



EUROPEAN LANGUAGE DATA SPACE



Common European Language Data Space The LDS Workshop in Luxembourg

Dimitra Anastasiou (LIST, Luxembourg) - Moderator
Dimitra.anastasiou@list.lu

19-06-2024 LDS Country Workshop, Luxembourg
<https://language-data-space.ec.europa.eu>

THE LUXEMBOURG INSTITUTE OF SCIENCE AND TECHNOLOGY
A REFERENCE IN RESEARCH AND INNOVATION FOR A DIGITALIZED, RESILIENT & SUSTAINABLE SOCIETY

Created in 2015

Active in 4 fields

ENVIRONMENT

+
200
employees

INFORMATICS

+
100
employees

MATERIALS

+
200
employees

SPACE

+
20
employees



**EUROPEAN
LANGUAGE
DATA SPACE**

Unleashing the potential of data – for EU businesses and citizens

Context: Language Technologies, Artificial Intelligence and (Large) Language Models

- Language Technologies (LT) and Artificial Intelligence (AI) need language data
- Large Language Models (LLMs) lead now the AI race, needing vast amounts of language data to train them
- Europe's languages are largely under-resourced, except English
- A concerted effort for the collection of enormous amounts of language data for all European languages is very much needed



**EUROPEAN
LANGUAGE
DATA SPACE**

LDS and the Digital Europe Programme

The Common European Language Data Space (LDS)

- The LDS builds upon but goes far beyond initiatives like the European Language Resource Coordination (ELRC) and the European Language Grid (ELG), also focusing on making language data available
- Data Spaces are an inherent part of the EU Data Strategy
- The Common European Language Data Space is one of the approx. 20 official EU data space projects: strong focus on industry
- The LDS contributes to establish the European data market:
 - supporting European private and public Language Technology players and Language Data providers
 - Focusing on data sharing and monetising



**EUROPEAN
LANGUAGE
DATA SPACE**

Workshop Programme

09:15-09:30	Welcome by Dimitra Anastasiou - LIST
09:30-10:00	European Commission: The Digital Europe Programme and the Common European Language Data Space by Philippe Gelin - European Commission
10:00-10:30	European Language Data Space: developing a market for language data and services and benefitting from a joint European effort by Georg Rehm - DFKI and LDS Coordinator
10:30-10:45	Coffee Break
10:45-11:15	NLP For Low-Resource Languages by Fred Philippy - Zortify
11:15-11:45	The humans in the machines. How do thousands of average people actively and consciously contribute to a more reliable and less biased AI. by Alexandros Poulis - Dataforce - Transperfect
11:45-12:15	Panel Discussion: Language Data and Language Technologies in Luxembourg and for Luxembourgish Moderated by Dimitra Anastasiou - LIST Panelists: Gauthier Crommelinck - Ministry of the Economy Fred Philippy - Zortify Alexandros Poulis - Dataforce
12:15-13:30	Lunch break - provided on site
13:30-14:00	Large Language Model Bias dashboard by Jordi Cabot - LIST
14:00-14:30	Training large scale decoders for multilingual machine translation by Raheel Qader - Lingua Custodia
14:30-14:45	Coffee Break
14:45-15:15	Panel Discussion: Data for the Languages of Luxembourg. Which challenges do we face and how can we overcome these? Moderated by Daniele Pagani - LIST Panelists: Cédric Lothritz - LIST Yves Maurer - Bibliothèque nationale du Luxembourg Christopher Morse - Zenter fir d'Lëtzebuurger Sprooch Alistair Plum - University of Luxembourg
15:15-15:30	Discussion and Closing remarks



**EUROPEAN
LANGUAGE
DATA SPACE**

Practical Information:

Please share your feedback with us: *Evaluation form*

Do you need a *Certificate of attendance?*



Common European Language Data Space

Thank you and enjoy the workshop!



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

Dimitra Anastasiou (LIST, Luxembourg) - Moderator
Dimitra.anastasiou@list.lu

19-06-2024 LDS Country Workshop, Luxembourg
<https://language-data-space.ec.europa.eu>