

Date:
24 March 2020
Location:
Utrech, The Netherlands

Organised by SSHOC, the workshop "Linking Social Survey and Linguistic Infrastructures through EOSC" will be in co-location with the 3rd SSHOC Consortium meeting.

Workshop Objectives

Survey Infrastructures systematically interview tens of thousands of individuals across Europe each year.
Randomly selected from the population, the chosen people provide the survey infrastructure with a wide range of data on themselves that is valuable to researchers and subsequently, policymakers. Yet, a large proportion of the information conveyed in an interview is lost. Coding complex events into structured taxonomies, needed for cutting edge sociological research, many aspects of an individuals' responses are thrown away.
The respondents' tone, their clarity, their fluidity, the depth of the vocabulary can all be used to provide insights into various concepts of interest to social scientists such as cognitive function, socio-economic status or verbal reasoning skills. To make use of this lost data however, it is necessary to integrate the tools of linguistic infrastructures into the analytical pipeline of survey infrastructures.

The cross-pollination and integrated usage of tools are precisely what the EOSC aims to do and the work on voice recorded interviews and audio analysis of SSHOC, therefore, seeks to provide a proof of concept and framework for future research that explores this approach.

The workshop will comprehend a series of presentations and collaborative discussions around the potential for integrating social survey and linguistic infrastructures.

- **Tom Emery** will present work conducted by the GGP on capturing audio data through existing survey software in online interviews, and will provide initial evaluations of data quality.
- **Henk van den Heuvel** from the Oral History team will then detail the tools used for the analysis of Oral History data which could be adapted for analysis of survey interviews. In particular, he will address the so-called Transcription Chain, the basis of which is automatic speech-to-text conversion. The resulting text can, if needed after manual correction, be processed by NLP tools to obtain more insights into the linguistic structure, or to carry out a topic detection or text summarisation, to mention a few options.

An interactive discussion between participants, regarding the potential application of these tools and new avenues of scientific enquiry that could be integrated into the next phase of work on voice

© Copyright 2020 – FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

recorded interviews and audio analysis will then follow.

Find here more details of the workshop, the agenda and the registration form.

© Copyright 2020 – FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.