



Contribution ID: 60

Type: **Short Presentation**

The ONTOLISST project on DDI metadata, vocabularies and NLP

Tuesday, 3 December 2024 14:15 (10 minutes)

The talk introduces the new 2-year ONTOLISST project starting in December 2024, funded by the first OS-CARS Cascading grant call. The project will develop a simplified multilingual ontology (LiSST) to describe social science research data, create a corpus of social science metadata, and research whether and how NLP tools can help with (semi)automated (meta)data curation. The aim is to better understand how social science archives assign thematic metadata to their datasets in order to describe their contents and how data curation practices shape social scientific understanding. ONTOLISST will build on metadata in DDI format in different languages from various sources and using different CVs. The presentation outlines the project tasks, expected outputs and relationships with existing standards and tools. It also discusses how AI could help to accelerate the tedious, resource-intensive but important work of metadata and data curation and improve (meta)data interoperability across languages and disciplinary barriers.

Primary authors: DANCIU, Alina (Sciences Po, Center for Socio-Political Data (CDSP)); GÁRDOS, Judit (Hungarian Research Network Centre for Social Sciences, Research Documentation Centre (KDK)); KLEEMOLA, Mari (Finnish Social Science Data Archive, Tampere University)

Presenters: DANCIU, Alina (Sciences Po, Center for Socio-Political Data (CDSP)); GÁRDOS, Judit (Hungarian Research Network Centre for Social Sciences, Research Documentation Centre (KDK)); KLEEMOLA, Mari (Finnish Social Science Data Archive, Tampere University)

Session Classification: Sustainable, Ethical and Economical Controlled Vocabularies