



A Language Data Space for Europe

Monica PRETTI

Programme Officer
Data Directorate
Directorate-General for Communications Networks, Content and Technology
European Commission

My Europe, My Language

Brokering Event DIGITAL Data for EU, 16th December 2021

Language Data Space in DIGITAL WP 21-22

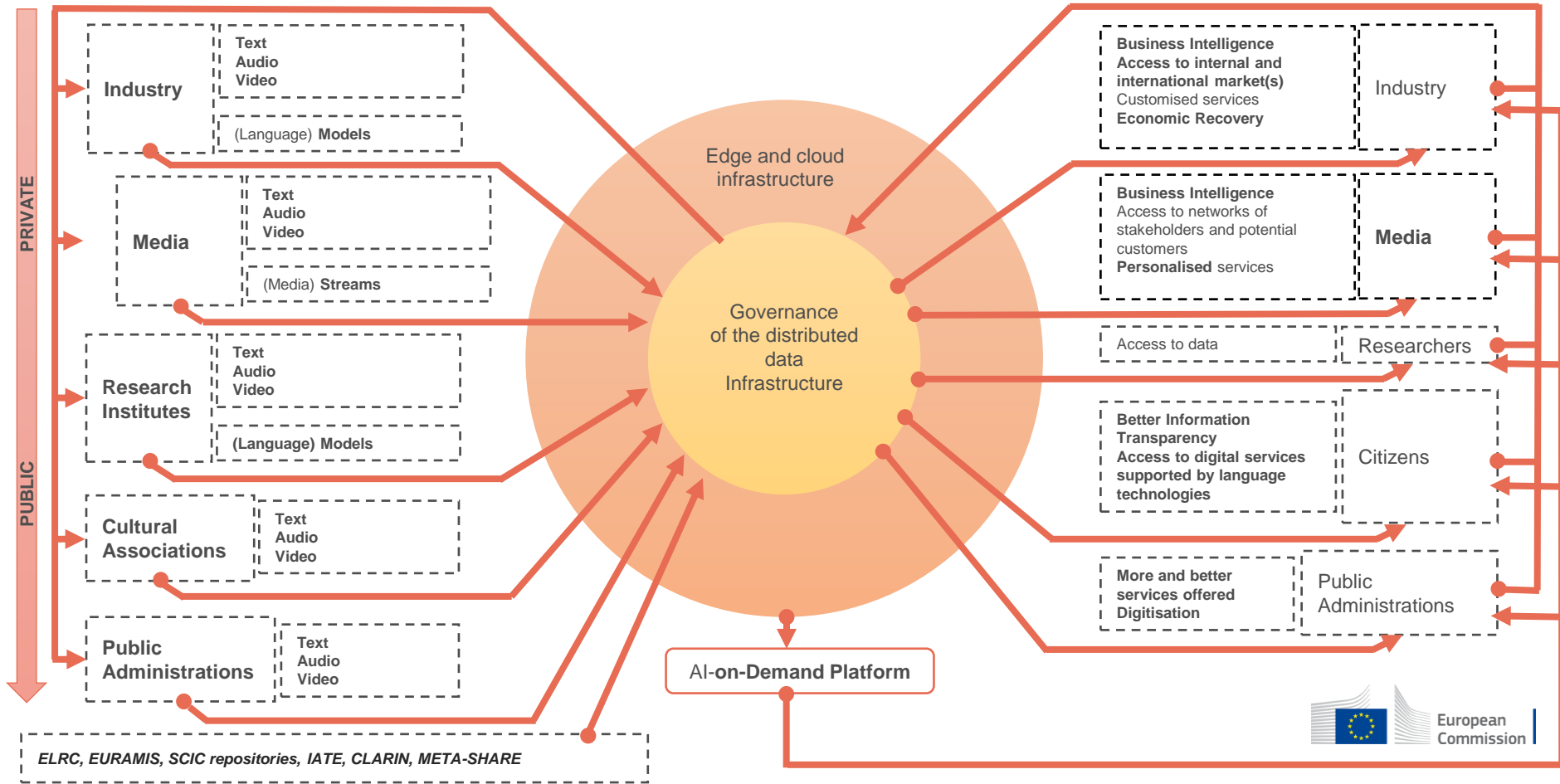
Type of Action	PROCUREMENT
Indicative Budget	EUR 6 million
Indicative Time of the Call Opening	2022
Indicative Duration of the Action	24 months
Implementation	European Commission

[DIGITAL WORK PROGRAMME 21-22](#)

Worth Noting !



Language Data Space – Visualisation



Language Data Space – Operational

Two Work Strands

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

Deploying the Language Data Space

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

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 – 1st Work Strand

General

CELT

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

This will make use of existing EU initiatives of language data collections such as ELRC, EURAMIS, SCIC repositories, IATE, CLARIN, META-SHARE.

Detail

Develop in close collaboration with the MSs a multi-stakeholder **data and services governance scheme**, bringing together large industrial entities, public stakeholders and small-and-mid-size enterprise stakeholders.

Identify and bring together existing European stakeholders and initiatives.

Elaborate a blueprint for the **language data ecosystem** based on existing EU legislation and language data policies to build innovative business models across all stakeholders.

Identify the **large multimodal language models** to be deployed.

Identify datasets and data streams (public, private, citizen-collected...) relevant to the creation of large language models to be brought into conformity with the new blueprint standards and principles.

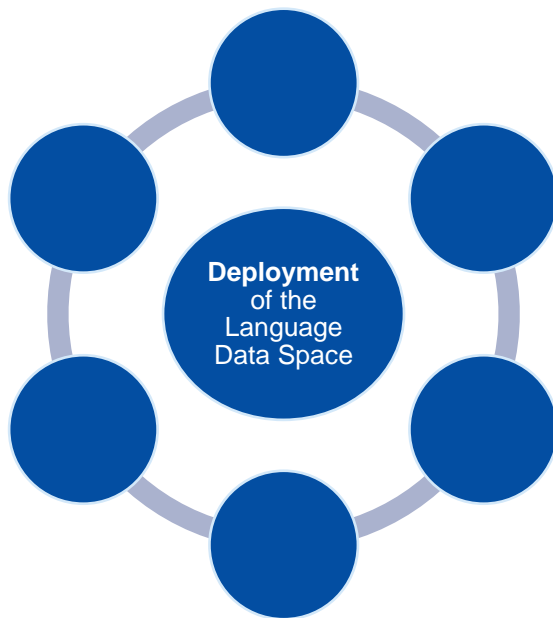
Establish a **detailed roadmap** on how to deploy the language data space.

Language Data Space – 2nd Work Strand

Implementing the **data & services governance**, business models and strategies as defined by CELT.

Deploying the necessary **infrastructure** for the collection and sharing of multimodal language data and models.

Deploying advanced **AI-based language processing services and tools** to be made available through the AI-on-demand platform.



Supporting the increased uptake and usage of the language data space and its derived language technologies services among European **private and public sectors**, including various DIGITAL platforms and endeavours/initiatives, also through **piloting and deployment projects**.

Establishing Member States and industrial collaboration to bring the set of identified language datasets and data streams into **conformity** with the new blueprint standards and principles.

Establishing Member States and industrial collaboration **to create and deploy the identified large multimodal language models**.

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

MONICA PRETTI

European Commission

DG CONNECT - Communications Networks, Content and
Technology

Dir G: Data

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

Monica.PRETTI@ec.europa.eu

