



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

META-SHARE LEARN • DISCOVER • PARTICIPATE • CONNECT • LOGIN

Search & exchange language resources

META-SHARE is an open and secure network of repositories for sharing and exchanging language data, tools and related web services

Share your own resources!

[JOIN OUR NETWORK NOW](#)

Already a member? [Login](#)

Search the META-SHARE Inventory

[OR LEARN MORE](#)

4,481 users | 2,887 language resources | 32% text corpora | 27,630 number of downloads

Virtual Language Observatory Search Contributors Help **CLARIN**

CLARIN Virtual Language Observatory

Welcome to the VLO!

Use the **search bar** below to start searching through hundreds of thousands of language resources, or **continue** to browse everything and use **facts** to narrow down to your area of interest or discover new resources.

[See all records](#) [Take a quick tour](#)

Search through 1,030,321 records

Home ELRC-SHARE Repository Help About Register Login

ELRC-SHARE Repository

Type in your keywords, please...

Welcome to the ELRC-SHARE repository!

The ELRC-SHARE repository is used for documenting, storing, browsing and accessing Language Resources that are collected through the European Language Resource Coordination and considered useful for feeding the CEF Automated Translation (CEF-AT) platform.

If you want to contribute resources, all you have to do is [register](#) (new user) or [login](#) (existing user) and go on to describe and upload your data with a simple form.

European Language Resource Association

1096 Language Resources (Page 1 of 55)

2006 CoNLL Shared Task – Arabic & Czech

Arabic-Czech
ID: ELRA-W0087
URL: P8-85-294.792.1

2009 CoNLL Shared Task – Arabic & Czech consists of dependency treebanks used as part of the CoNLL 2009 shared task on multi-lingual dependency parsing. The Conference on Computational Natural Language Learning (CoNLL) is accompanied every year by a shared task intended to promote natural lan.

2009 CoNLL Shared Task - Ten Languages

Bulgarian, Danish, Dutch, Finnish, German, Japanese, Portuguese, Slovakian, Spanish, Catalan, Swedish, Turkish

ID: ELRA-W0086
URL: P8-207-333.884.6



Language Technologies

Discover, try out, use and download LT services and resources for all European languages.

Browse ELG and find the LT services, resources, developers and providers you are looking for.

ELG RELEASE 3.0 (OCTOBER 2023) — GRID WORKING — NO MAINTENANCE SCHEDULED

8000
Corpora

3884
Tools & Services

2812
Conceptual Resources

510
Models & Grammars

1775
Organizations

513
Projects

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

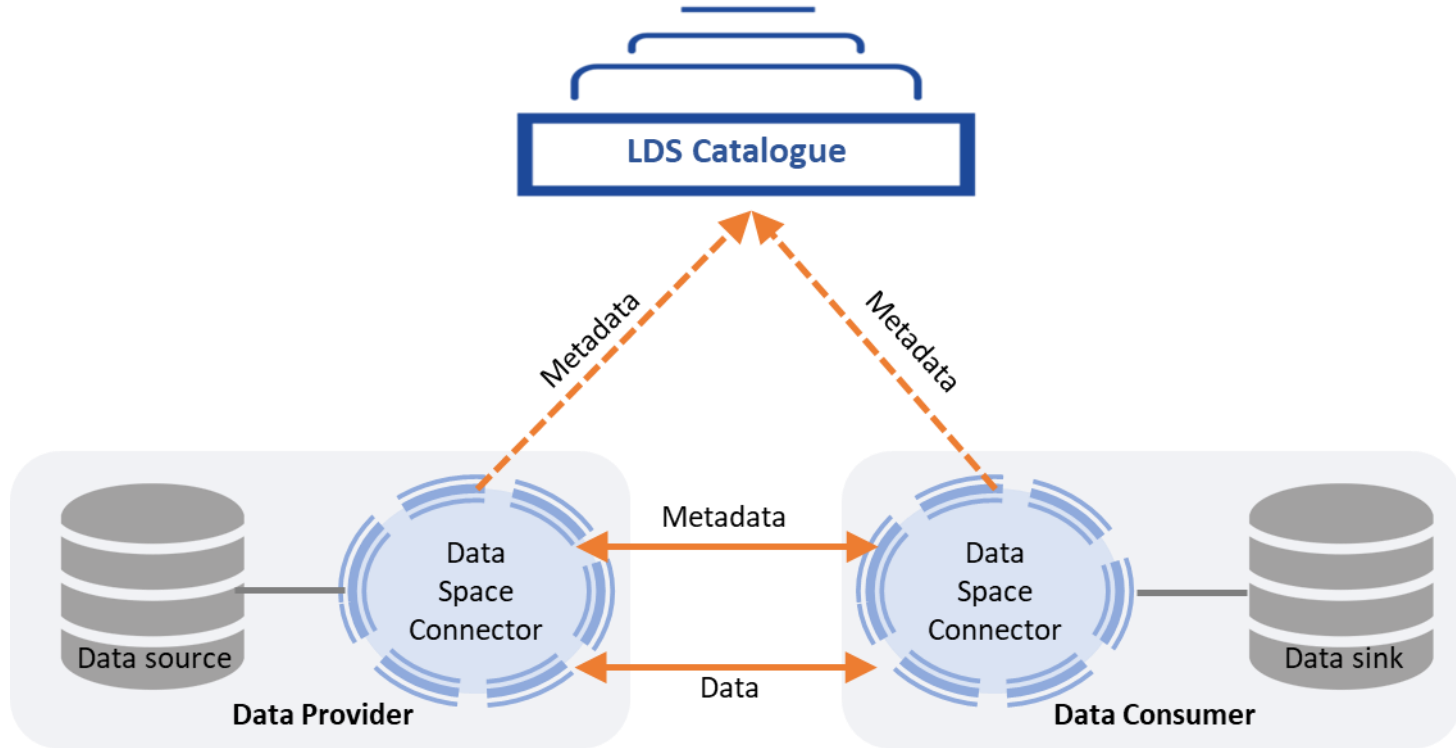
By language data we mean

- 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

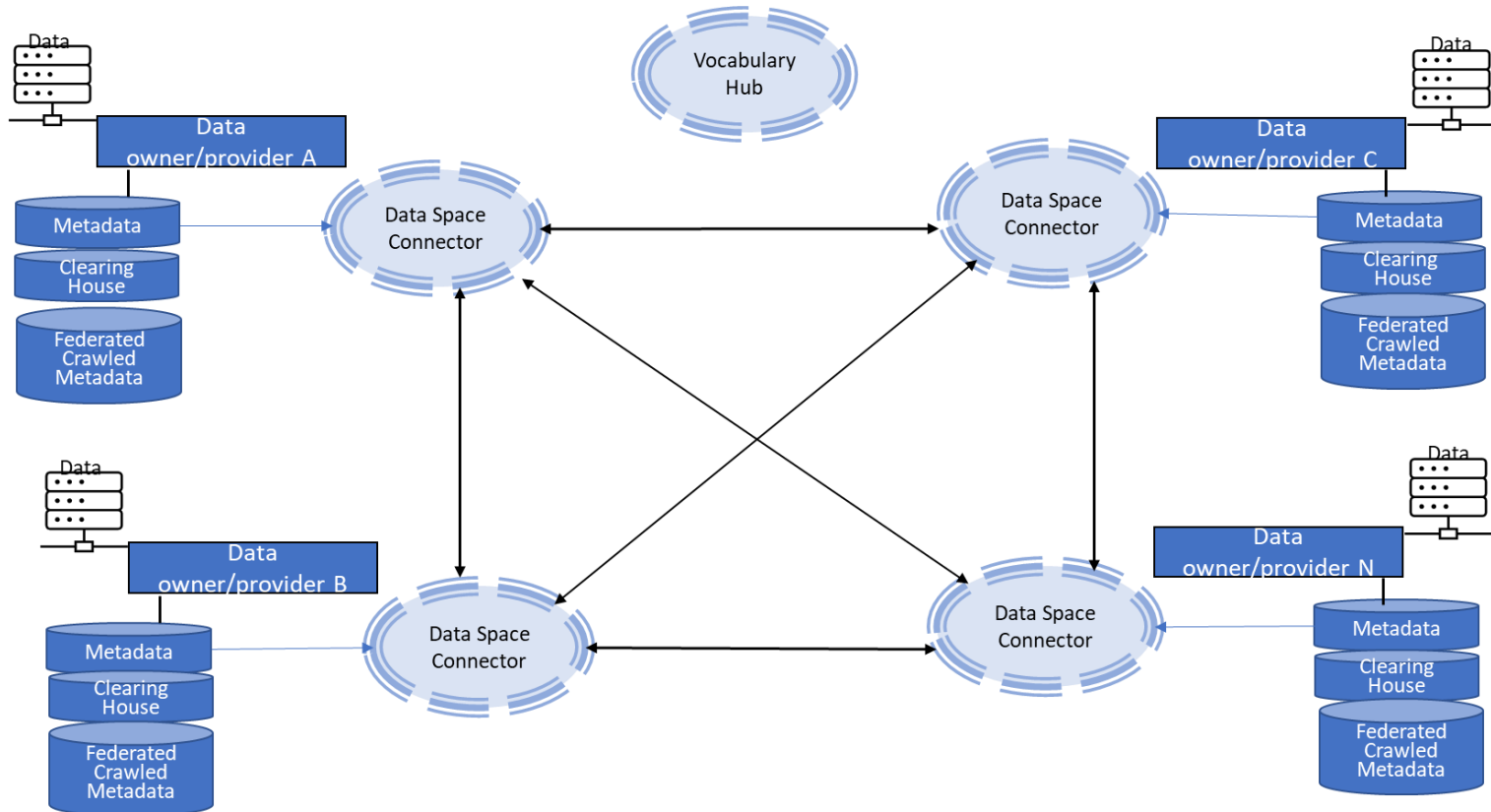
EU Data Strategy & Data Spaces

- 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

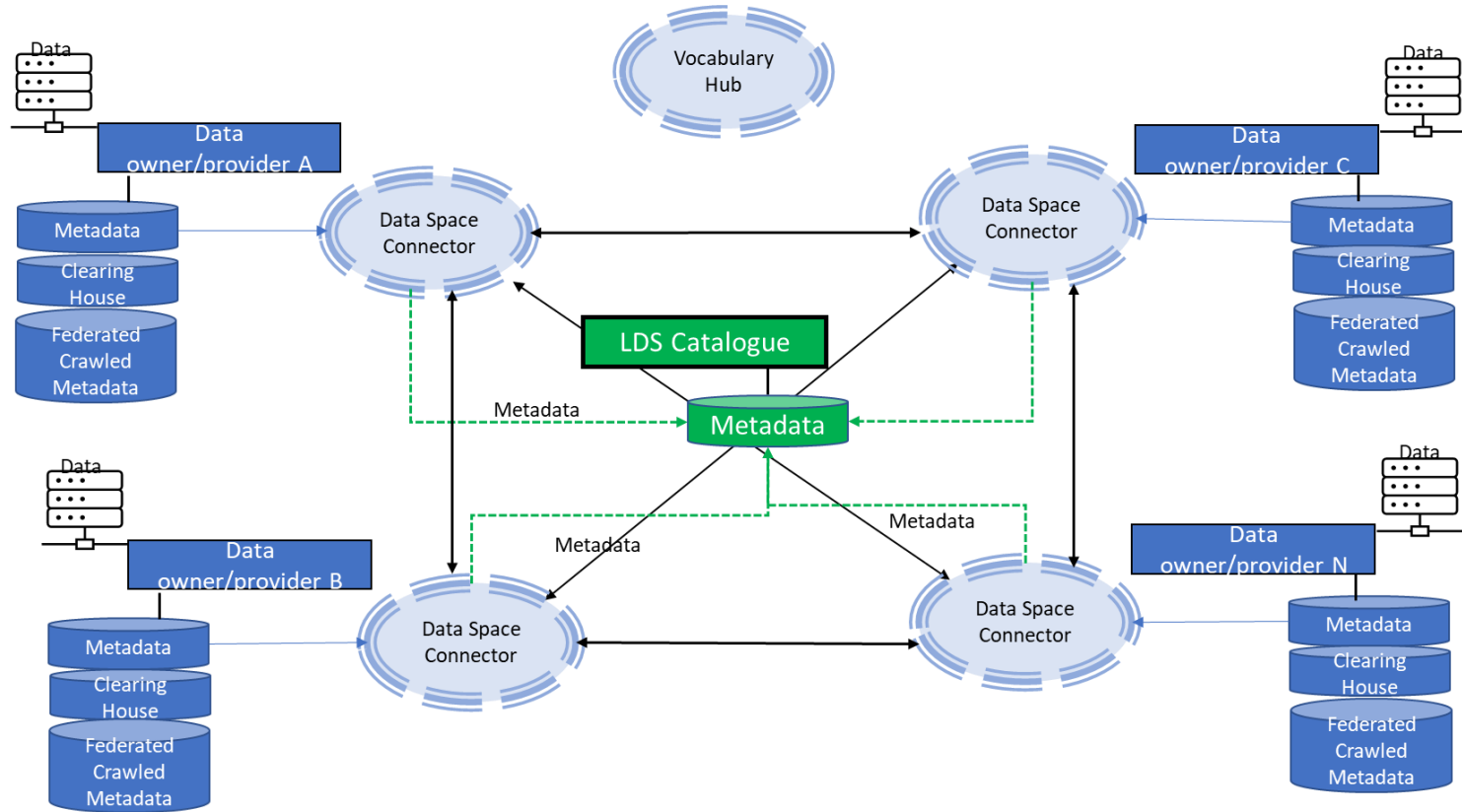
LDS Data Flows



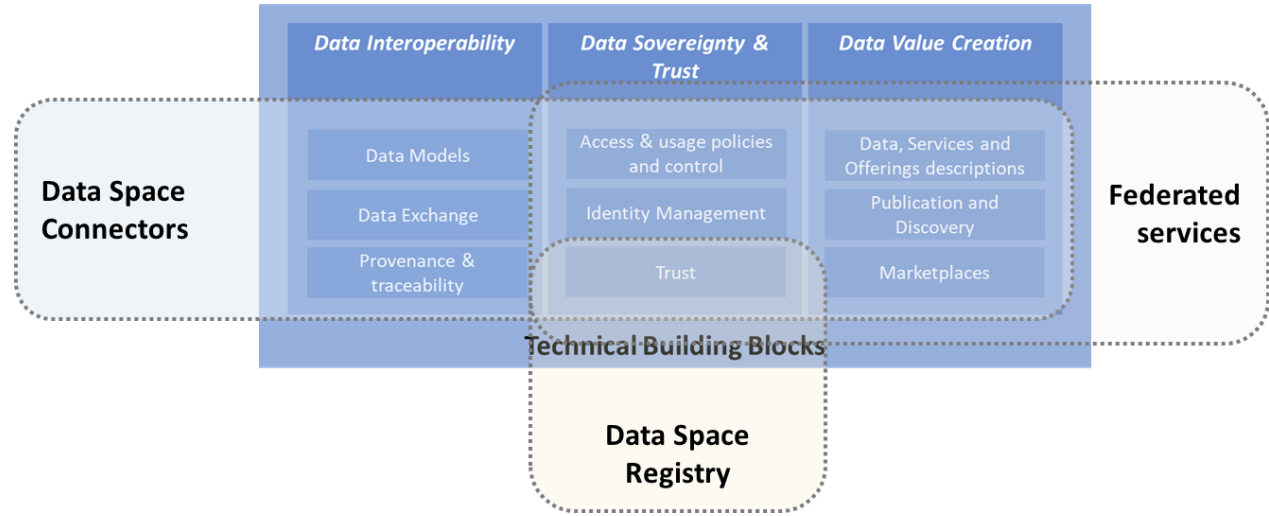
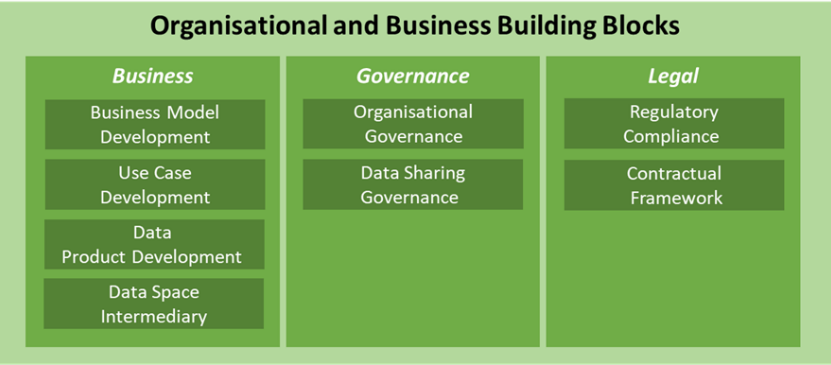
LDS Architecture



LDS Architecture



LDS/DS Building Blocks



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

The screenshot shows the homepage of the European Language Data Space. At the top, there is a navigation bar with the European Commission logo, the text 'European Language Data Space', and a search bar. Below this is a main banner with the title 'European Language Data Space' and the tagline 'Data for everyone - Share. Connect. Benefit.' A yellow button invites users to 'Subscribe to the LDS Newsletter'. The main content area features three columns with icons and text: 'Share' (sunburst icon), 'Connect' (group of people icon), and 'Benefit' (clouds icon). Below this is a 'Latest News' section with three articles: 'What is a Data Space?' (13 June 2023), 'LDS at META-FORUM 2023' (7 June 2023), and 'LLM workshop review' (5 June 2023). At the bottom, there is a footer with contact information, social media links, and logos for the European Commission and the LDS.

About



On 19 January 2023, the service contract between the European Commission and a consortium of four partners on the creation and implementation of a Common European Language Data Space came into force.

The goal of this project is to establish a European platform and marketplace for the collection, creation, sharing and re-use of multilingual and multimodal language data.

With the European Language Data Space (LDS), relevant stakeholders, e.g., from the publishing, language technology or press industry, will be able to share and also monetise their language data and other language resources (e.g., language models) through a single platform, taking EU values and compliance with EU rules fully into account. As a result, the LDS will significantly increase the much-needed availability of quality data to support the creation and deployment of large language models (LLMs) and other AI-based language technology services for a range of businesses.

The creation of the LDS platform will represent a true turning point in the approach to the collection of language resources, as it will help the European industry to compete globally with the LT services provided by US companies, and to build trust throughout the language data sharing process.

The LDS Consortium, coordinated by the Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) [\[2\]](#), includes the R.C. "Athens" with its Institute for Language and Speech Processing (ILSP) [\[3\]](#), the Evaluations and Language Resources Distribution Agency (ELDA) [\[4\]](#) and the technology and localisation company Tilde [\[5\]](#).

News (6)

Showing results 1 to 6

13 June 2023

What is a Data Space?

Find out more about the concepts behind Common European Data Spaces.

7 June 2023

LDS at META-FORUM 2023

The second ELE Conference will take place in Brussels on 27 June.

5 June 2023

LLM workshop review

The CONNECT University online session on Large Language Models, co-organised by the European Language Data Space consortium, took place on 6 June.

16 May 2023

Experience and Expertise

For the realisation of the ambitious goals set by the project, the LDS Consortium builds on the experience of its partners, which covers language technologies research and development – from data collection up to platform, infrastructure and processing aspects.

30 March 2023

Data Spaces Symposium 2023

The Data Spaces Symposium took place from 21 to 23 March 2023 in The Hague, and it was a resounding success. The event featured 150 expert speakers on five different stages and attracted over 700 visitors on the first day alone.

22 January 2023

The Common European Language Data Space

On 19 January 2023, the service contract between the European Commission and a consortium of four partners on the creation and implementation of a Common European Language Data Space (LDS) came into force.

<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>