DATAWEEK

JOIN.LEARN.SHARE.GET VALUE

EDIC (European Digital Infrastructure Consortia) and Data Spaces

EUHUBS4DATA

BDV BIG DATA VALUE ASSOCIATION

Introduction and setting the scene

- Kimmo Rossi, EC DG CNECT.G. EDIC and Data Spaces: Overview from the European Commission
- Mikael Blomberg, Government Offices of Sweden EDIC: Viewpoint from a Member State
- Ilari Pulli, ERIC A perspective from the ERICs (EuroBioImaging Give a picture of what has worked well and what can be improved in EDIC to capitalise on experiences and provide the best conditions for the new initiative

European Commission

2030 DIGITAL COMPASS

THE EUROPEAN WAY FOR THE DIGITAL DECADE

Digital Decade MCPs and EDICs

Kimmo Rossi, DG CNECT.G.1

Data Week, Luleå, 14/6/2023

The key to understanding EDIC

- **What is an EDIC?**

  A ***European Digital Infrastructure Consortium (EDIC)*** is a governance structure designed to support **Multi-Country Projects** (MCPs).

- **What is an MCP?**

  A ***Multi-Country Project (MCP)*** supports the achievement of the **Digital Decade** targets and objectives.

- **What is the DDPP?**

  The ***Digital Decade Policy Programme*** outlines the strategy and roadmap to guide **Europe's digital transformation**

Digital Decade targets 2030 (Art 4 of DDPP)

DIGITAL SKILLS

Adults with basic digital skills

NOW 56%

TARGET 80%

Employed ICT specialists

NOW 8.4 million

TARGET 20 million

DIGITAL INFRASTRUCTURES

Gigabit network coverage

NOW 59%

TARGET 100% of households

5G coverage

NOW 14%

TARGET 100% of populated areas

The EU production of semiconductors, including processors, makes up

NOW 10%

TARGET ≥20% of world production in value

There are **10,000 edge nodes** in the EU for **better, secure and sustainable data processing.**

By 2025, the first **EU computer with quantum acceleration** is paving the way for cutting-edge quantum capabilities.

DIGITAL TRANSFORMATION OF BUSINESSES

BUSINESSES USING

Cloud computing services

NOW 26%

TARGET 75%

Big data

NOW 14%

TARGET 75%

Artificial Intelligence

NOW 25%

TARGET 75%

SMEs with at least a basic level of digital intensity

NOW 60%

TARGET >90%

There are

NOW 122

TARGET **Double** the number of unicorn startups in Europe

DIGITALISATION OF PUBLIC SERVICES

Online access to key public services (related to career, studying, family, regular business operations, moving)

NOW 75/100 services for citizens; 84/100 services for businesses

TARGET **All** services for citizens and businesses

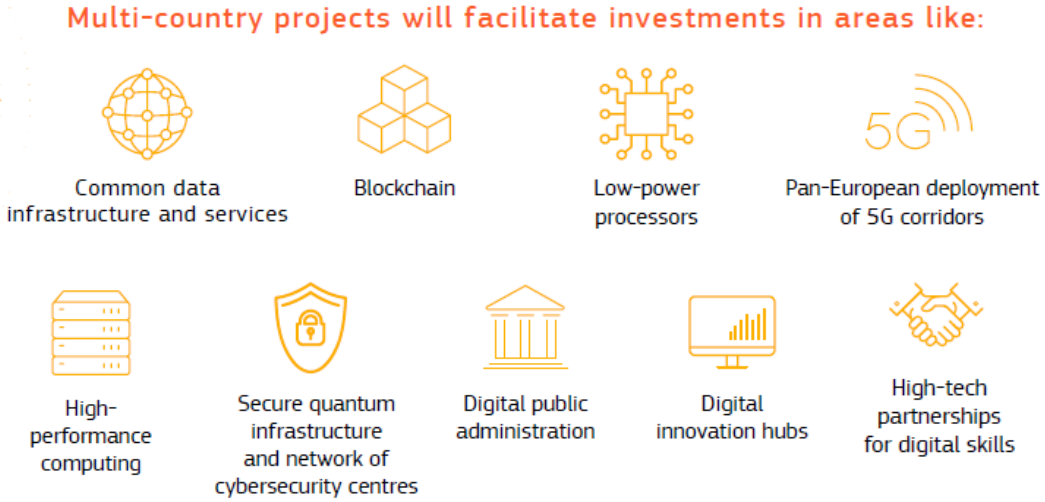
TARGET 100% of Europeans with access to digital health records &

80% using a digital id

European Commission

DDPP Decision: multi-country projects (Articles 10-12 of DDPP)

- **MCPs**: large scale projects for the EU digital transformation, supporting the achievement of the Digital Decade objectives and targets.
- **MCPs**: Addressing strategic vulnerabilities and dependencies along the supply chain.
- **MCPs**: pool EU, national, and private resources to achieve progress that no MS could do on its own.



European Commission

Initial list of areas for multi-country projects (as listed in the DDPP Annex)

- **European common data infrastructure and services;**
- Endowing the Union with the next generation of low-power trusted processors;
- Developing the pan-European deployment of 5G corridors;
- Acquiring supercomputers and quantum computers, connected with the EuroHPC;
- Developing and deploying ultra-secure quantum and space-based communication infrastructures;
- Deploying a network of security operations centres;
- Connected public administration;
- European blockchain services infrastructure;
- European digital innovation hubs;
- High-tech partnerships for digital skills through the Pact for Skills;
- Skills and training in cybersecurity;
- Other projects which meet the criteria of the Decision, and which become necessary to the achievement of the objectives of the Digital Decade due to emerging social, economic or environmental developments.

European Digital Infrastructure Consortia (DDPP Art. 13-21)

EDIC is

- a legal entity set up by a decision of the Commission upon a request of a group of (at least 3) Member States. Its legal personality and full legal capacity is recognised in all Member States.
- one of the many (Art. 11 of DDPP) possible mechanisms for **implementing MCPs**.
- inspired by the success of European **Research** Infrastructure Consortia (E**R**ICs).
- **pooling** Commitments (money or in-kind contributions) of the members and grants from EU instruments as the main sources of financing.
  - The EDIC will be able to apply for funding from the centrally-managed EU programmes (e.g., DEP, CEF, Horizon Europe), RRF and funds under shared management.
  - The rules of each programme apply, and the overall requirements arising from applicable state aid legislation, Financial Regulation (especially, the no-profit rule), etc.

European Digital Infrastructure Consortia (DDPP Art. 13-21)

Characteristics/advantages:

- Tailored to the needs of (infrastructure) **deployment** projects (as opposed to research&innovation).
- Provide **continuity** and long-term sustainability (as opposed to EU-funded projects of 3-4 years).
- Member States who provide financial/non-financial contributions will have voting rights. Other MSs may join EDICs as observers.
- Other public/private entities can become members, but they may not be able to outvote MSs.
- Speedy setup and flexible implementation, remaining open for new members throughout the lifecycle of the EDIC.
- Flexible internal structure defined in the **statutes**, adopted by EDIC members.
- The Commission participates in the deliberations of the assembly of members, without voting rights.

Sources of information

- Digital Decade Policy Programme Decision (32022D2481)
- Digital Decade web page
- Digital Decade targets

PRESENTATIONS:

Language Data Spaces

Thibault Grouas, French Government

Philippe Gelin, EC

MINISTÈRE
DE LA CULTURE
Liberté
Égalité
Fraternité

MINISTÈRE
DE L'ÉCONOMIE
DES FINANCES
ET DE LA SOUVERAINETÉ
INDUSTRIELLE ET NUMÉRIQUE
Liberté
Égalité
Fraternité

European Digital Infrastructure consortium

Alliance for Language Technologies ALT EDIC proposal

Dataweek 2023, June 14th

MINISTÈRE
DE LA CULTURE
Liberté
Égalité
Fraternité

MINISTÈRE
DE L'ÉCONOMIE
DES FINANCES
ET DE LA SOUVERAINETÉ
INDUSTRIELLE ET NUMÉRIQUE
Liberté
Égalité
Fraternité

Alliance for Language Technologies EDIC

Key objectives

Preserve linguistic and cultural diversity in Europe

Technological leadership and strategic autonomy

Respect European rules and values

Cooperation

Raising awareness

Linguistic diversity

MINISTÈRE
DE LA CULTURE
Liberté
Égalité
Fraternité

MINISTÈRE
DE L'ÉCONOMIE
DES FINANCES
ET DE LA SOUVERAINETÉ
INDUSTRIELLE ET NUMÉRIQUE
Liberté
Égalité
Fraternité

Main thematic areas of action plan

DW

MINISTÈRE
DE LA CULTURE
Liberté
Égalité
Fraternité

MINISTÈRE
DE L'ÉCONOMIE
DES FINANCES
ET DE LA SOUVERAINETÉ
INDUSTRIELLE ET NUMÉRIQUE
Liberté
Égalité
Fraternité

Area 1 **Existing models**

Existing models
- Gather existing open-source language models

Existing models
- Fine-tuning existing models for SMEs

Existing models
- Reduce and optimize existing models (in particular bloom)

Existing models
- Provide evaluation, certification and normalisation methodologies

MINISTÈRE
DE LA CULTURE
Liberté
Égalité
Fraternité

MINISTÈRE
DE L'ÉCONOMIE
DES FINANCES
ET DE LA SOUVERAINETÉ
INDUSTRIELLE ET NUMÉRIQUE
Liberté
Égalité
Fraternité

Area 2 Development of models

Development of models

- Launch new open-source model projects

Development of models

- Coordination of action of European sovereign supercomputer (**Euro HPC**)

Development of models

- Provide logistical and operational support for public and private experts

MINISTÈRE
DE LA CULTURE
Liberté
Égalité
Fraternité

MINISTÈRE
DE L'ÉCONOMIE
DES FINANCES
ET DE LA SOUVERAINETÉ
INDUSTRIELLE ET NUMÉRIQUE
Liberté
Égalité
Fraternité

Area 3
Data

Data

- A central platform for European language resources

Data

- Collect quality data sets (continuity with the ***LDS***, ***ELG***, ***ELRC-SHARE***, … actions)

Data

- Creation of data (strategic for languages with few speakers)

MINISTÈRE
DE LA CULTURE
Liberté
Égalité
Fraternité

MINISTÈRE
DE L'ÉCONOMIE
DES FINANCES
ET DE LA SOUVERAINETÉ
INDUSTRIELLE ET NUMÉRIQUE
Liberté
Égalité
Fraternité

Area 4
Ecosystem

- Ecosystem
  - Host a start-up incubator
- Ecosystem
  - Develop and foster relationships between **industry** and research
- Ecosystem
  - Act as a key player of the European coordinated plan on AI
- Ecosystem
  - Bring together and strengthen the LT community
- Ecosystem
  - Provide dedicated support to institutions for investing in LTs
- Ecosystem
  - Cultural program based on AI for languages

MINISTÈRE
DE LA CULTURE
Liberté
Égalité
Fraternité

MINISTÈRE
DE L'ÉCONOMIE
DES FINANCES
ET DE LA SOUVERAINETÉ
INDUSTRIELLE ET NUMÉRIQUE
Liberté
Égalité
Fraternité

Alliance for Language Technologies EDIC

Timeline

Expression of interest	Pre-notification to the Commission	Formal application of ALT EDIC
February 2023	May 31st, 2023	Autumn, 2023
France + 23 members states interested in language technology	France as coordinator, 10 MS participating, 7 MS observers Draft statutes, budget, action and implementation plan	France as coordinator Working group meetings Final statutes, budget, action and implementation plan

ALT EDIC should be up and running before the end of 2024

MINISTÈRE
DE LA CULTURE
Liberté
Égalité
Fraternité

MINISTÈRE
DE L'ÉCONOMIE
DES FINANCES
ET DE LA SOUVERAINETÉ
INDUSTRIELLE ET NUMÉRIQUE
Liberté
Égalité
Fraternité

Thank you for attending !

DW

The Common European Language Data Space

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

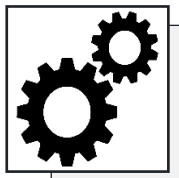
Data Week – Lulea – 14/06/2023

Data Spaces at a Glance

Green Deal

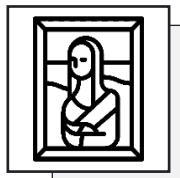
Smart Communities

Mobility



Manufacturing

Agriculture

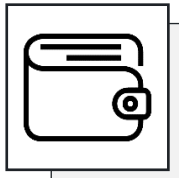


Cultural Heritage

Health, Genomics

Health, Cancer Images

Media



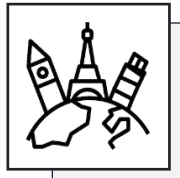
Financial

Skills

Language

Public Procurement

Security and Law Enforcement



Tourism

Why?

Goal

Increase Data Exchange

What is needed

Text, audio and video files are a large part of the data we produce daily.

To process and extract relevant information from these data, we need to have complex AI-based language services available.

Complex AI-based language services need large sets of data to be trained.

What we will offer

Multimodal language data

Language Data Space

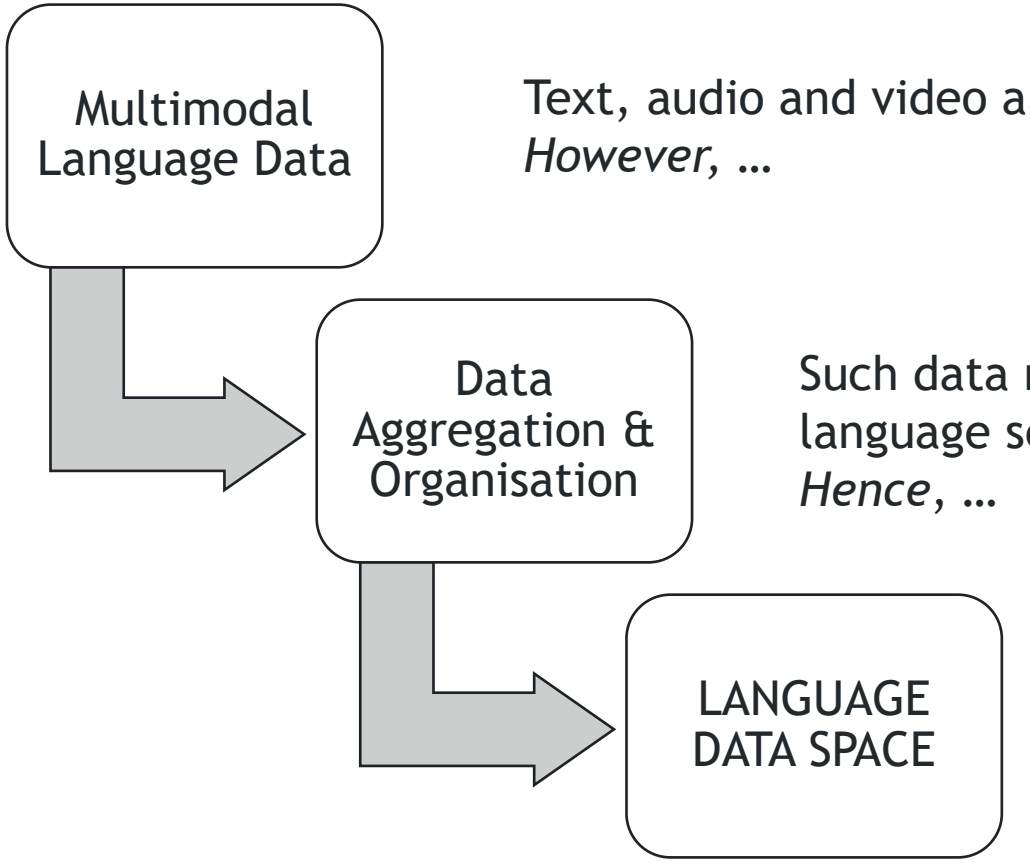
Collection

Creation

Sharing

Re-use

What ?



Text, audio and video are a large part of the data produced every day.
However, …

Such data need to be aggregated and organised for AI-based language services to process and extract relevant information.
Hence, …

1. for the collection, creation, sharing and reuse of multimodal language data;
2. as a support to the deployment of large multimodal language models;
3. as a support to the implementation of AI language technology services through the AI on-demand platform.

How - Design Principles

Data control

Data holders will retain full control over their data, giving or revoking authorisation and changing data access rights over time.

Governance Framework

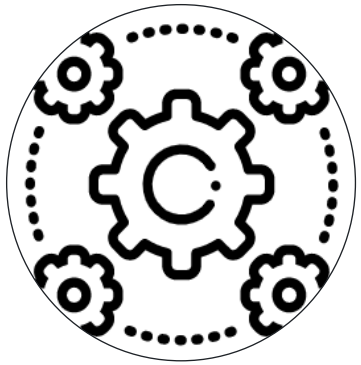
It will ensure fair, transparent, proportionate and non-discriminatory access to, sharing and use of data.

Respect of EU Rules & Values

In particular, personal data protection, consumer protection legislation and competition law.

Technical Data Infrastructure

It will consist of building blocks allowing building the DSs in an efficient and coordinated manner.



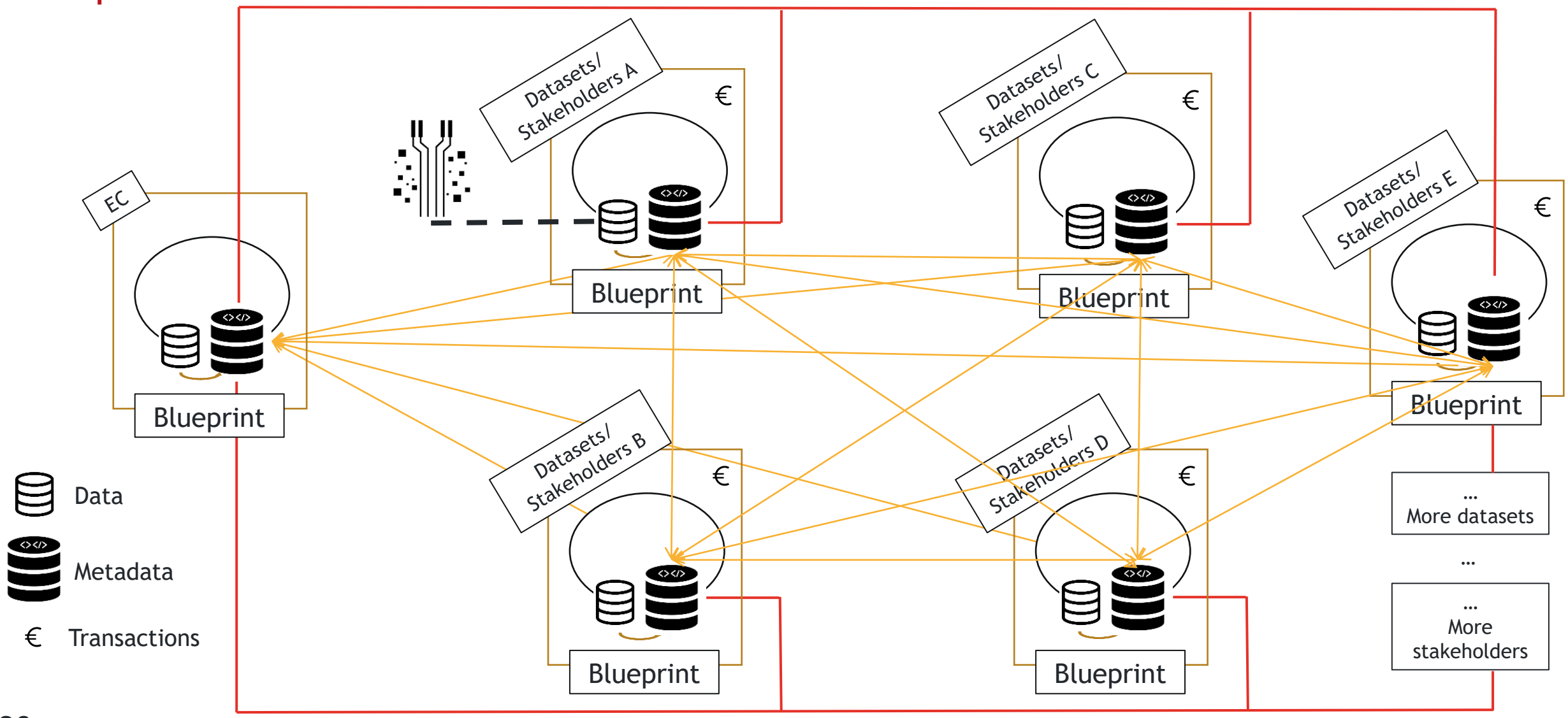
Interconnection & Interoperability

To favour exploitation of data, DSs will be progressively interlinked, thus leading to a fully integrated ecosystem.

Openness

Participation in DSs will be open to all actors, organisations or individuals, fostering competition and avoiding lock-ins.

HOW - Distributed Architecture



How ? Landscape

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

SO3
Cyber-security

SO4
Advanced Skills

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

European Data Spaces Technical Framework

consisting of

Cloud-to-Edge Infrastructures and Services

including

- Smart Middleware Solutions
- Marketplace

DSSC

Sectoral Data Spaces

Energy
Manufacturing
Agriculture
Health
Smart Cities
Mobility
...

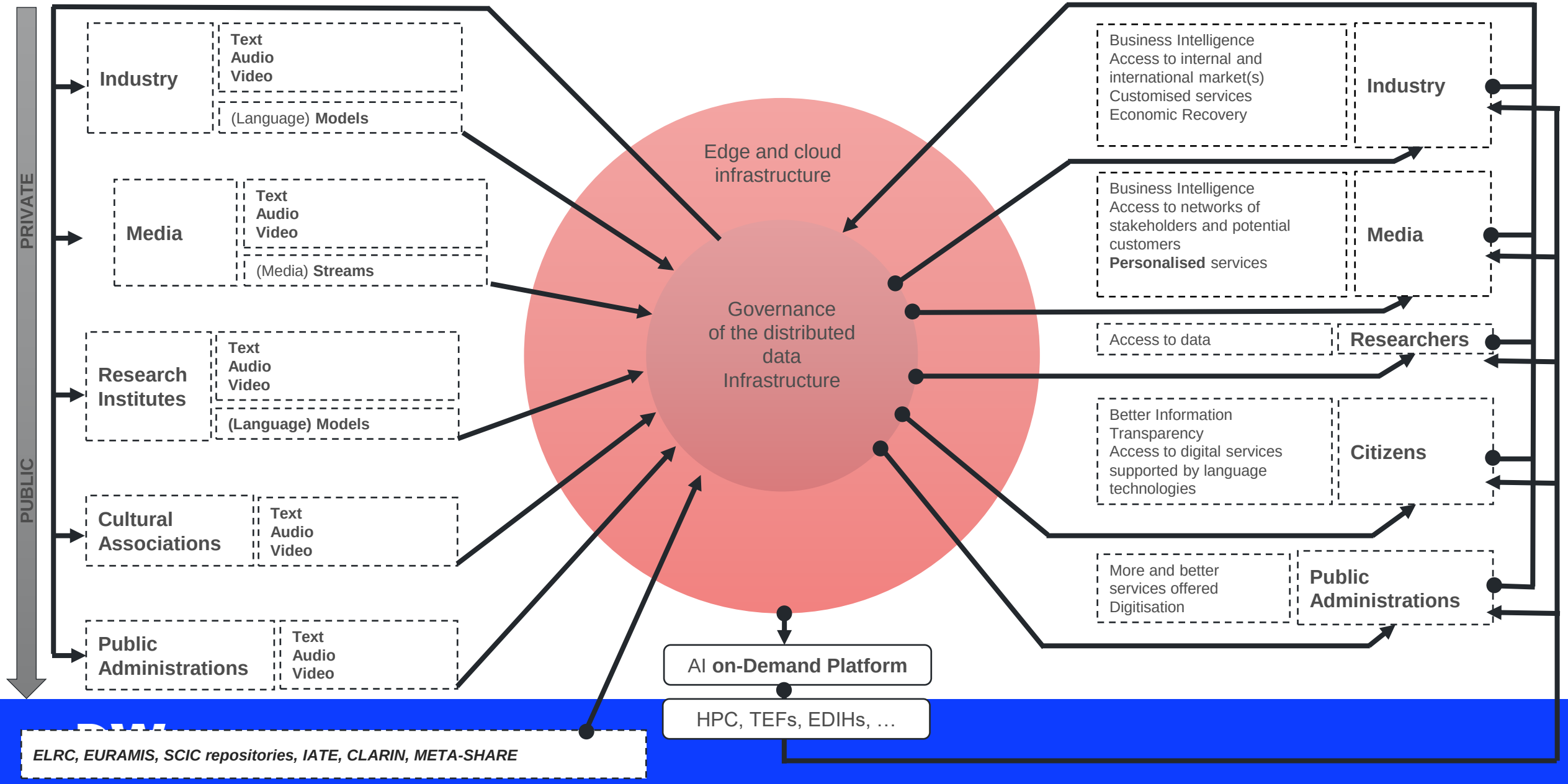
LANGUAGE DATA SPACE

DSSC - Data Spaces Support Centre
coordinating all relevant actions on sectoral DSs and making architectures and data infrastructure requirements available for the DSs.

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

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

WHO? Potentially Interested Stakeholder Groups



From Where?

ELRC

European Language Resource Coordination

under CEF - 'Connecting Europe Facility'

ELG

European Language Grid

under HORIZON 2020

Language Data Space

It has focused on collecting open language resources and language processing services, defining collection guidelines and promoting the ELRC-SHARE, a repository used for documenting, storing, browsing and accessing language data and tools.

It develops and deploys a scalable cloud platform, providing access to hundreds of commercial and non-commercial Language Technologies for all European languages, including running tools, services, data sets and resources.

How: LDS Work Strands

1 Establishing a **Governance** Scheme - **CELT (+)**

Collection, creation, sharing and re-use 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 on-demand platform.

How: LDS Work Strands

1 Establishing a **Governance** Scheme - **CELT (+)**

Collection, creation, sharing and re-use 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 on-demand platform.

How: Establishment of the Governance CELT (+)

CELT Coordinating across the MS	*Governance*	Developing the **Multi-Stakeholder Data and Services Governance Scheme (with the advice of CELT +)**
	Networking	Identifying and bringing together **public** European **stakeholders/initiatives** around: • Datasets and data streams • *Large language models*
CELT + Coordination across the Industry	*Governance / Planning*	**Blueprint for the Language Data Ecosystem** The plan will include a detailed **roadmap** on how to deploy the language data space
	Networking	Identifying and bringing together **private** European **stakeholders & business opportunities** around • Datasets and data streams • *Large language models*

How: Establishment of the Governance CELT (+)

CELT Coordinating across the MS	*Governance*	Developing the **Multi-Stakeholder Data and Services Governance Scheme (with the advice of CELT +)**
	Networking	Identifying and bringing together **public** European **stakeholders/initiatives** around: • Datasets and data streams • *Large language models*
CELT + Coordination across the Industry	*Governance / Planning*	**Blueprint for the Language Data Ecosystem** The plan will include a detailed **roadmap** on how to deploy the language data space
	Networking	Identifying and bringing together **private** European **stakeholders & business opportunities** around • Datasets and data streams • *Large language models*

secretariat@language-data-space.eu

How: LDS Deployment

Governance Level

Implementing the strategies defined by the CELT (+) → MS and industrial collaboration to deploy the identified language resources and models.

Technical Level

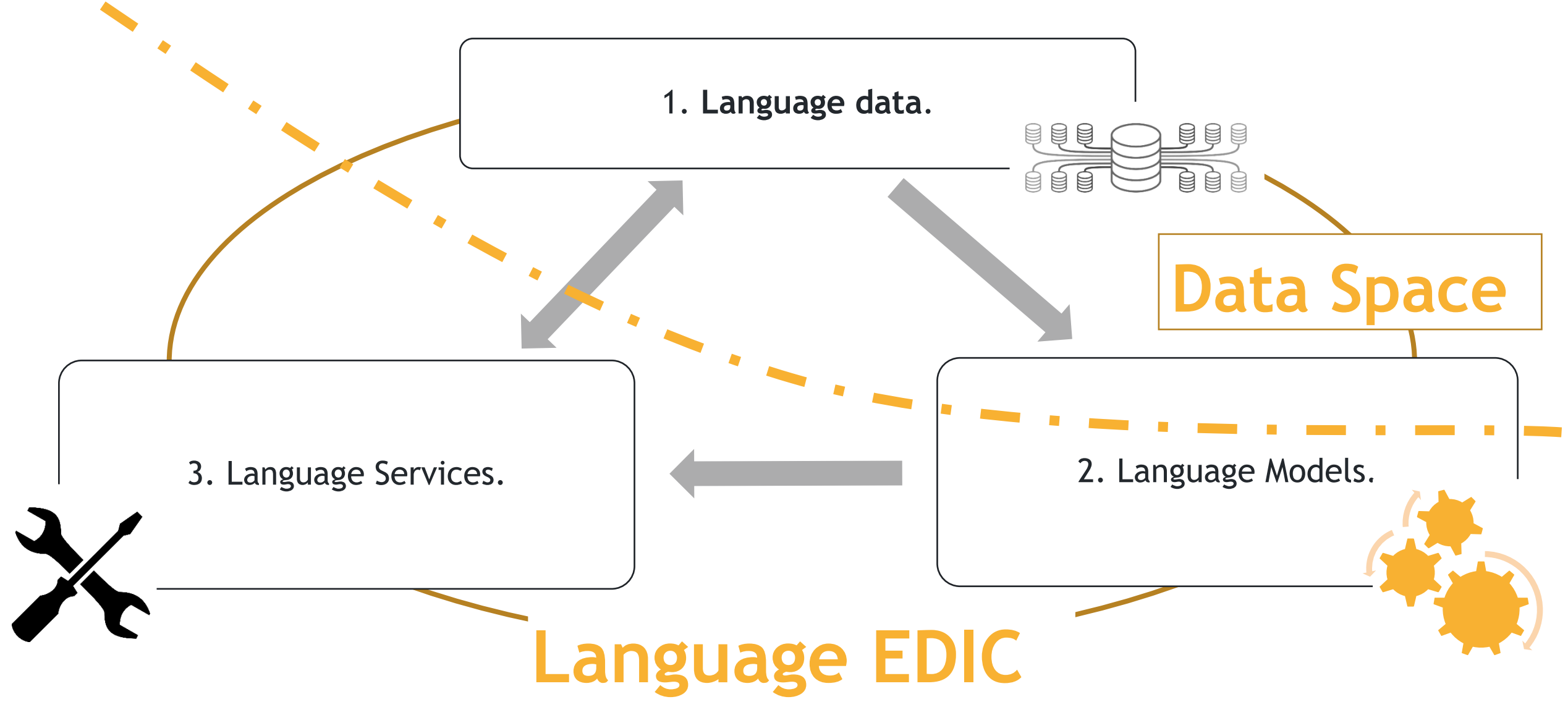
- Deploying the necessary **infrastructure** (cf. European Data Spaces Technical Framework).
- Bringing the set of **identified language data** into conformity with the Blueprint.
- Deploying advanced AI-based **language processing services and tools** to be made available through the AI on-demand platform.

Promotional Level

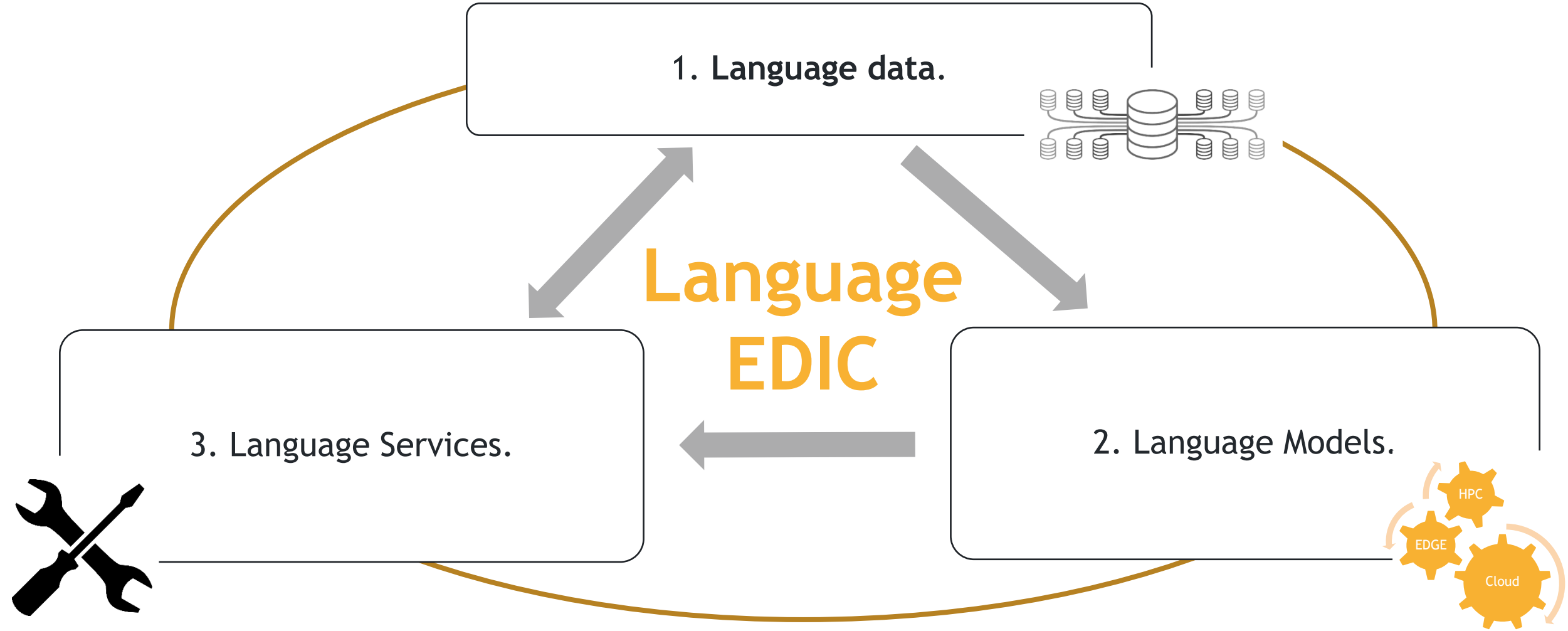
- Supporting the **increased uptake and use of the LDS** and its derived LT services among European private and public sectors, including various DIGITAL platforms and endeavours/initiatives, also through piloting and deployment projects.

Where to?

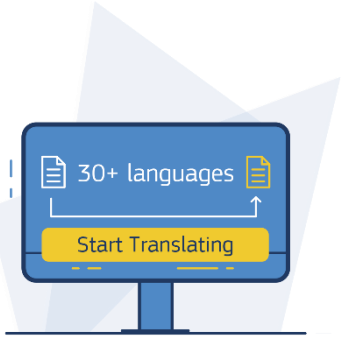
The European Language Technologies Landscape



European Language Technologies Landscape



Thank you very much!



Join us through: secretariat@language-data-space.eu

For any follow-up questions, remarks and/or queries, please feel free to send us an email at: CNECT-LANGTECH@EC.EUROPA.EU.

Need eTranslation or other language tools?
Visit: https://language-tools.ec.europa.eu/.

AgriFood

FABIO PIANESI

Chief Strategy Officer for European Affairs at Fondazione Bruno Kessler

RAFFAELE GIAFFREDA

Chief IoT Scientist at Fondazione Bruno Kessler

Mobility

Lucie Kirstein, Acatech

Discussion panel, Jeanette Nilsson RISE
What is the vision?
What is the value of EDICs?
Why and how to engage ?
How can private players engage and contribute (Industry and research)

- -Philippe Gelin (EC)
- -Tohid Ardeshiri (RISE)
- -Fabio Pianesi (FBK)
- -Kimmo Rossi (EC)

THE DATA IS OUT THERE

DATAWEEK

JOIN.LEARN.SHARE.GET VALUE

Thank you!

EUHUBS4DATA

BDV BIG DATA VALUE ASSOCIATION