

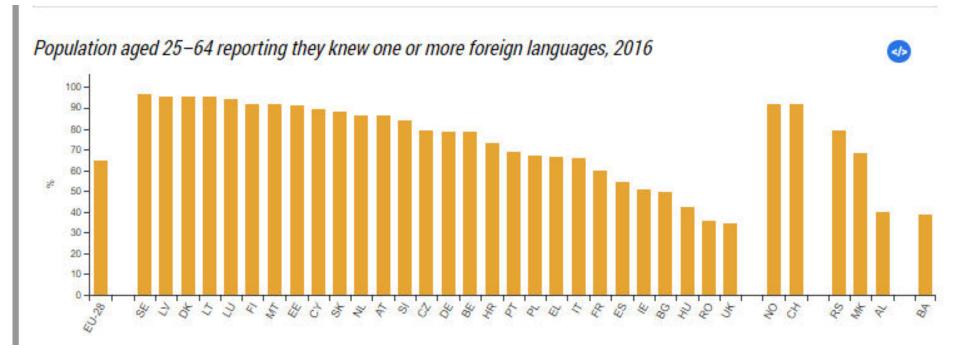
Philippe Gelin

Head of Sector Multilingualism

Directorate-General for Communications Networks, Content and Technology **European Commission**

My Europe, My Language

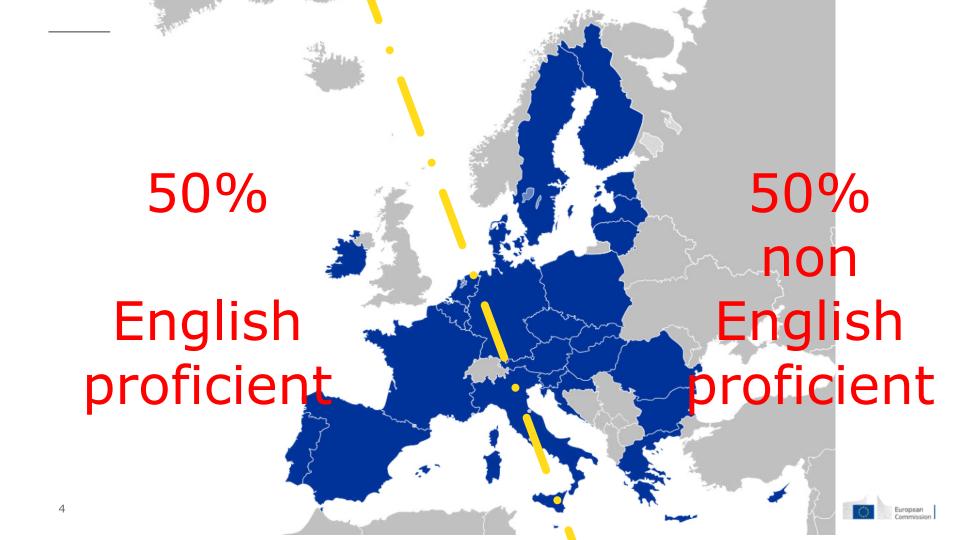
Virtual Berlin, 15th November 2021 - 9:15 - 9:40

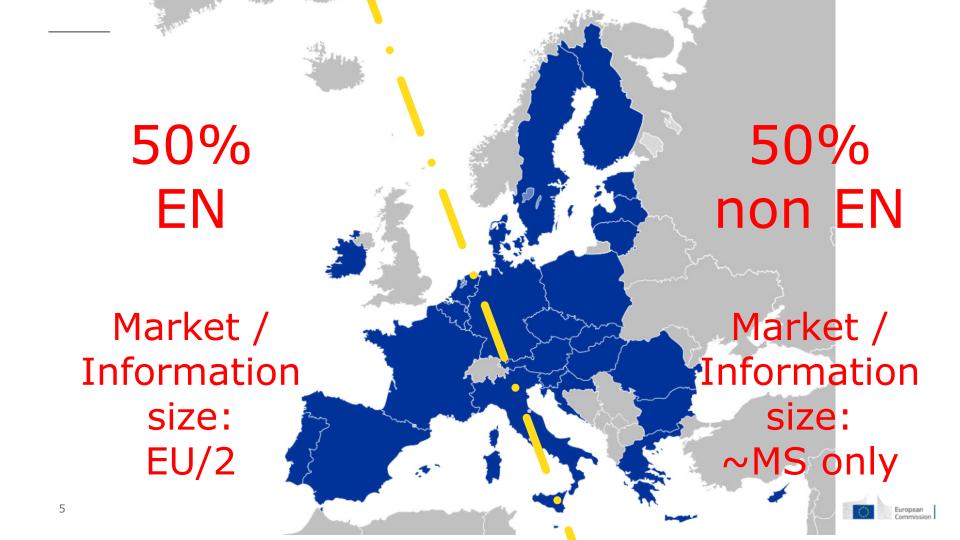


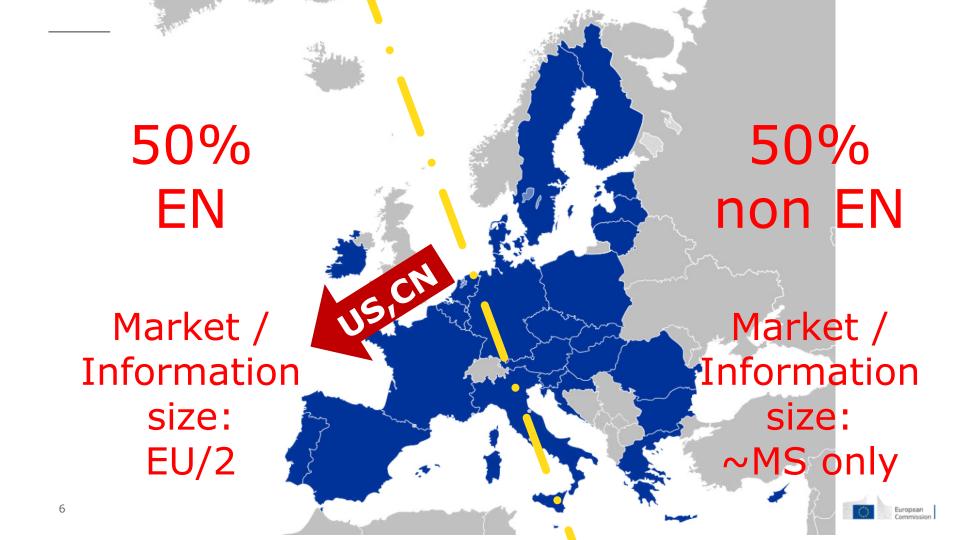


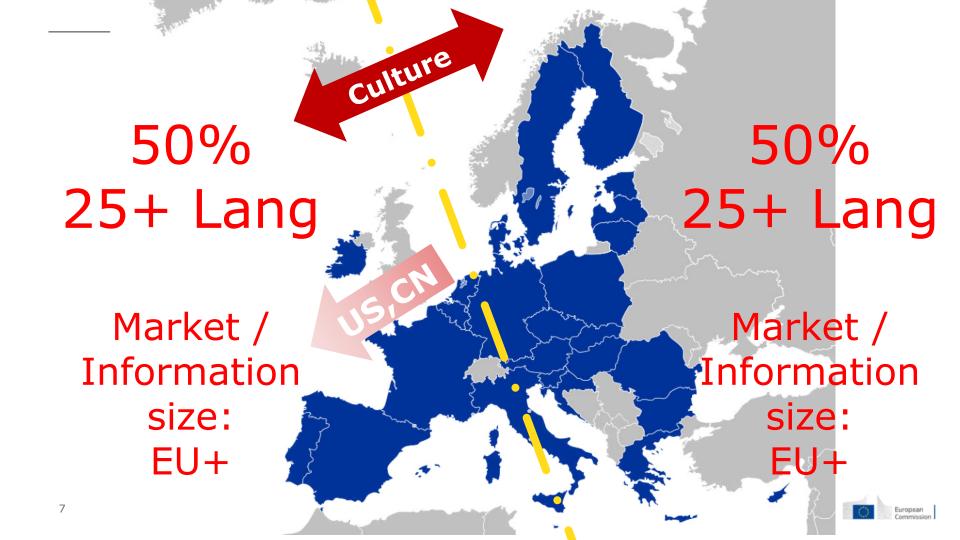
Which are the most studied foreign languages? (% of pupils at lower secondary level) O EU 97.3% English 33.8 % French 23.1% German Spanish 13.6% 2.7% Russian 1.1% Italian Data from 2015. Hello Hallo Salut



















HIGH PERFORMANCE COMPUTING





ADVANCED DIGITAL SKILLS



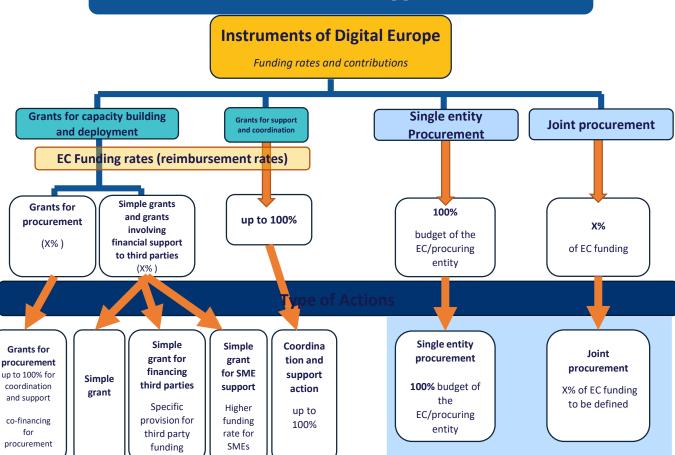
CLOUD, DATA AND AI



BEST USE OF DIGITAL TECHNOLOGIES



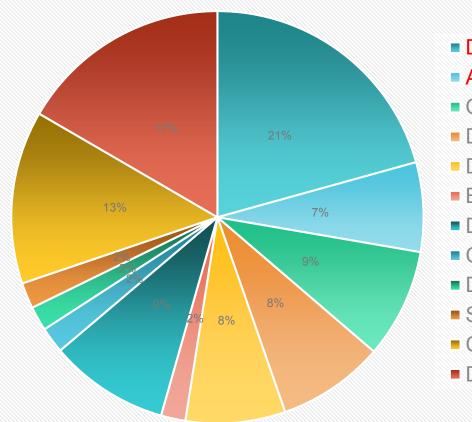
DEP Instruments and Type of Actions





Status of December 2019

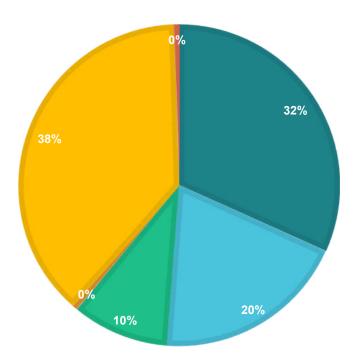
DIGITAL Work Programme 2021-2022 (1,98 Bio)



- Data spaces & Cloud-to-edge
- Al work-strand
- Quantum communication
- Digital skills
- Destination Earth
- Blockchain
- Digital Government
- Confidence
- Digital Technologies
- Support actions
- Cybersecurity
- Digital Innovation Hubs

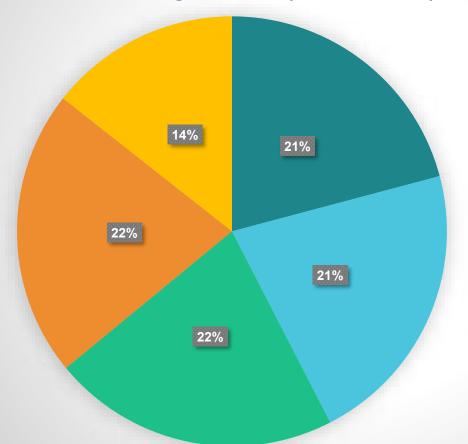
Cloud to Edge (205 Mio)

■ Smart middleware ■ Large Scale Pilot ■ Market Place ■ Alliance C2E ■ TEF - Edge Al ■ Alliance processor

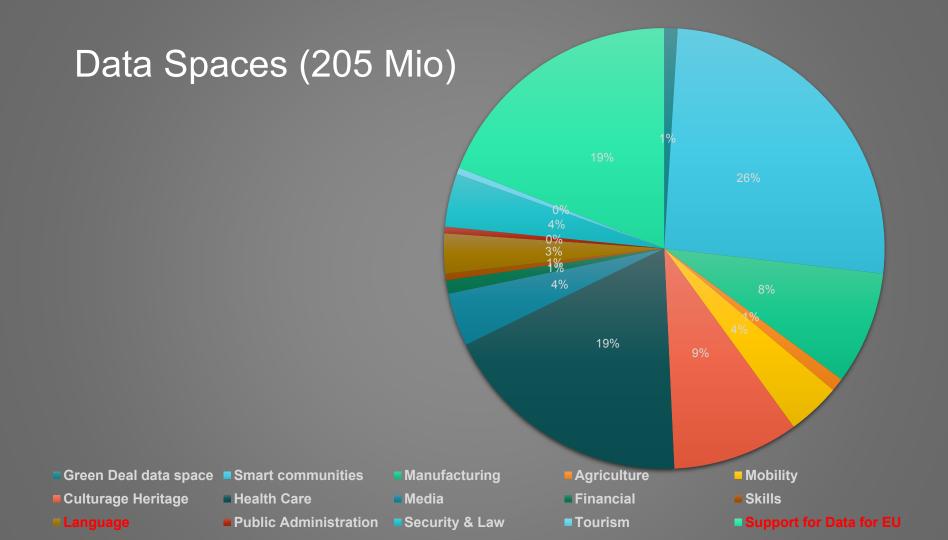




Artificial Intelligence (205 Mio)



- Al on demand
- Al TEF Manufacturing
- AI TEF Health
- AI TEF Agri-food
- Al TEF Smart Cities





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 Al language technologies services to be offered through the Al platform.

Two Strands:

- Establish an institutional Centre of Excellence for Language Technologies (CELT)
- Deployment of the language data space



1st work strand

 Establish an institutional Centre of Excellence for Language Technologies (CELT) to coordinate across the Member States 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.



1st Work Strand - Detail

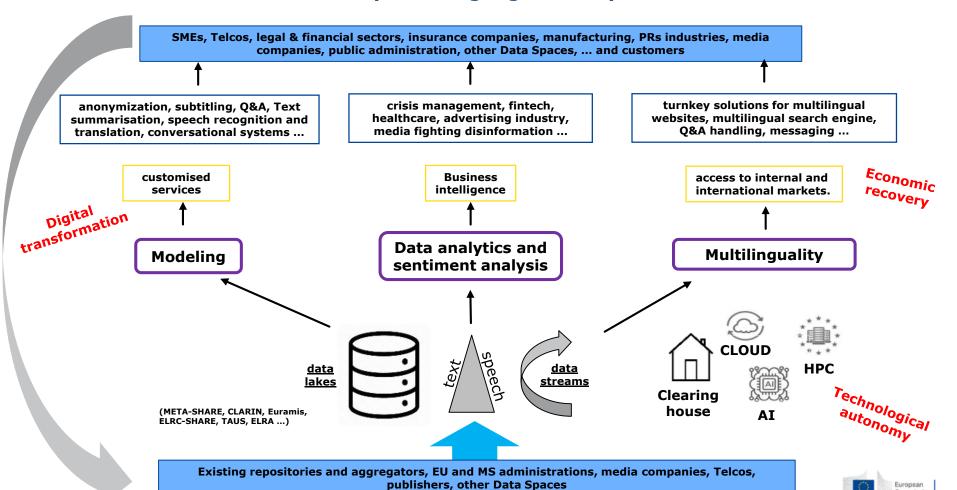
- Develop in close collaboration with the Member States a multi-stakeholder data and services governance scheme, bringing together large industrial entities, public stakeholders and small-and-mid-size enterprise stakeholders.
- Identify and bring together existing European stakeholders and initiatives.
- 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.

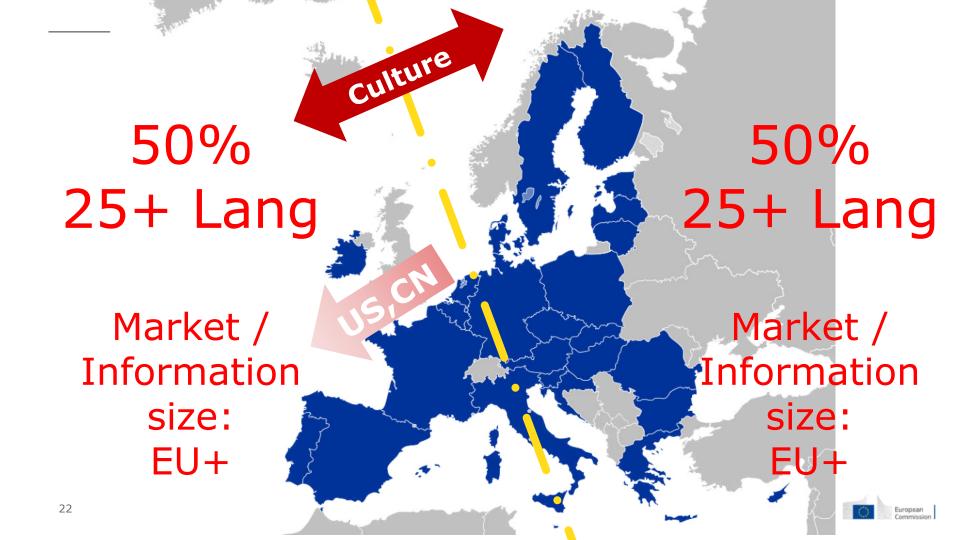


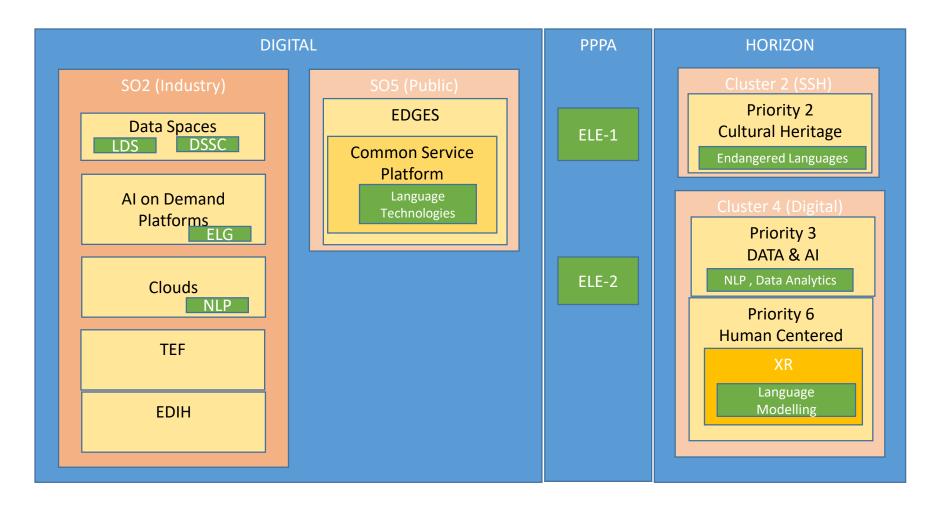
2nd work strand

- deploying the necessary infrastructure for the collection and sharing of multimodal language data and models;
- implementing the data & services governance, business models and strategies as defined by CELT;
- 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;
- deploying advanced Al-based language processing services and tools to be made available through the Al-on-demand platform.

The European Language Data Space







Links & contacts



PHILIPPE GELIN

European Commission

DG CONNECT - Communications Network, Content and Technology

Dir G: Data

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



Philippe.Gelin@ec.europa.eu

