Contribution ID: 3

Type: Talk

Achieving data interoperability through harmonized metadata for joint data analysis: Lessons learnt from ENPADASI, INTIMIC-KP and NFDI4Health

Wednesday 11 October 2023 11:10 (20 minutes)

Joint data analyses may overcome challenges of traditional literature-based meta-analysis owing to the use of harmonized exposure and outcome definitions as well as statistical modelling. It also allows to re-use existing data for other research purposes in more flexible ways to increase scientific impact. Achieving FAIR (meta)data by generating interoperable, harmonized, high quality metadata and data harmonization is mandatory for joint data analysis. DataSHIELD is a software tool allowing remote federated analysis of harmonized datasets across studies without physically sharing individual-level data, thereby substantially reducing burdens and challenges for data sharing that are especially common in ongoing observational studies. As part of the European Nutritional Phenotype Assessment and Data Sharing Initiative (ENPADASI), and the Knowledge Platform Intestinal Microbiomics (INTIMIC-KP) within the Joint Programming Initiative a Healthy Diet for a Healthy Life (JPI-HDHL) as well as the National Research Infrastructure for Personal Health Data (NFDI4Health) we are collecting and harmonizing metadata on observational studies in the field of nutrition, biomarkers, omics, and chronic diseases. We also established a searchable Mica database to make harmonized metadata publicly available (https://mica.mdc-berlin.de). Based on study-level metadata, studies eligible for federated DataSHIELD analyses of multiple study data can be identified and consent for participation in a federated analysis can be requested from the principal investigators. We provide standard operating procedures (SOPs) for Opal/DataSHIELD infrastructure installation, data upload and setting permissions for eligible studies. Alternatively, harmonized datasets can be hosted at the MDC Opal database. We also provide SOPs for semi-automated data harmonization using the R package harmonizR developed by Maelstrom Research. Federated analysis of studies is performed centrally at the MDC with DataSHIELD. Currently, we are extending this work as part of NFDI4Health to provide a central access point at the MDC for interested researchers to conduct federated data analysis of multiple epidemiological studies. In addition, the handling of credentials will be optimized by central access solutions (keycloak) in NFDI4Health. Our experiences show that harmonized and searchable study-level metadata are useful for identifying eligible studies for a federated analysis of a specific research question, yielding successful publication. Lessons learnt are also that data harmonization as well as the set-up of Opal and DataSHIELD needs time and resources at the study-level. With these lessons learnt we are contributing to shaping the research infrastructures being built by NFDI4Health.

Please assign your contribution to one of the following topics

Data interoperability through harmonised metadata and interoperable semantics

Please specify "other" (stakeholder)

In addition please add keywords.

metadata; data harmonization; data re-use; FAIR principles; DataSHIELD; federated analysis

Please assign yourself (presenting author) to one of the stakeholders.

Researchers

Primary author: NIMPTSCH, Katharina

Co-authors: Dr SCHWEDHELM, Carolina; SIAMPANI, Sofia M.; Dr PINART, Mariona; Prof. PISCHON, Tobias

Presenter: NIMPTSCH, Katharina

Session Classification: Parallel Track 1

Track Classification: Facilitating connectivity of research data: Data interoperability through harmonised metadata and interoperable semantics