

ONTOLISST – Thematic ontologies in social science research data



ONTOLISST combines existing social science resources with new algorithms to optimise topic assignment in multilingual contexts. It mainly explores thematic metadata but also how standard background variables, such as gender or age, can be made interoperable. The ultimate goal is to support cost-effective FAIR discovery tools in multiple EU languages, improving access to and visibility of social science research for a diverse group of users, from metadata experts and researchers to business and policymakers.



SSHOC
Social Sciences and Humanities

Challenge

There is a lack of widely used, streamlined, multilingual tools to assign topics and concepts to surveys on variable level across social scientific data from various archives, which impedes data discoverability and accessibility.

Solution

The Light Social Science Thesaurus (LiSST), developed and employed using selected NLP algorithms to produce a cross-language annotated dataset (gold standard corpus). All these tools can be exploited for automated or semi-automated topic assignment in research data repositories.

Scientific Impact

ONTOLISST builds theoretical knowledge about thematic metadata creation and curation, which is put to practical use in developing and integrating a multilingual thesaurus (LiSST), NLP tools, and AI-driven methods, fostering interoperability across research infrastructures and enabling improved discoverability of data for diverse users.

Partners

Research Documentation Centre (coordinator) and PolTextLab at HUN-REN Centre for Social Sciences, Center for Socio-political Data at Fondation Nationale des Sciences Politiques, and Finnish Social Science Data Archive at Tampere University.

<https://bit.ly/OSCAR-fundedproject-ONTOLISST>