



EUROPEAN LANGUAGE DATA SPACE



Common European Language Data Space The LDS Workshop in Slovakia

Miroslav Zumrík (Ľ. Štúr Institute of Linguistics of Slovak Academy of Sciences, Slovakia)
miroslav.zumrik@korpus.juls.savba.sk

07-11-2024 LDS Slovakia Workshop, Bratislava
<https://language-data-space.ec.europa.eu>



**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

09.30 – 09.45	REGISTRATION & COFFEE
09.45 – 09.55	Welcome and Introduction <i>Miroslav Zumrík, Ľ. Štúr Institute of Linguistics of Slovak Academy of Sciences</i>
09.55 – 10.00	Welcome <i>Barbora Tricoire, Representative of the European Commission's Directorate General for Translation in Slovakia</i>
10.00 – 10.30	The Digital Europe Programme and the Common European Language Data Space <i>Mateusz Szoturma, Representative of the European Commission</i>
10.30 – 11.00	Common European Language Data Space: Developing a Market for Language Data and Services and Benefitting from a Joint European Effort <i>Simon Ostermann, Representative of the Language Data Space & DFKI – German Research Center for Artificial Intelligence</i>
11.00 – 11.30	COFFEE BREAK
11.30 – 12.00	Importance and Advances in Language Data Processing in Slovakia <i>Radovan Garabík, Ľ. Štúr Institute of Linguistics of Slovak Academy of Sciences</i>
12.00 – 12.30	Horizon Europe Project DisAI <i>Marián Šimko, Kempelen Institute of Intelligent Technologies</i>

12.30 – 13.15	<p>PANEL SESSION Language Data and Language Technologies in Slovakia and for Slovak</p> <p><i>Moderator: Miroslav Zumrík</i></p> <p><i>Michal Gregor, Kempelen Institute of Intelligent Technologies</i></p> <p><i>Marek Šuppa, Sli.do & Faculty of Mathematics, Physics and Informatics of Comenius University Bratislava</i></p>
13.15 – 14.15	LUNCH BREAK
14.15 – 15.00	<p>PANEL SESSION Language Data Production, Management and Market Development: Overcoming Obstacles</p> <p><i>Moderator: Miroslav Zumrík</i></p> <p><i>Filip Bednárík, Essential Data</i></p> <p><i>Andrej Greguš, Nettle.ai</i></p>
15.00 – 15.15	<p>Conclusions</p> <p><i>Miroslav Zumrík, Ľ. Štúr Institute of Linguistics of Slovak Academy of Sciences</i></p>
15.15 – 16.00	COFFEE & NETWORKING



**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.

Miroslav Zumrík (Ľ. Štúr Institute of Linguistics of Slovak Academy of Sciences, Slovakia)
miroslav.zumrik@korpus.juls.savba.sk

07-11-2024 LDS Slovakia Workshop, Bratislava
<https://language-data-space.ec.europa.eu>