SOCIETY FOR LONGITUDINAL AND LIFECOURSE STUDIES (SLLS)

WORKSHOP 4

Investigating Change Across Time and Place: Using AI to Facilitate Retrospective Data Harmonization

Convenor: Eoin McElroy, Ulster University, UK

SUMMARY

Pooling data from multiple longitudinal population surveys offers significant advantages for social, health, and economic research. By integrating datasets across different cohorts, researchers can enhance statistical power, enabling more precise estimates of trends and associations. This approach also facilitates cross-national and cross-cohort comparisons, revealing how demographic, policy, and environmental differences influence outcomes over time. Additionally, combining datasets increases sample diversity, improving the generalizability of findings and allowing for subgroup analyses that might be underpowered in individual studies. However, harmonizing variables across surveys often presents challenges, including differences in data collection methods and instruments.

This half-day workshop will discuss these challenges and provide guidance on how they can be addressed. We will focus on how developments in AI, specifically Natural language processing (NLP), can be used to facilitate the harmonization of questionnaire data.

Workshop objectives:

- Outline common challenges when harmonizing data
- Demonstrate different approaches to data harmonization
- Provide a demonstration of Harmony (https://harmonydata.ac.uk/), an open-source NLP tool for harmonizing questionnaire data
- Introduce statistical methods for testing the comparability of data across studies

Target audience: Researchers from across the social and health science, particularly those who routinely use questionnaire data in their research. Basic understanding of data management, multivariate analyses will be useful. A basic knowledge of R code would be useful, but not essential.