The Common European Language Data Space

…for the Language Industry

LREC Conference 2022

22 June 2022

Why are we here today?

A **new opportunity** for more **language data**, **tools** and **services** and for their **monetisation**.

EC

We will present…

- what the Language Data Space (LDS) is

- how the Language Technology Industry can benefit from it

Language Technology Industry

You can share…

- insights into your specific needs to be met in the LDS

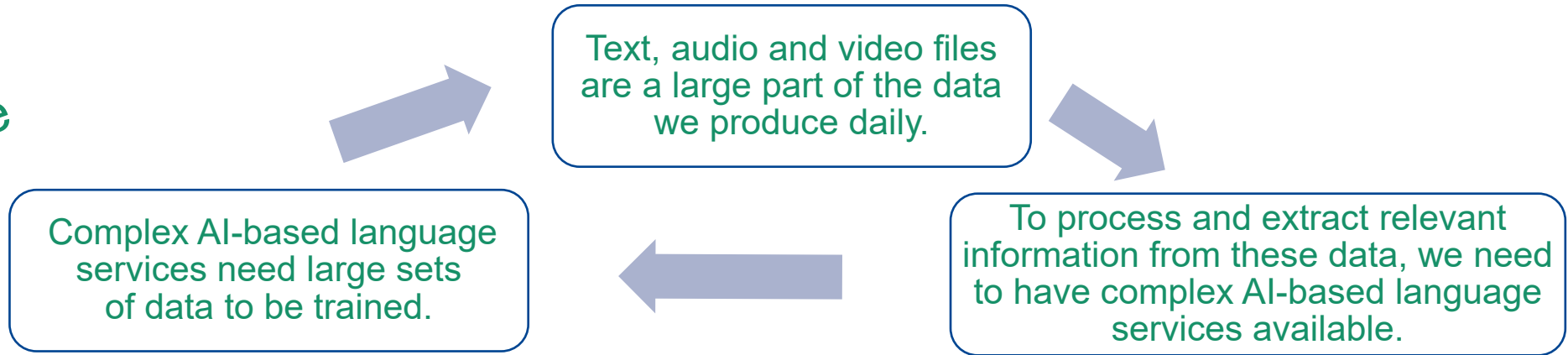- your thoughts on the LDS set-up and how to improve it

The Common European Language Data Space

Cross-fertilisation for a mutual gain

European Commission

Language Data Space… What is it?

Goal

Increase Data Exchange

Where we are

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 offer

Multimodal language data

Language Data Space

| Collection | Creation | Sharing | Re-use |
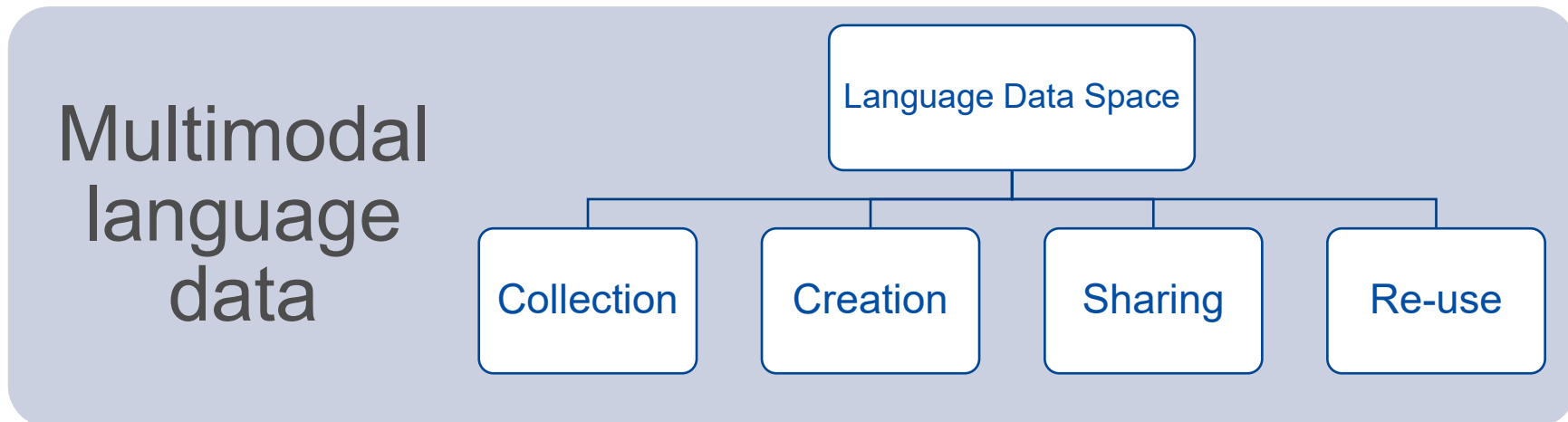
What's in it for you
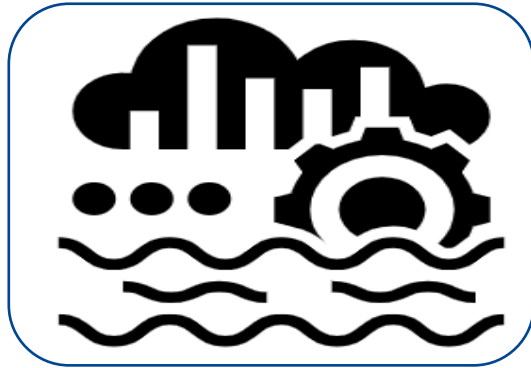
Data sharing against remuneration

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

Developing, selling and purchasing language-based services

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

E.g., multilingual and/or multimodal data

E.g., automatic speech transcription

E.g., automatic text summarisation

E.g., historical data

European Commission

Your involvement in the LDS initiative

Short-Term Today's Meeting	• Presenting and exchanging ideas on the LDS to give further shape to the project
Medium-Term June – September 2022	• Finding out more about the LDS (call for tenders publication) • Taking part in (consortia) proposals
Long-Term From January 2023 onwards	• Participating in the deployment of the LDS • Using the LDS as a platform and marketplace • Being part of the LDS community

European Commission

'Origins' of the Common European Data Spaces

Common European Data Spaces

are the centrepiece of the

European Strategy for Data

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

Objectives

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.

Framework Programme of DSs and LDS

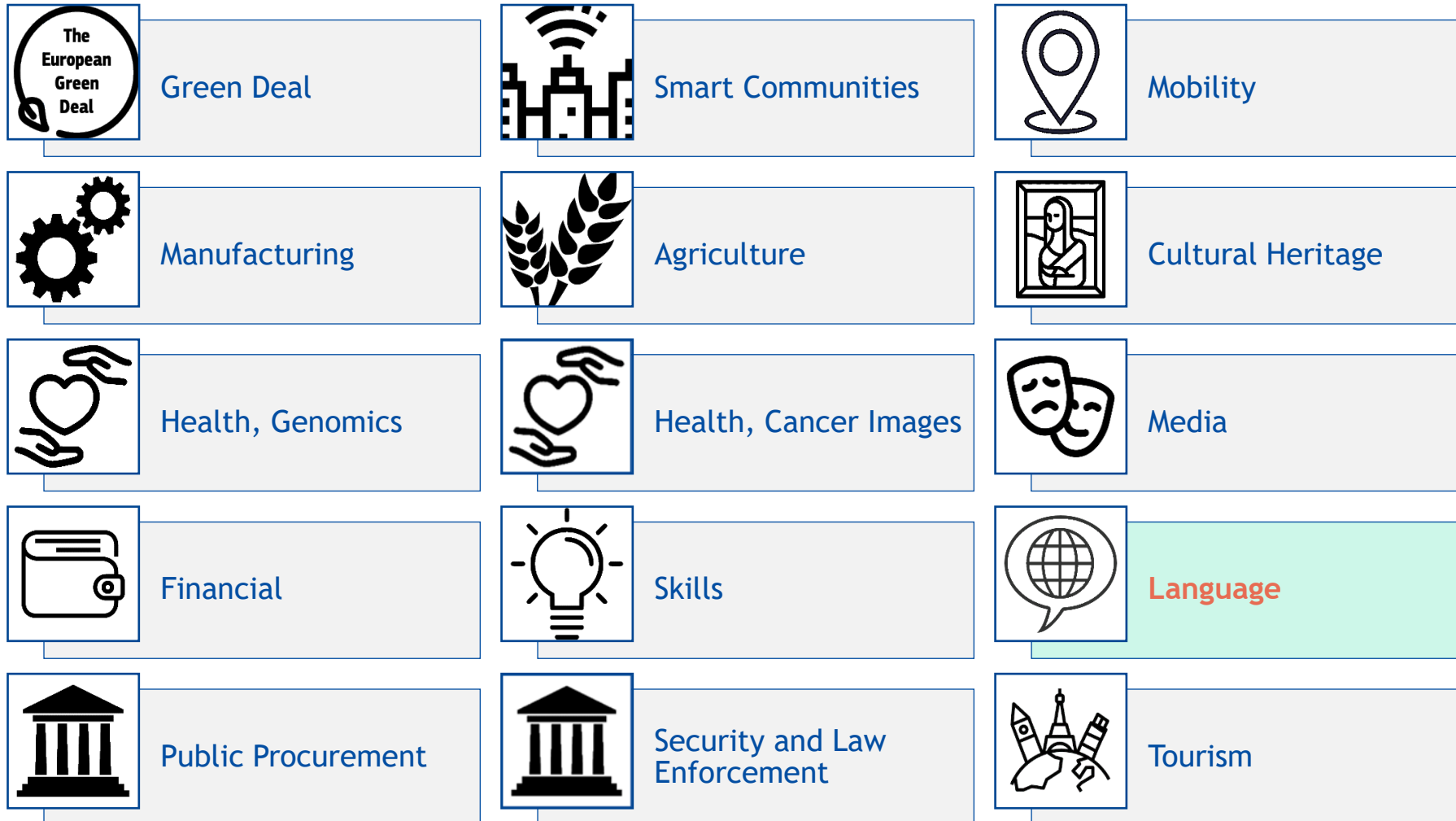The [Digital Europe Programme (DIGITAL)](#) provides funding for projects in five crucial areas:

SO1

SO4

SO3

HIGH PERFORMANCE COMPUTING

ADVANCED DIGITAL SKILLS

CYBERSECURITY

DIGITAL EUROPE PROGRAMME

SO5

SO2

CLOUD, DATA AND AI

BEST USE OF DIGITAL TECHNOLOGIES

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*.

European Commission

DIGITAL WP 21-22 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

European Commission
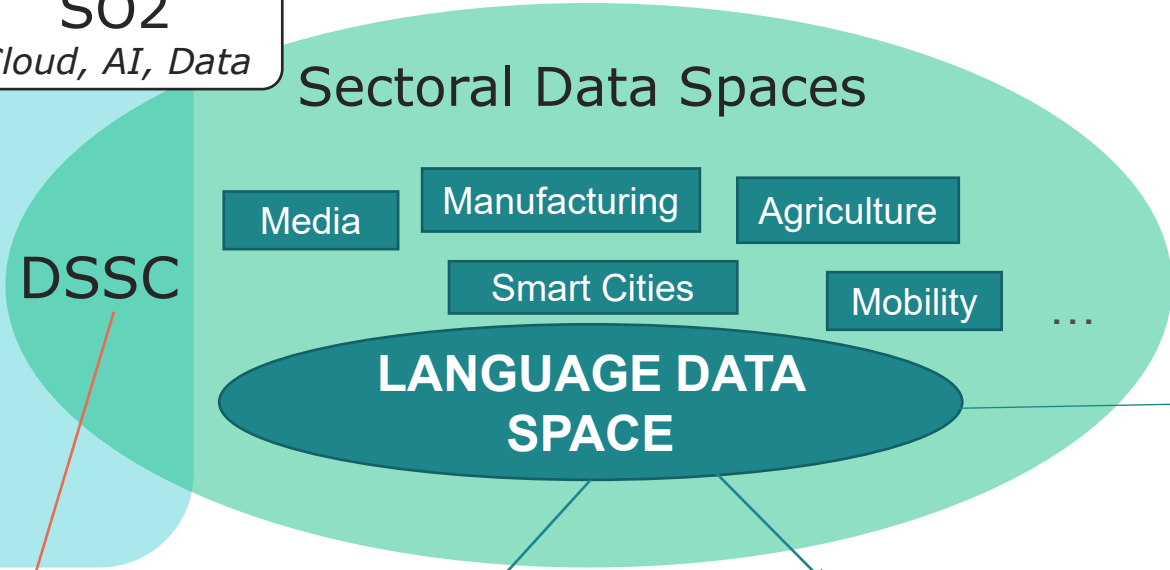
European DS Ecosystem – DIGITAL SO2

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

SO3
Cyber-security

SO2
Cloud, AI, Data

European Data Spaces Technical Framework

consisting of

Cloud-to-Edge Infrastructures and Services

including

- Smart Middleware Solutions
- Marketplace

DSSC

Sectoral Data Spaces

Media   Manufacturing   Agriculture

Smart Cities

Mobility   …

LANGUAGE DATA SPACE

SO4
Advanced Skills

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

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
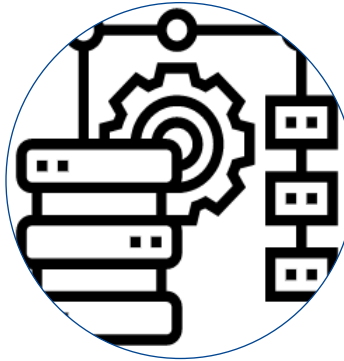for experimenting with AI-based solutions in real-world environments.

Design Principles of the LDS
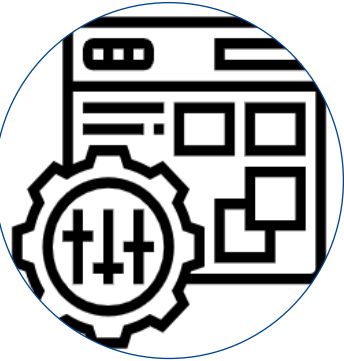
Data control

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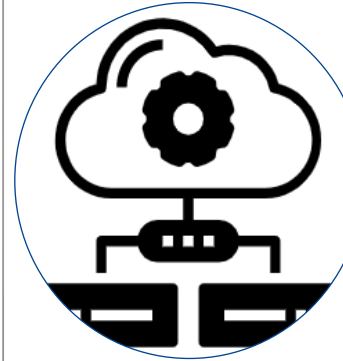- Option to change data access rights over time.

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
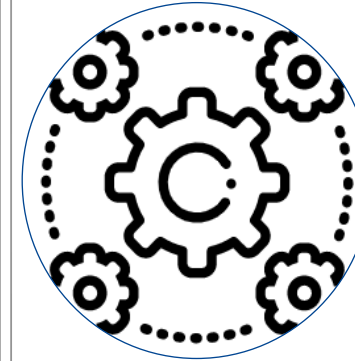
Interconnection & Interoperability

- 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
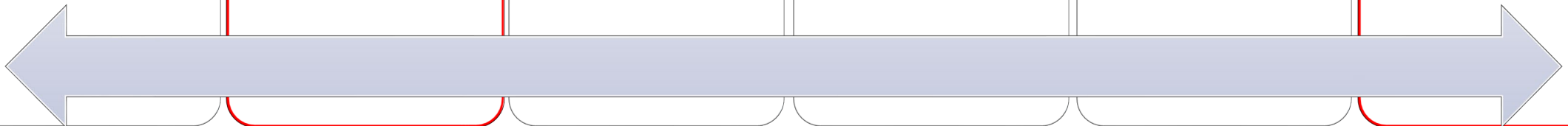
Financing

In DIGITAL WP **21-22**

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

In DIGITAL WP **23-24**

Planned types of actions: *procurement* & *grants*.

LDS Work Strands

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

2   **Deploying** the Language Data Space

Actions:

Governance layer

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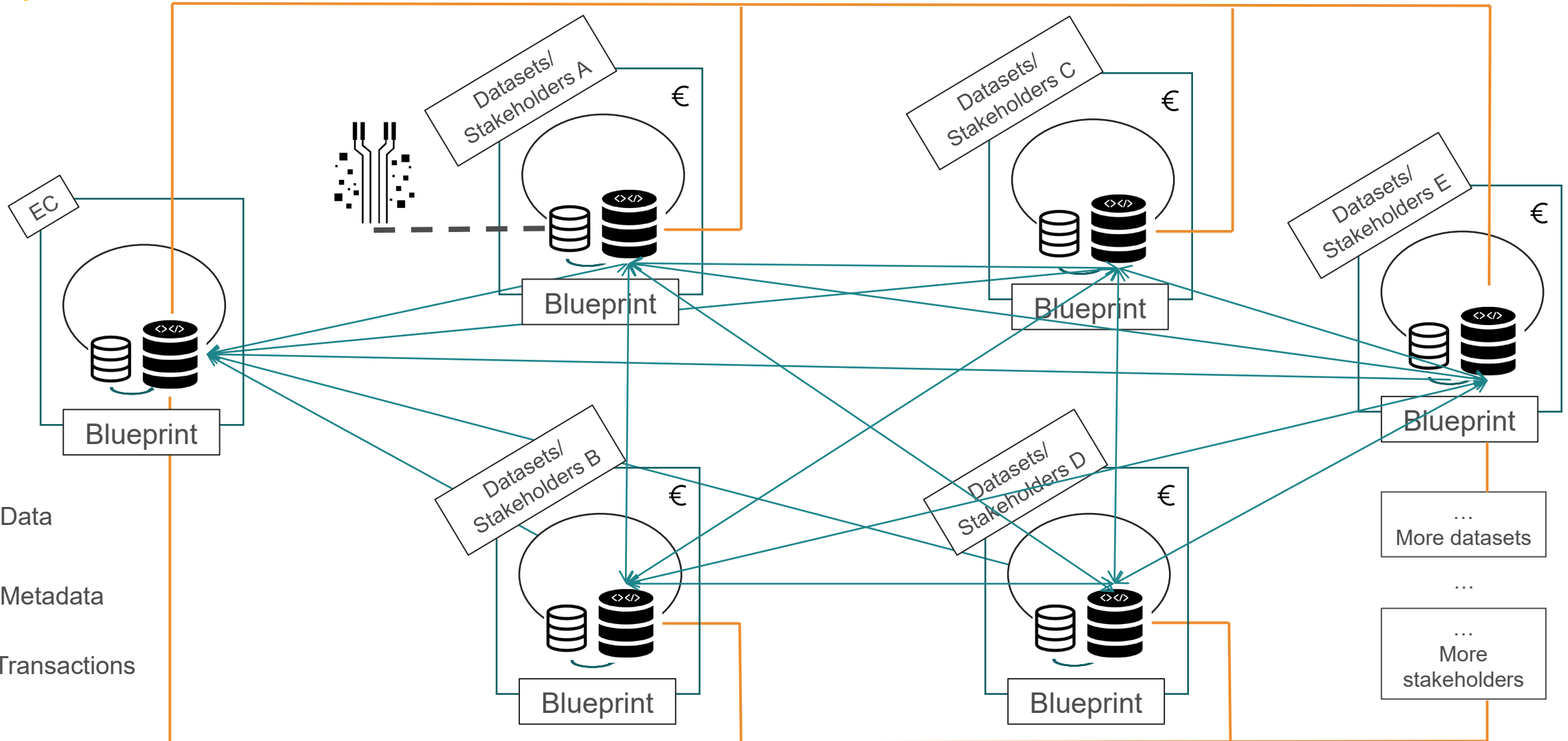
Operational layer

→ 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	**CELT**
	Coordinating across the MSs the creation and collection of multimodal language data and models.
		Networking	Identifying and bringing together existing European **stakeholders/initiatives** and **language resources**
			   • Datasets and data streams
			   • *Large language models*
			
			ELRC, EURAMIS, SCIC repositories, IATE, CLARIN, META-SHARE, etc.

		Governance	Developing a **Multi-Stakeholder Data and Services Governance Scheme**
Operational level	**CELT+** Driven by Industry	*Planning*	Elaborating a **Blueprint for the Language Data Ecosystem** based on existing EU/national legislation, incl. a detailed **roadmap** on how to deploy the language data space

EC Role: Facilitator

LDS Distributed Architecture

Stakeholders' Landscape

Other Data Spaces

PRIVATE

Industry
- Text
- Audio
- Video
- (Language) Models

Media
- Text
- Audio
- Video
- (Media) Streams

PUBLIC

Research Institutes
- Text
- Audio
- Video
- (Language) Models

Cultural Associations
- Text
- Audio
- Video

Public Administrations
- Text
- Audio
- Video

Smart Middleware = cloud-to-edge infrastructure

Governance of the distributed data Infrastructure

Business Intelligence
Access to internal and international market(s)
Customised services
Economic Recovery — Industry

Business Intelligence
Access to networks of stakeholders and potential customers
Personalised services — Media

Access to data — Researchers

Better Information
Transparency
Access to digital services supported by language technologies — Citizens

More and better services offered
Digitisation — Public Administrations

AI on-Demand Platform

HPC, TEFs, EDIHs, …

ELRC, EURAMIS, SCIC repositories, IATE, CLARIN, META-SHARE

European Commission
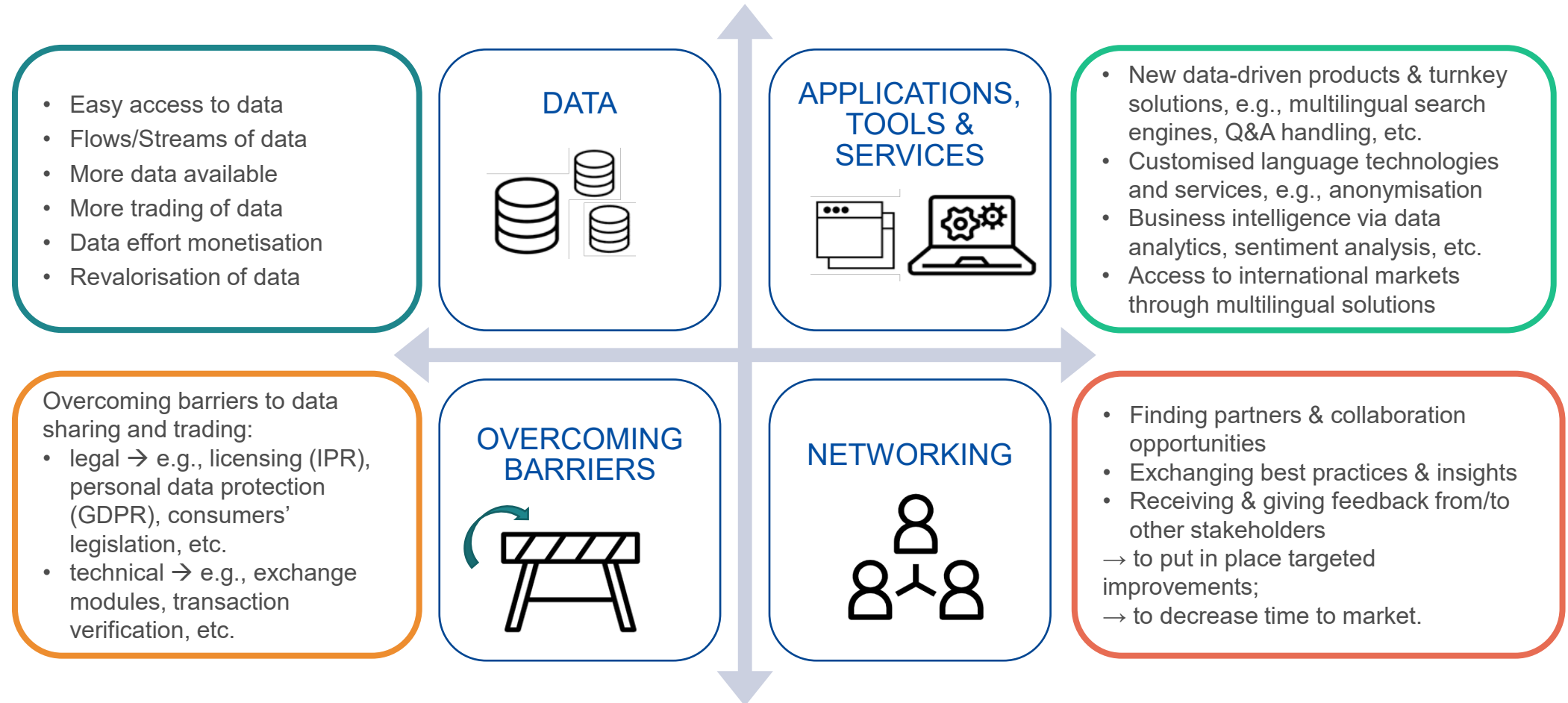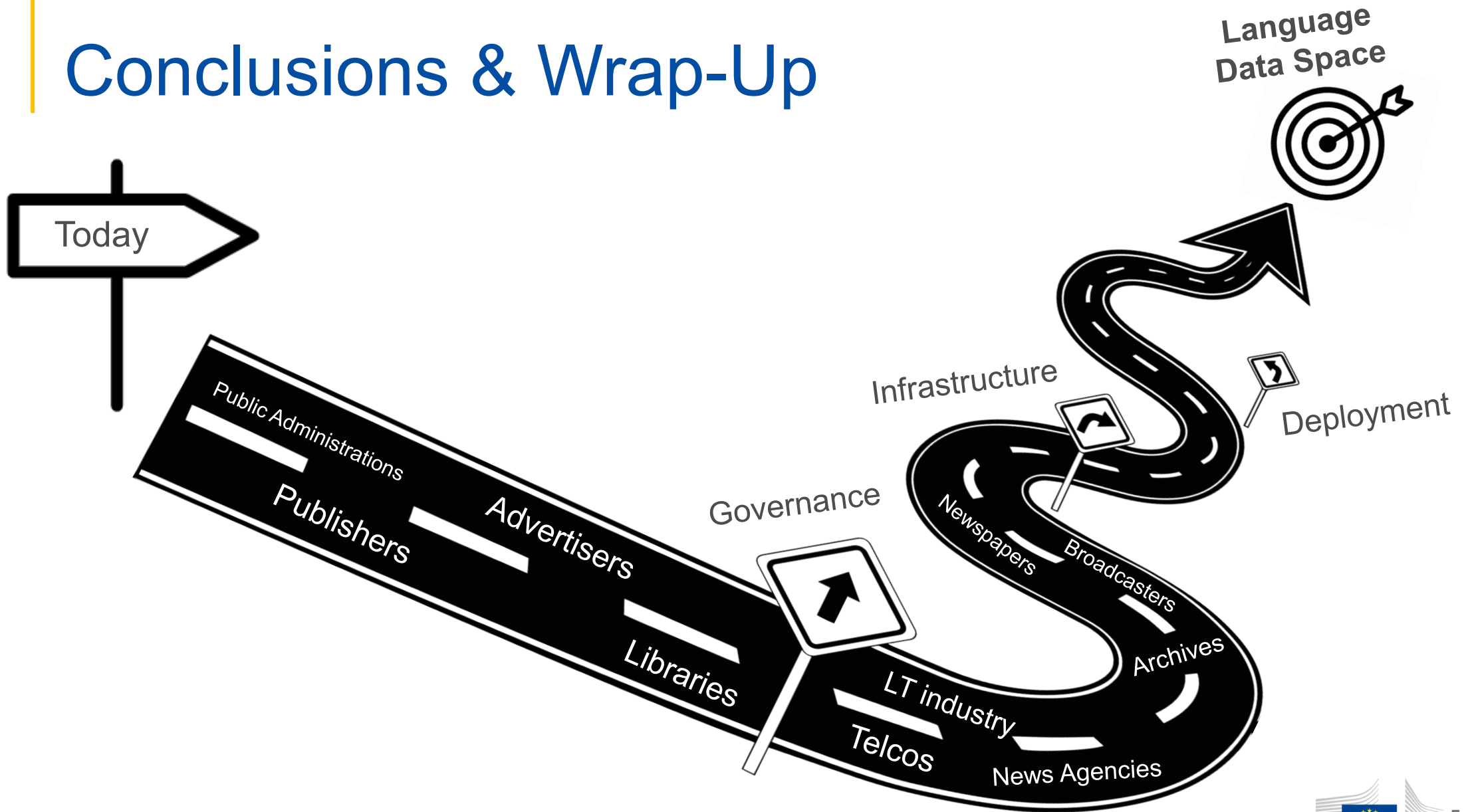
Benefits of the LDS

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

- Easy access to data
- Flows/Streams of data
- More data available
- More trading of data
- Data effort monetisation
- Revalorisation of data

DATA

APPLICATIONS, TOOLS & SERVICES

- New data-driven products & turnkey solutions, e.g., multilingual search engines, Q&A handling, etc.
- Customised language technologies and services, e.g., anonymisation
- Business intelligence via data analytics, sentiment analysis, etc.
- Access to international markets through multilingual solutions

Overcoming barriers to data sharing and trading:
- legal → e.g., licensing (IPR), personal data protection (GDPR), consumers' legislation, etc.
- technical → e.g., exchange modules, transaction verification, etc.

OVERCOMING BARRIERS

NETWORKING

- Finding partners & collaboration opportunities
- Exchanging best practices & insights
- Receiving & giving feedback from/to other stakeholders

→ to put in place targeted improvements;
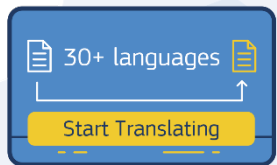→ to decrease time to market.

Conclusions & Wrap-Up

Thank you very much for sharing your ideas!

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/

eTranslation

European Commission