Linked Open Research Data for Social Science

A concept registry for granular data documentation

Pascal Siegers[1][https://orcid.org/0000-0001-7899-6045], Antonia May[1][https://orcid.org/0000-0002-4979-0977], Claudia Saalbach[2][https://orcid.org/0000-0002-1748-908X], Jana Nebelin[2], Dagmar Kern[1][https://orcid.org/0000-0003-1794-625X], Andreas Daniel[3][https://orcid.org/0000-0002-0111-8858], Ben Zapilko[1][https://orcid.org/0000-0001-9495-040X], Fakhri Momeni[1][https://orcid.org/0000-0002-5572-575X], Knut Wenzig[2][https://orcid.org/0000-0002-2259-0203], and Jan Goebel[2][https://orcid.org/0000-0002-3243-1935]

[1] GESIS Leibniz-Institute for the Social Sciences, Germany

[2] Deutsches Institut für Wirtschaftsforschung eV

[3] German Centre for Higher Education Research and Science Studies

Abstract. The re-use of research data is an integral part of research practice in the social and economic sciences. To find relevant data, researchers need adequate search facilities. However, a comprehensive, thematic search for research data is made more difficult by inconsistent or missing semantic indexing of data at the level of social science concepts (e.g., representing the theory language). Either the data is not documented at a granular level, or primary investigators use their ad-hoc terminology to describe their data. Consequently, researchers have to make great efforts to find relevant or comparable data. From the user's perspective, the lack of theory language in data documentation impedes effective data searches and thus significantly limits the research potential of existing data collections. Because there is currently no semantic model for indexing the data content, the specific challenge for improving data search lies in establishing concept-based indexing of research data. Research infrastructures need technology for the harmonized semantic indexing of their research data. The LORD concept registry aims at closing this gap by developing a registry of sociological and economic concepts and, following the FAIR principles, making this concept registry generally available to the scientific community. As a first step, we developed a basic data model for the Concept Registry using United Modeling Language (UML). All links between are created and managed in the form of so-called RDF triples. An annotation application allows for linking specific questions/variables to concepts. The application also includes the two SKOS-compliant thesauri, "Thesaurus Social Sciences" (TheSoz) and "Standard Thesaurus Economics" (STW) but could be extended to other resources like ELSST.

We illustrate the application of the LORD concept registry with examples from three large-scale survey programmes (German Socio-Economic Panel, German General Social Survey, National Academics Panel Study). The initial focus is on variables and questions with overlapping content in the three survey programmes, as they form a sound basis for cross-linking with concepts.

Keywords: Linked Open Data, Indexing, Interoperability, Survey research

Data availability statement

Data produced for this study is not ready for publication yet. And there is no practical use of it.

Funding