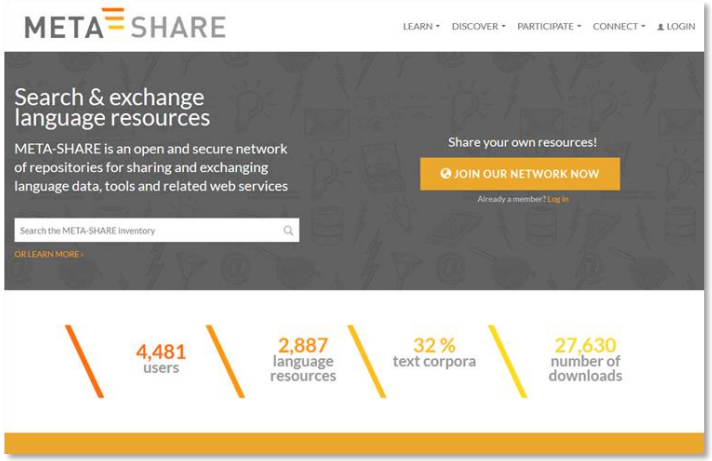
EUROPEAN
LANGUAGE
DATA SPACE

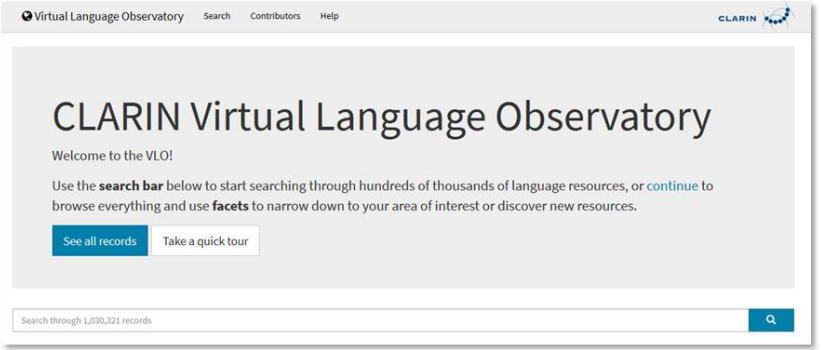
Common European Language Data Space: An Overview

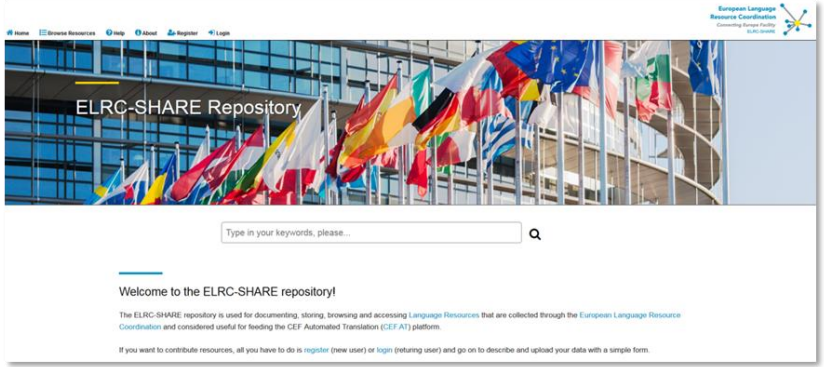
Stelios Piperidis (ILSP/Athena RC, Greece)
spip@athenarc.gr

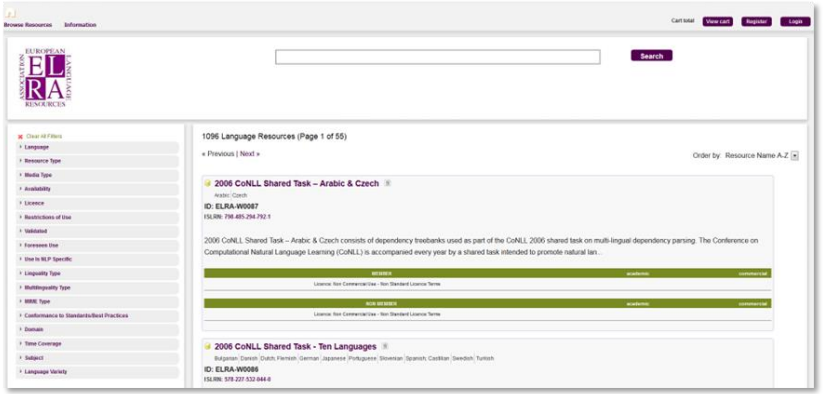
10-10-2023 Language Technologies-Beyond Digital Products, DG CONNECT University
https://language-data-space.ec.europa.eu

Existing Platforms/Infrastructures

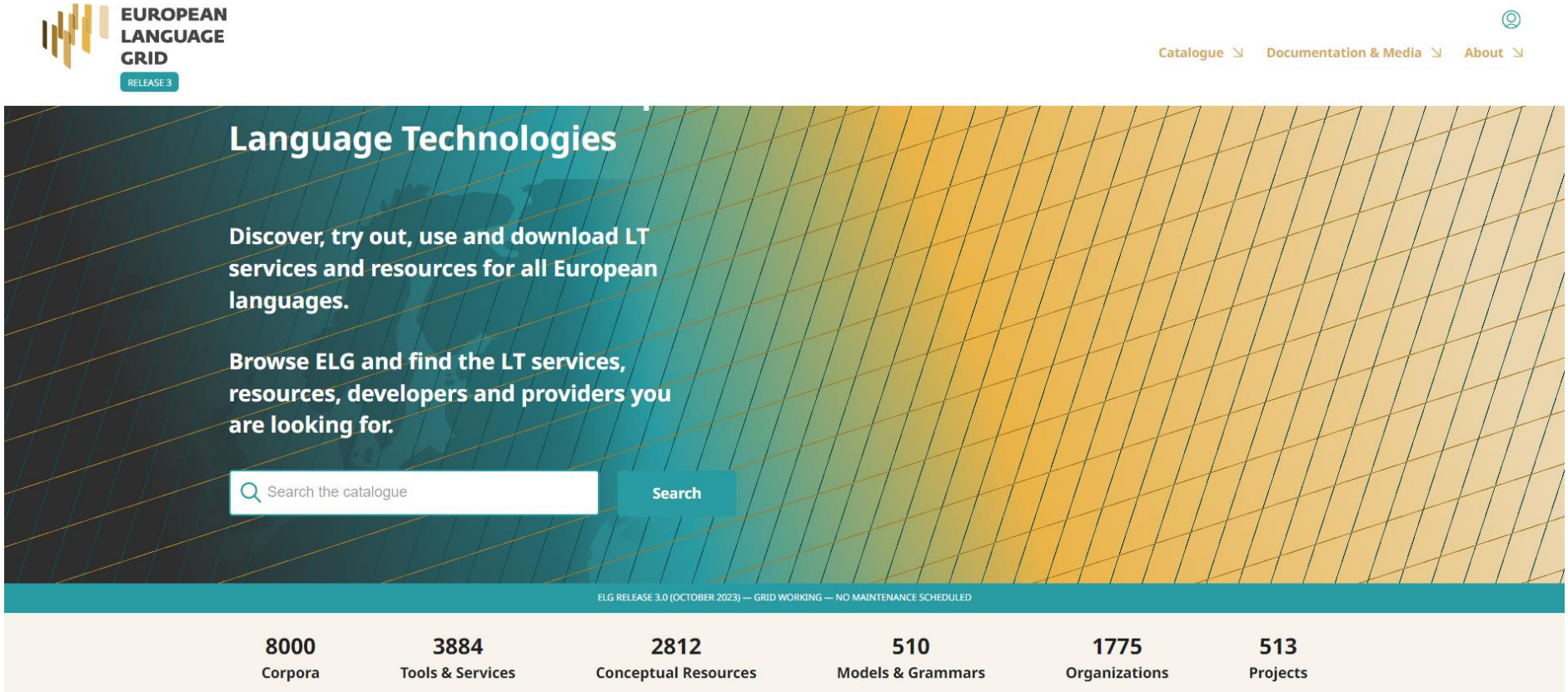








European Language Grid



Language Data Space

- Type of action: procurement (CNECT/LUX/2022/OP/0026)
- Budget: 6M€ (+ 2M€ if renewed)
- Runtime: 36 months (+ 12 months if renewed)
- Objective: Develop and deploy a European data space and marketplace for the collection, creation, sharing and re-use of multilingual and multimodal language data
- Salient features: governance framework, technical architecture and infrastructure, openness, promotion
- Stakeholders: **industry**, research, public administrations, cultural associations, NGOs and citizens

Consortium and Subcontractors

Lead Partner and Coordinator		
Deutsches Forschungszentrum für Künstliche Intelligenz GmbH	DFKI	DE
Partners and Operation Leads		
R.C. “Athena”, Institute for Language and Speech Processing	ILSP	GR
Evaluations and Language Resources Distribution Agency	ELDA	FR
TILDE	TILDE	LV
Main Subcontractors		
3pc GmbH Neue Kommunikation	3pc	DE
Capgemini Deutschland GmbH	CapG	DE
CLARIN ERIC	CLARIN	NL
Data, AI and Robotics AISBL	BDVA	BE

Plus legal experts (Delcade, France) and approx. 30 organisations for the logistics of multiple country workshops

EU Data Strategy & Data Spaces

- Data Spaces are an inherent part of the EU Data Strategy
- Various data economy and data infrastructure initiatives in Europe with slightly different goals and individual positioning but conceptual, technical, legal and operational overlap:
  - Data Spaces Business Alliance (DSBA): Gaia-X, IDSA, FIWARE, BDVA
  - EU: DSSC (incl. DSBA), SIMPL, approx. 20 data spaces (in preparation status)
- The Common European Language Data Space is one of the approx. 20 official EU data space projects – focus on industry

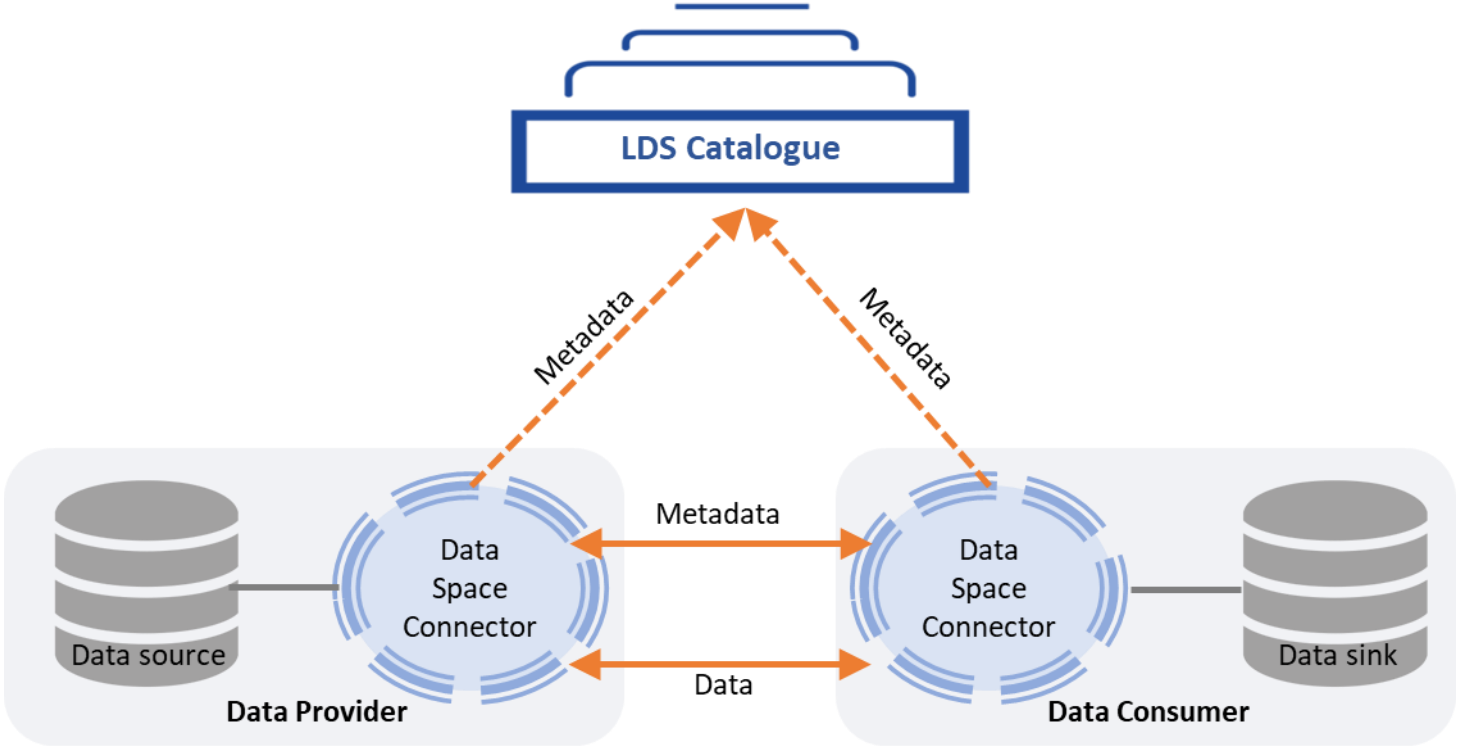
By language data we mean

- Text, audio, video - multimodal
- Language models, language descriptions
- Lexical and conceptual resources
- In different languages, monolingual/multilingual/parallel
- Different domains
- Different communication scenarios
- subjective, ambiguous as is language itself
- Scope of language data is evolving

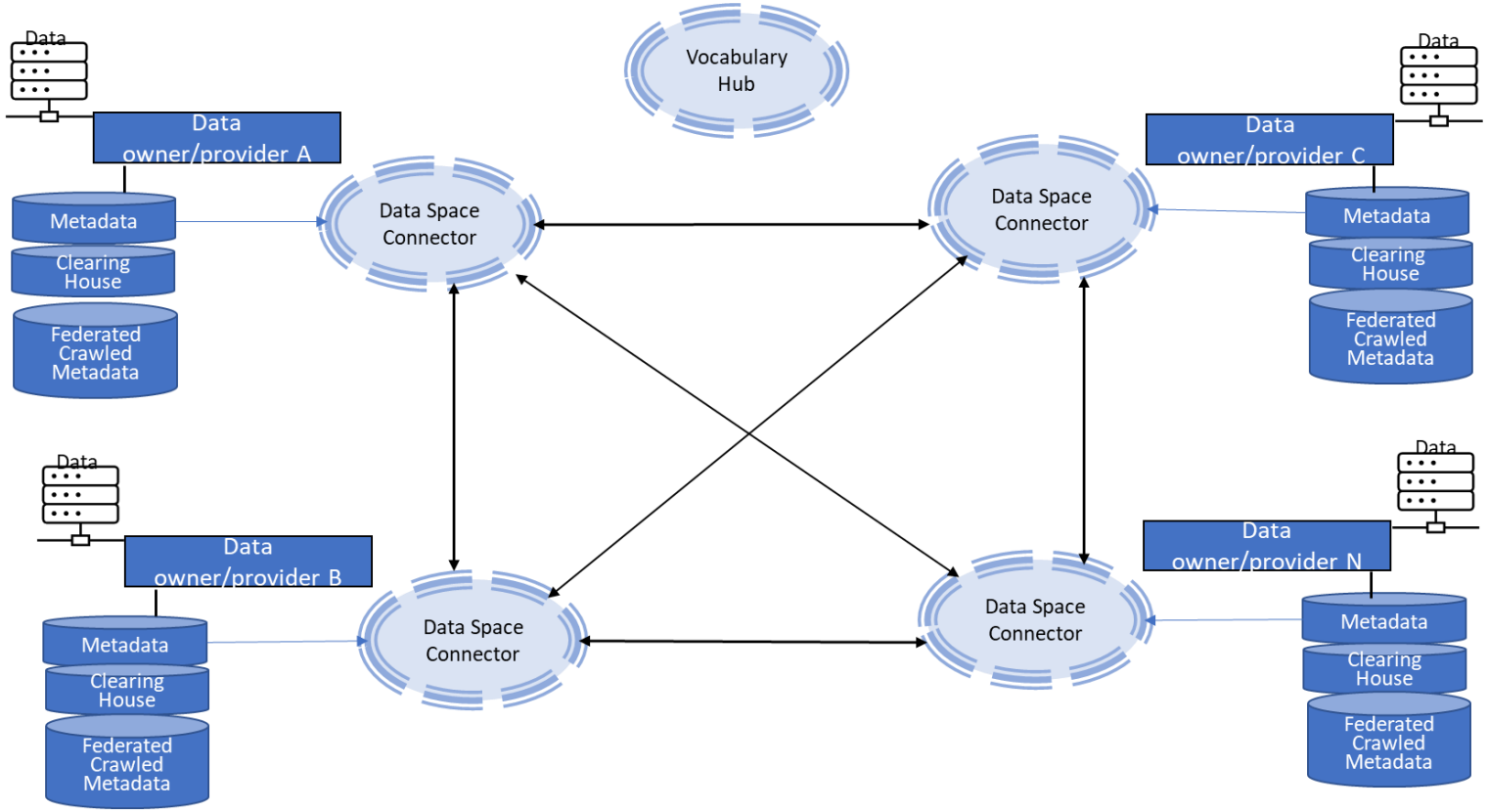
EU Data Strategy & Data Spaces

- Data **Sovereignty** is the ability of a natural or legal person to exclusively and sovereignly decide concerning the usage of data as an **economic** asset. (*IDS-RAM 4.0, ch. 2 Context*)
- Consequences for DSs design and implementation: DSs should allow for
  - data transactions, either monetary or not
  - exercising full control of data owner/provider over access to data, terms, pricing etc.
  - tracking data provenance and lineage
  - Technically enforcing policies, standards, rules, contracts
  - establishing trust between participants

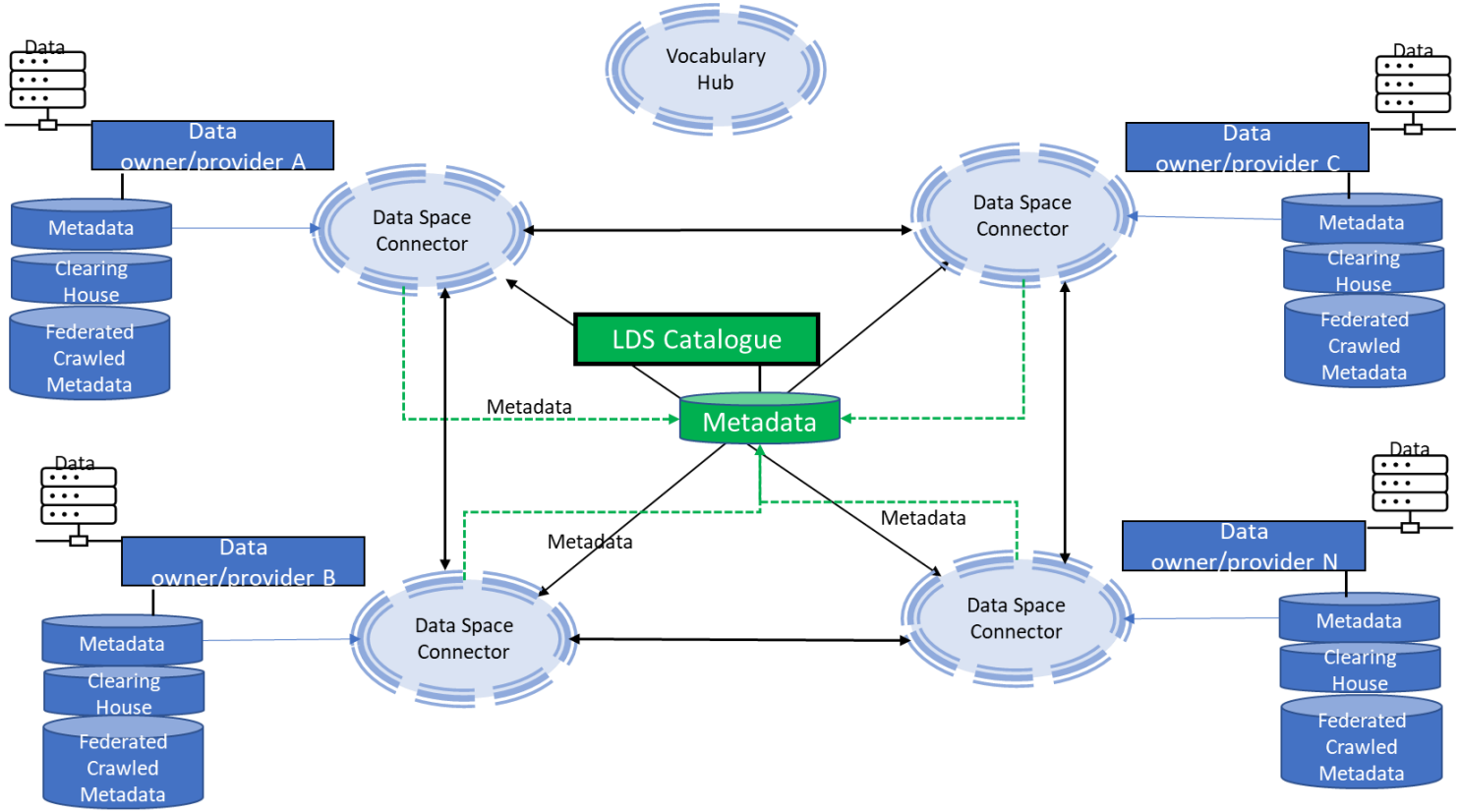
LDS Data Flows



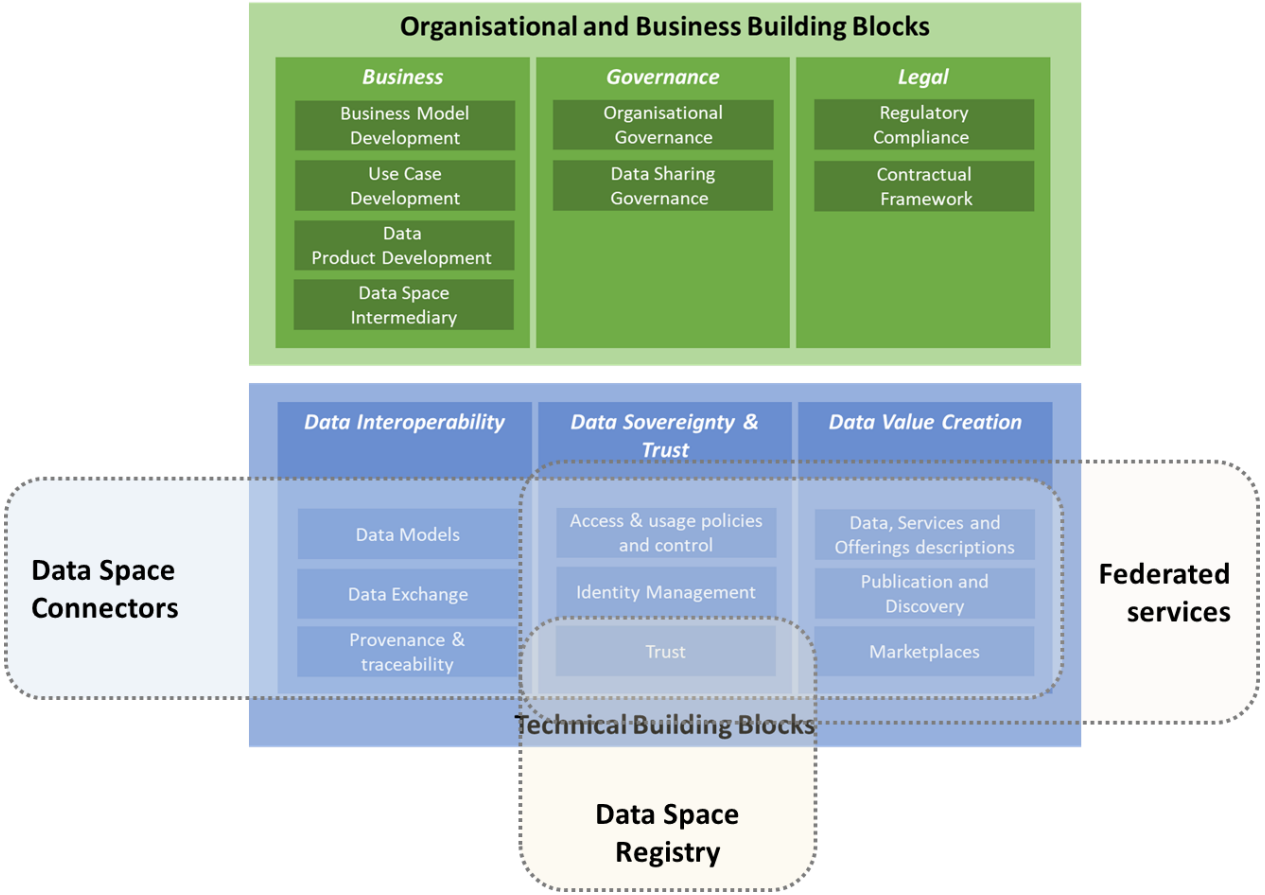
LDS Architecture



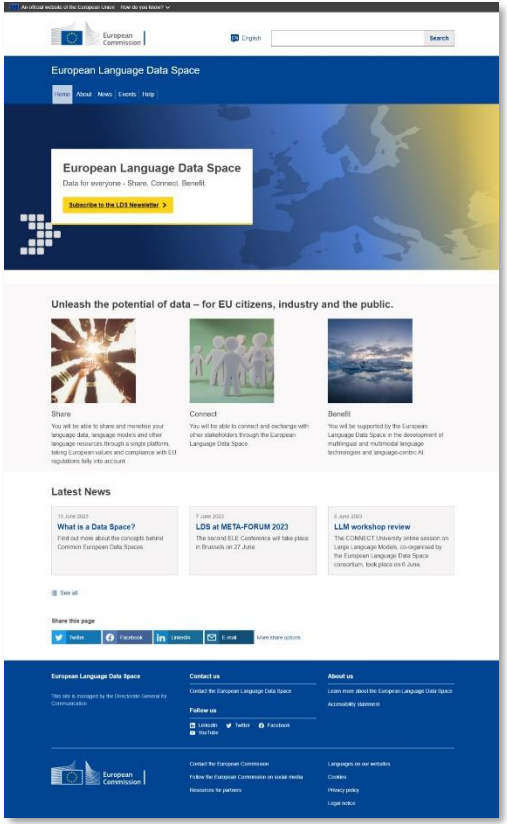
LDS Architecture

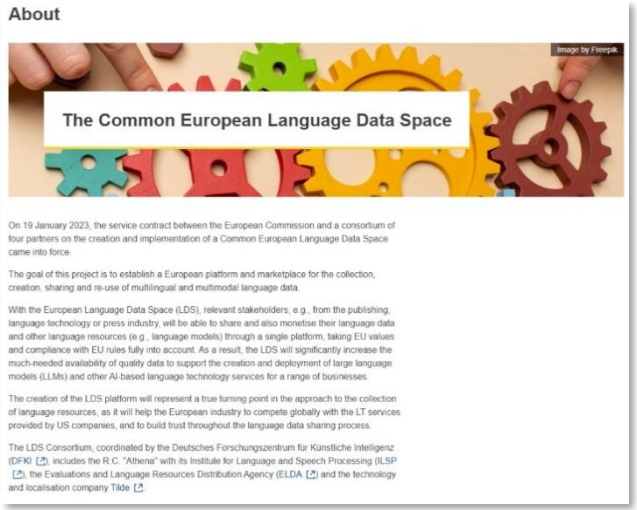


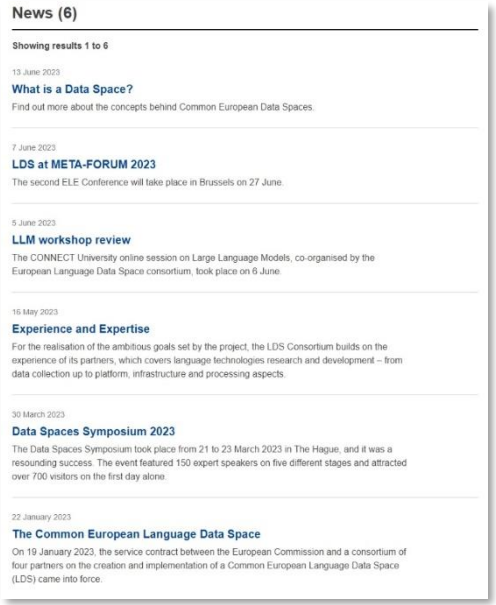
LDS/DS Building Blocks



Source: DSSC Data Spaces Blueprint, Version 0.5, September 2023







https://language-data-space.ec.europa.eu

Common European Language Data Space

Thank you!

A Common European Language Data Space – funded under contract LC-01936389 with the European Union.

Stelios Piperidis (ILSP/Athena RC, Greece)
spip@athenarc.gr

10-10-2023 Language Technologies-Beyond Digital Products, DG CONNECT University
https://language-data-space.ec.europa.eu