



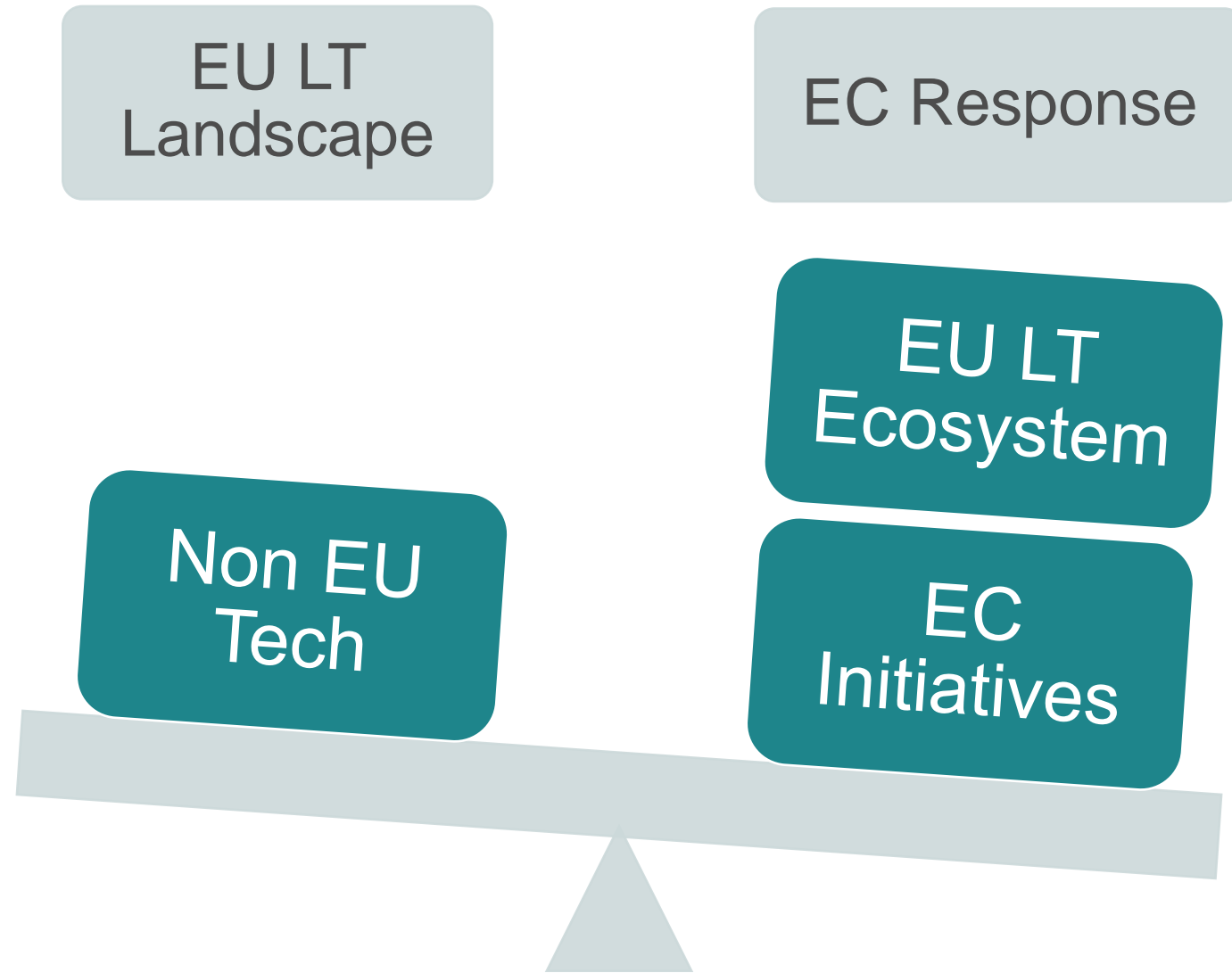
# The Digital Europe Programme and the Common European Language Data Space

LDS workshop in Finland

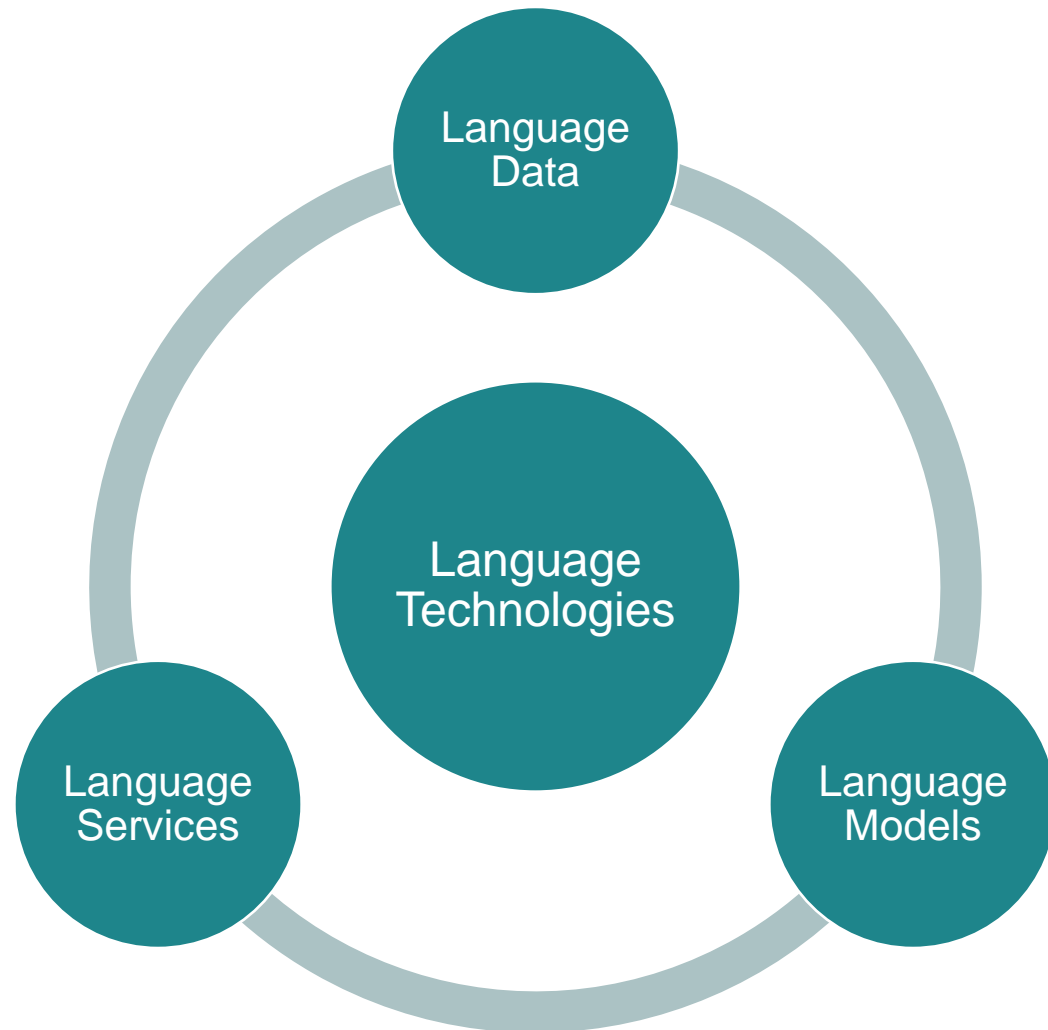
Unleashing the potential of data – for EU businesses and citizens

*Philippe Gelin*  
DG CONNECT – G3

# Background



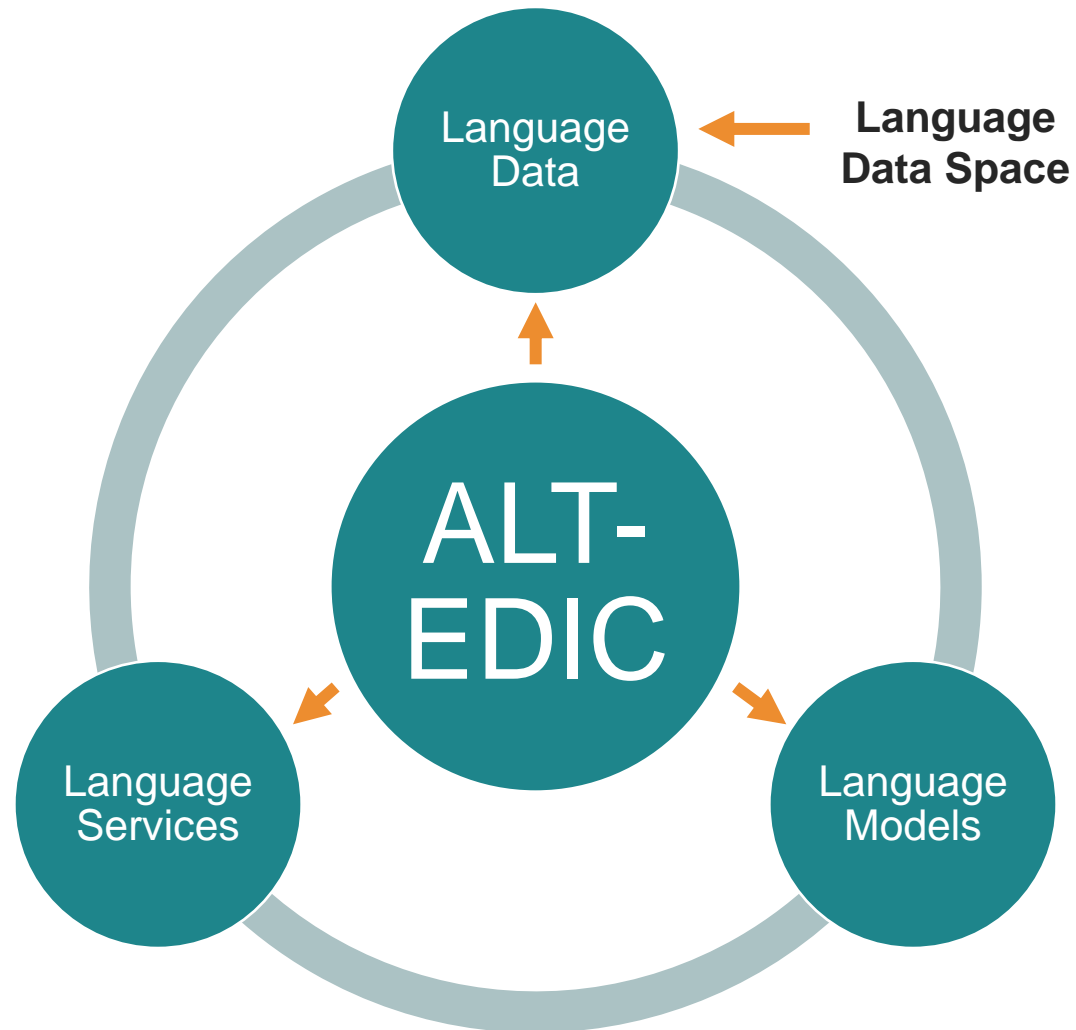
# The Language Technology Ecosystem



Language Technology ecosystem rests upon three fundamental dimensions:

- **Data,**
- **Models,**
- and **Services**

# Outline



**1**

Language Data Space

**2**

Alliance for Language  
Technology EDIC

**3**

DIGITAL Calls

# The Common European Language Data Space

# Language Data Space... What is it?

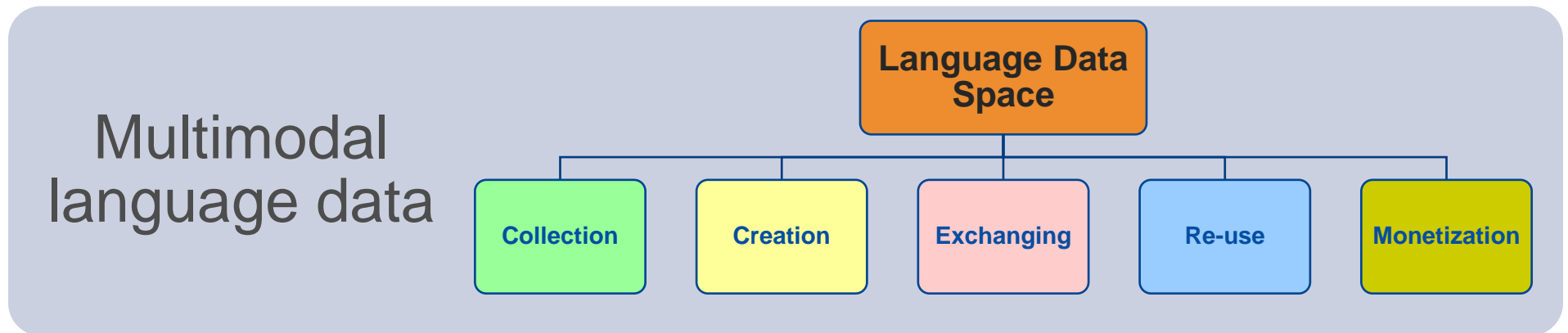
Goal

## Increase Data Exchange

Where we are

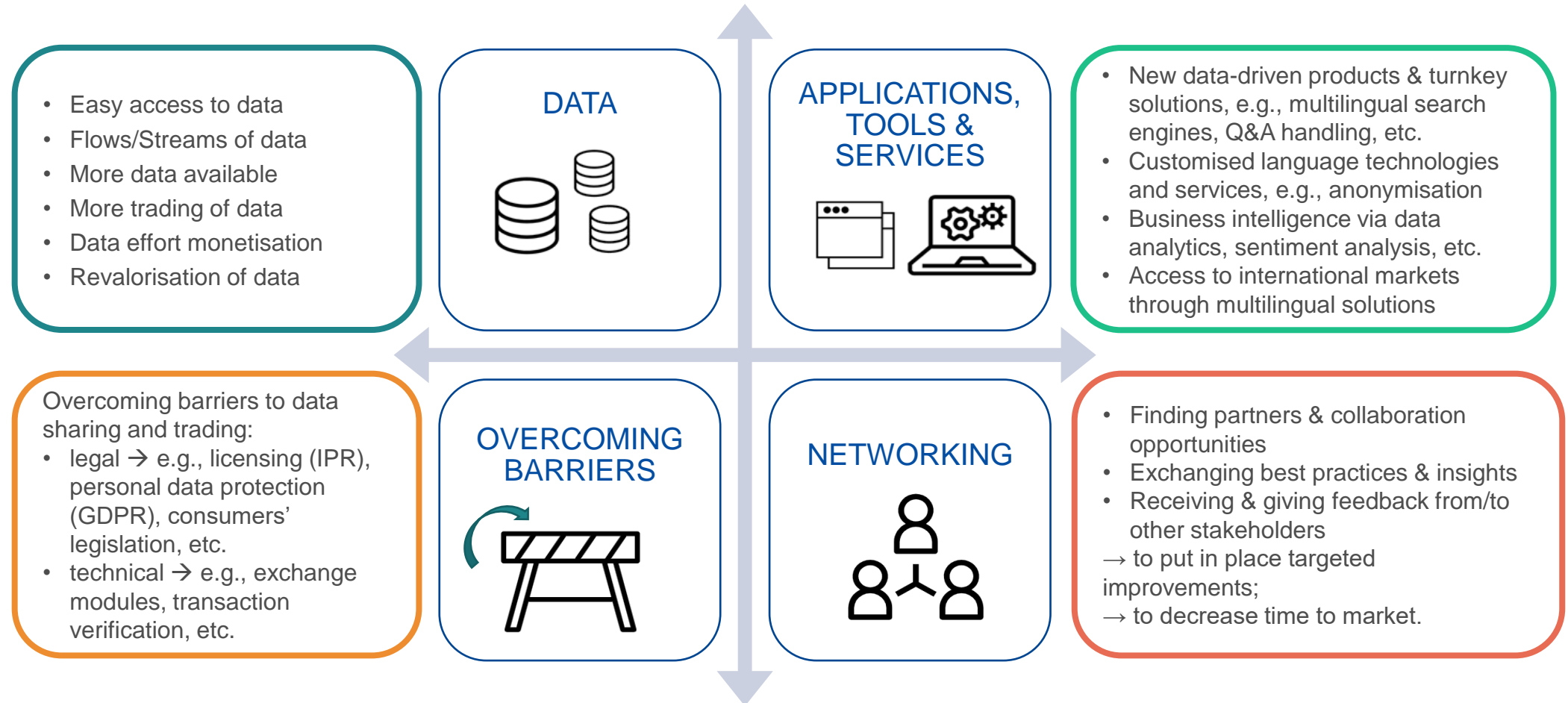


What we offer



# 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.*



# 'Origins' of the Common European Data Spaces

## Common European Data Spaces

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

are the centrepiece of the



Objectives



## European Strategy for Data

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.**



# Digital Programme

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*.

# Alliance for Language Technology EDIC (ALT-EDIC)

# Alliance for Language technologies EDIC

## What

[Multi-Country Project](#) proposed within the [Digital Decade Policy Program 2030](#)

Build a European ecosystem around Language Data and Technologies (Large Language Models, Translation, Chatbots, Speech Transcriptions, Subtitles, Text Analysis and Generation ...)

## Why

Federate data and resources from EU and Member States to boost competitiveness

Promote European technological leadership and strategic autonomy in the domain

## Who

**Members:** FR, BG, ES, GR, HR, IE, IT, LV, LT, NL, PL, SI

**Observers:** AT, DK, EE, MT, RO, SK–     **In process:** BE

**Interested:** HU, PT, CZ, DE, FI, SE

## How

~25 Mio € Member States contribution (**Financial Contribution:** € 12 Mio + **In-Kind contributions:** € 14 Mio)

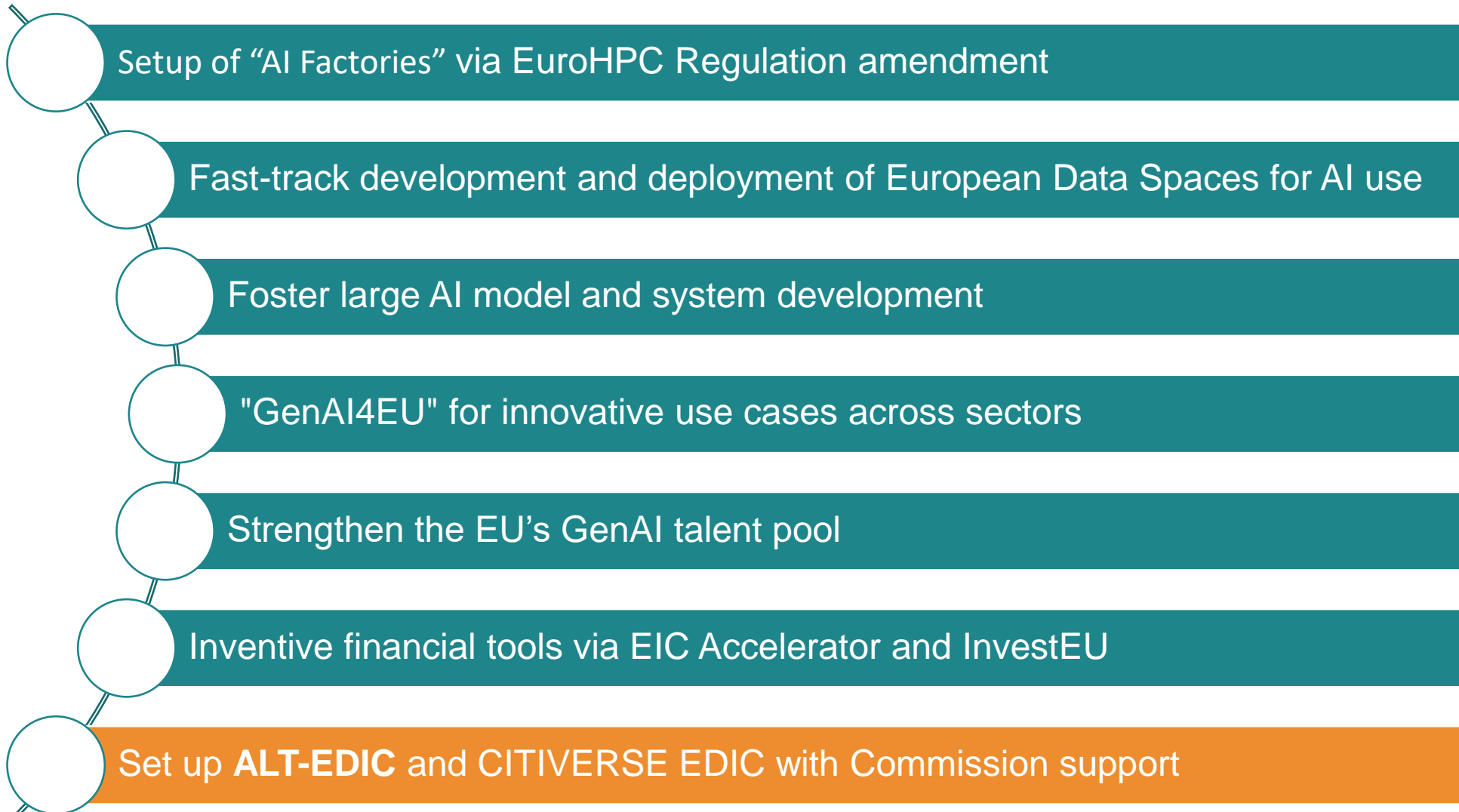
Potential EU funding (DIGITAL Programme)

## When

**7 February 2024:** [Commission Implementing Decision \(EU\) 2024/458](#) on setting up ALT-EDIC

**24 January 2024:** ALT-EDIC announced as part of **the AI Package**

# AI Innovation Package



# Alliance for Language technologies EDIC

## Objectives



# Alliance for Language technologies EDIC

## Work Plan

### Data

- Support collecting high quality Language data
- Synergies with the Language Data Space and MS initiatives

### Existing Models

- Gather existing European Language Models
- Fine-tune existing models for SMEs

### Development of Models

- New Open source Large Language Models
- Coordination with European initiatives on HPC

### Ecosystem

- Build an ecosystem around Language Technologies
- Dissemination, startup incubator, ...

# ALT-EDIC and Finland

- **Finland is not yet a member or observer of ALT-EDIC**

## Relevant initiatives in Finland:

- **Language data:** Language Bank of Finland, OPUS-CAT, OPUS-MT, Donate Speech project
- **HPC support for Language Technologies: CSC – IT Center for Science, LUMI** (EuroHPC Joint Undertaking)
  - LT projects in the pilot phase projects and the LUMI Extreme Scale Call :
    - **VLLMN:** Very Large Language Models in the Nordics
    - **LumiNMT:** Multilingual neural machine translation
    - **FinW2V2:** Pre-trained Finnish Speech Model
    - **MMATT:** Large-scale multi-modal data analysis with attention models
  - **HPLT Project** (High-Performance Language Technologies) where European LLMs will be trained on on one of the world's largest text collections using **LUMI**
- **SITRA**
  - “Large language models and European needs” is one of the focus area that have been recently discussed with the EC

# Digital Europe Programme

Work Programme 2024



# DIGITAL Work Programme 2024 (3 calls)

## DIGITAL-2024-AI-06-LANGUAGE-01

Alliance for Language Technologies:

### Data and Finetuning

€ 20 millions (Simple Grant)

- Support the Collection of Language Data
- Support Fine-tuning of Large Language Models

## DIGITAL-2024-AI-06-LANGUAGE-02

Alliance for Language Technologies:

### Ecosystem

€ 4 millions (CSA)

Support the Coordination of a European Language Technologies Ecosystem.

## DIGITAL-2024-AI-06-FINETUNE

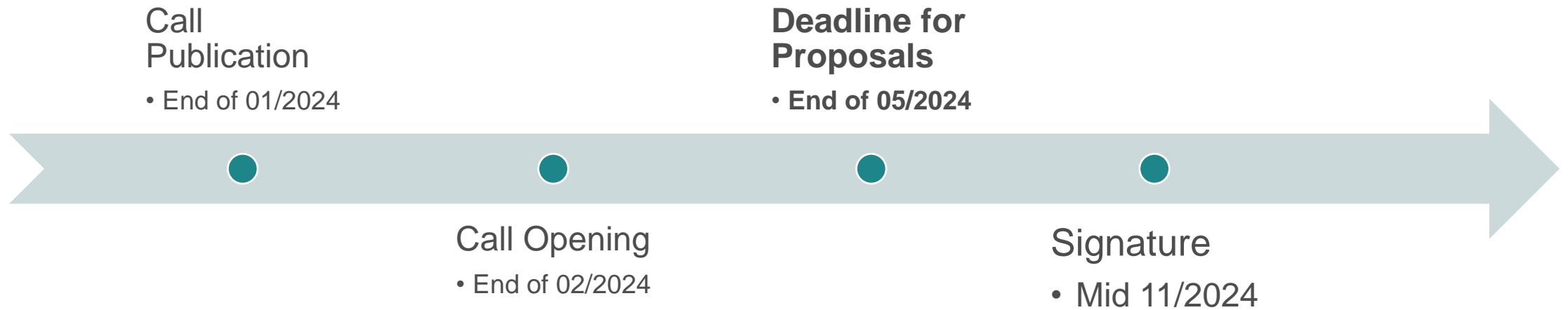
Open-source European foundation model

€ 25 millions (SME Grant)

Develop and make available one open-source large language foundation model as an infrastructure designed to be largely used by public or private users.

# DIGITAL Work Programme 2024

## Timeline



# Thank you



© European Union 2023

Unless otherwise noted the reuse of this presentation is authorised under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.

Slide xx: [element concerned](#), source: [e.g. Fotolia.com](#); Slide xx: [element concerned](#), source: [e.g. iStock.com](#)