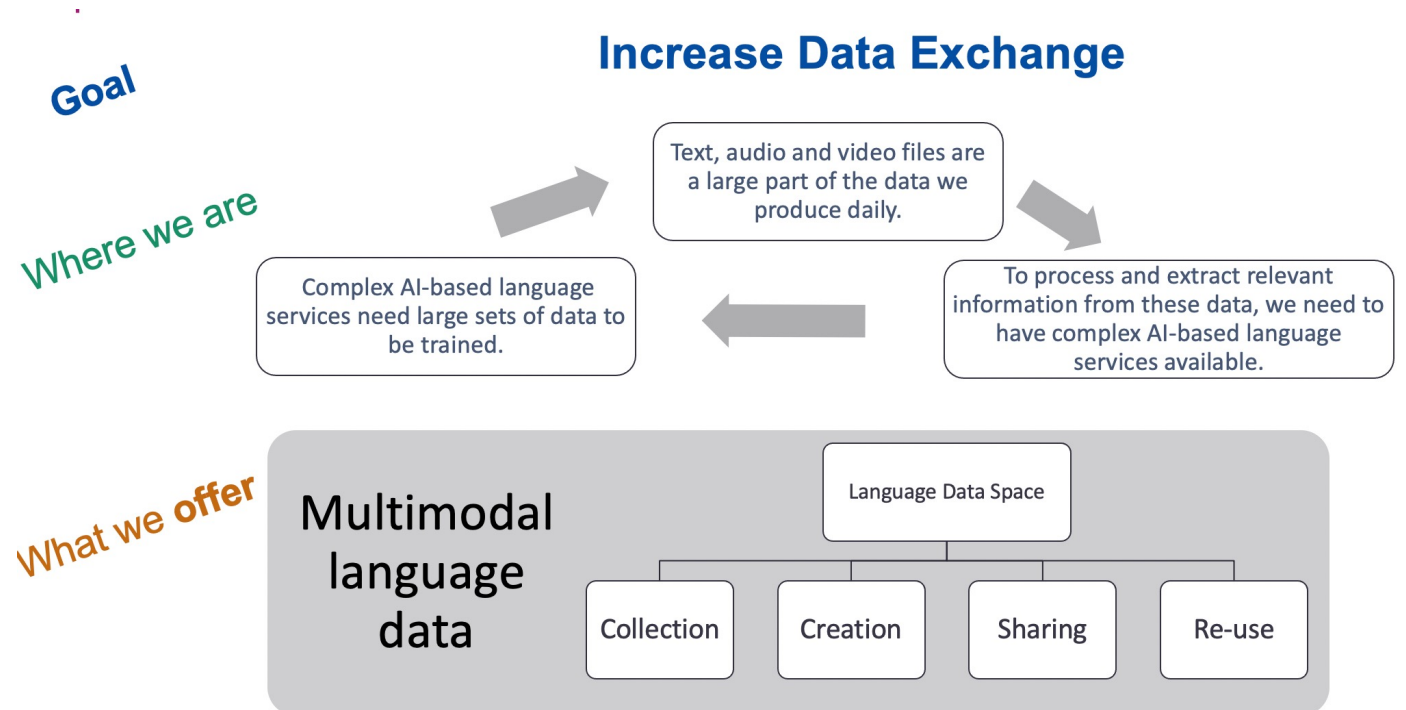
The LDS aims to give stakeholders the opportunity to monetise their efforts in terms of language resources (data, tools, services, models, etc.), while also supporting the deployment of language models and language technology services for their businesses, in one single marketplace.

The objective is to create an interconnected and competitive European data economy for the valorisation and re-use of language resources to spur innovation and economic growth.

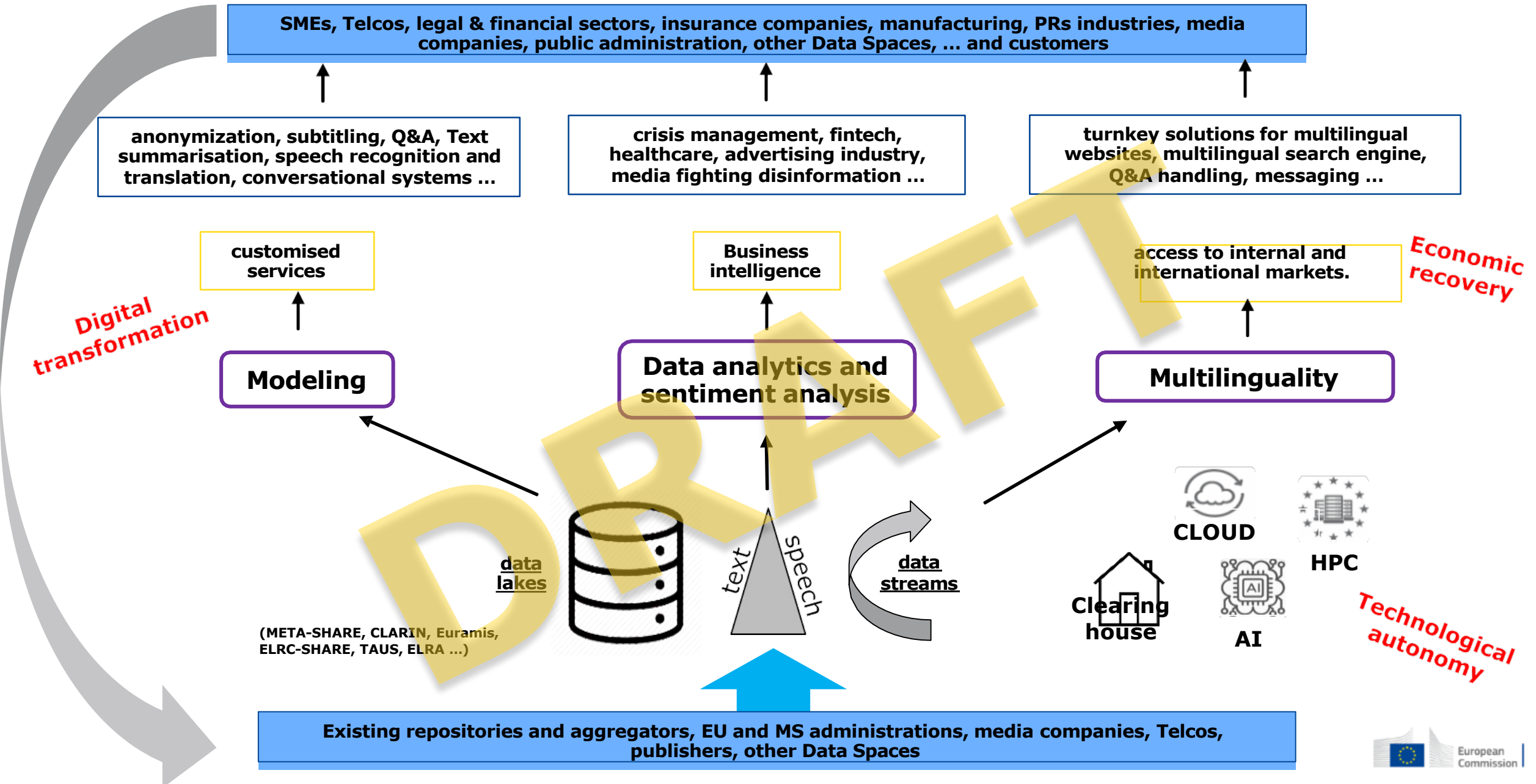
The Language Data Space - *What*

Collection, creation, sharing, reuse and monetisation of European multimodal language data through **a platform** and **a marketplace** *where the request for language resources and services matches the offer.*

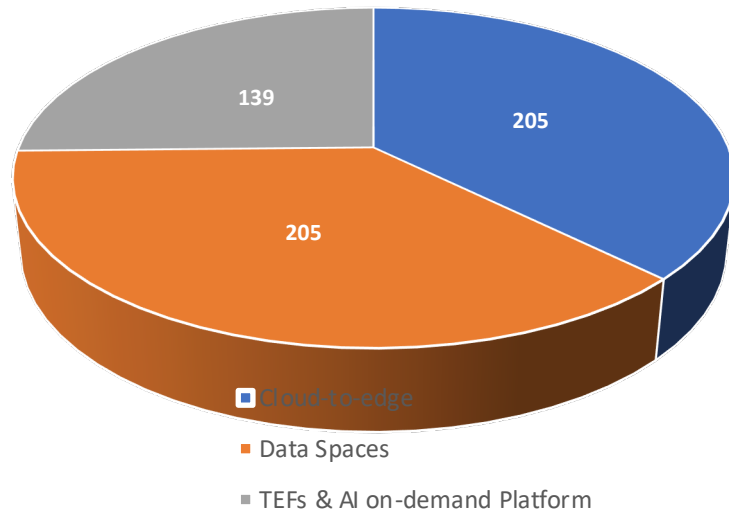
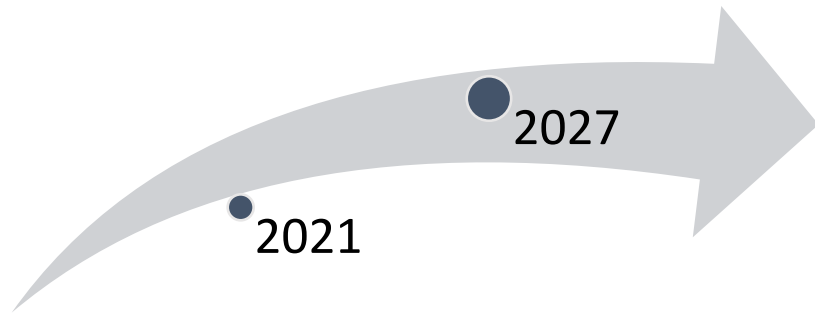
LDS is a thematic data space building on an **existing service offer** for research, public sector, and industry.



Language Data Space – **What**: a Multitude of Functions



Where in the MFF budget



DIGITAL in MFF 2021-2027

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

[Funding and Tenders](#)

DIGITAL WP 2021-2022

Total DIGITAL WP financing €1.4 Bio	<p>SO2 'Could, Data and AI'</p> <p>Cloud-to-Edge Infrastructure and Services: €205 Mio</p> <p>Data Spaces: €205 Mio</p> <p>Language Data Space: €6 Mio</p> <p>TEFs & AI on-demand Platform: €139 Mio</p> <p>Total financing for <i>Cloud, Data and AI</i>: €549 Mio</p>
---	--

LDS Financing - When

2021-2022

Type of Action	PROCUREMENT
Indicative Budget	EUR 6 million
Call Opening	July – September 2022
Indicative Starting Date	Early 2023

2023-2024

Planned types of actions: *procurement* and *SME support grants*.

How will the LDS come about – main steps

1

Establishing an institutional Centre of Excellence for Language Technologies (**CELT**), making use of existing EU initiatives for language data collection such as ELRC, CLARIN, and META-SHARE

Actions:

Governance layer

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

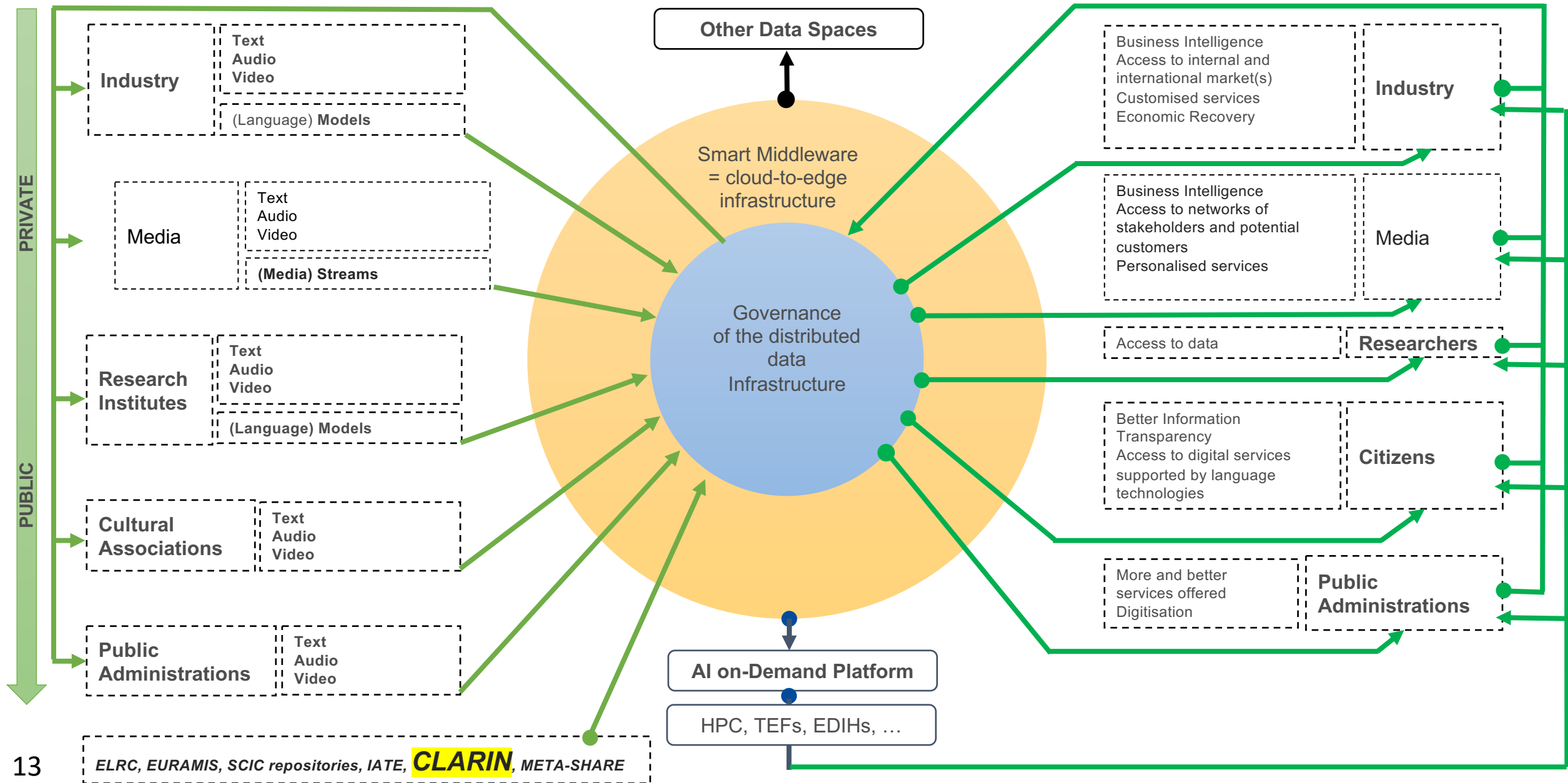
2

Deploying the Language Data Space, making use of existing platform technology for the exchange of distributed data such as the European Language Grid, CEF eTranslation and the CLARIN VLO

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.

Language Data Space – Stakeholders' Landscape



What CLARIN RI will contribute to the LDS



1. The technology behind CLARIN's discovery platform VLO
2. Models and expertise for the creation and harvesting of metadata for distributed multilingual and multimodal language resources
3. Expertise with a range of licensing models
4. Link with the interoperability framework emerging from  EOSC
5. Experience with the design and implementation of a governance model for multi-country collaboration.

see you @

www.clarin.eu

