Contribution ID: **34** Type: **not specified**

Assessing Data Management and Compliance in Large Research Collaborations for Consent, Data Sharing, and Knowledge Transfer

Tuesday, September 30, 2025 2:15 PM (1h 30m)

This paper addresses the complex challenges of data management, compliance with legal and ethical standards, and knowledge transfer in large-scale collaborative research centers (CRCs), particularly in the context of linguistic data. The management of sensitive information, such as personal and health data, and the need for proper anonymization of audio and video recordings, pose challenges for researchers and data management practices. These legal issues, together with ethical requirements, complicate data sharing and reuse, which are essential components of adherence to FAIR (findable, accessible, interoperable, reusable) data principles. However, legal and ethical requirements must be complied to when it comes to study planning, data management plans, and further good research practices, even before the data have been collected (Jorschick et al. 2024). Structured research data management (RDM) can improve project documentation and collaboration efficiency (Mittal et al. 2023), but RDM strategies need to be tailored to the needs of the individual projects in order to allow for effective collaboration and reduced time spent on data management (Kanza & Knight 2022, Pascquetto et al. 2017).

In order to meet the issues outlined above, we introduce the development of an integrated platform to support structured RDM within a large Collaborative Research Centre (CRC) in linguistics. This platform includes (1) a comprehensive formal *ontology*, a conceptual schema, composed of definitions of abstract classes and their properties as well as the relation between them, which serves as the foundation to data and metadata representation according to FAIR principles, (2) a consent form *wizard* that supports researchers in setting up studies in accordance with RDM, e.g., by automatically generating legally compliant consent forms tailored to the specific requirements of each experiment, and (3) a *knowledge base* that records the collected information using the classes and relations of the ontology as well as guidelines for good research practices. In later phases, the combination of knowledge base and ontology will ground the platform's (meta)data, in that every dataset will be automatically linked to its participant-level consent, allowing researchers to query and monitor status for data sharing and (re-)use (Jorschick et al., 2024).

As the initial phase of platform development, we designed a semi-standardized interview protocol to assess each CRC project's data management and data protection practices. The guided questions trace the entire data life-cycle –from collection through storage, stewardship, legal-ethical safeguarding, and sharing –while also mapping the team's technical skills, workflows, and perceived training gaps. By confronting researchers with the interview questions, the method guided not only to evaluate current practice but also to raise awareness of legal and ethical obligations and identifies where training and infrastructure support were still needed. The collected data were then used to start refining the platforms knowledge base and the ontology, building on existing GDPR-compliant ontologies like GConsent (Pandit et al. 2018). We present knowledge graphs as a visualization of the information gained, representing the interrelationships between the projects in the CRC, such as the type and use of collected data, or research goals. These graphs allowed for the identification of potential synergies and connections between projects, promoting inter-project collaboration, and facilitates the clearer identification and continued pursuit of future research directions and management.

By integrating technical, legal, and ethical considerations into the research infrastructure, both the interview process and the ontology development for the platform aim to improve the overall sustainability and compliance of collaborative research. This presentation describes the development of this user-oriented solution. It will outline the structure of the platform, the role of the ontology, and the key benefits of implementing such a solution for large-scale interdisciplinary research collaborations.

References

Jorschick, A., Schrader, P., & Buschmeier, A. (2024), What can I do with this data point? Towards modeling legal and ethical aspects of linguistic data collection and (re-)use. *Proceedings of the Workshop on Legal and Ethical Issues in Human Language Technologies at LREC-COLING 2024*. ELRA and ICCL, 47–51. https://aclanthology.org/2024.legal-1.8

Kanza, S. & Knight, N. J. (2022), Behind every great research project is great data management. In: *BMC Research Notes*, 15(20). https://doi.org/10.1186/s13104-022-05908-5

Mittal, D., Mease, R., Kuner, T., Flor, H., Kuner, R., & Andoh, J. (2023), Data management strategy for a Collaborative Research Center. In: *GigaScience*, 12, https://doi.org/10.1093/gigascience/giad049

Pandit, H. J., Debruyne, C., O'Sullivan, D., & Lewis, D. (2018). *GConsent –A Consent Ontology based on the GDPR*. Retrieved from https://openscience.adaptcentre.ie/ontologies/gconsent/main.html.

Pasquetto, I. V., Randles, B. M., & Borgman, C. L. (2017), On the reuse of scientific data. *Data Science Journal*, 16(8), 1–9. https://doi.org/10.5334/dsj-2017-008

Abstract

Poster

Authors: JORSCHICK, Annett (Universität Bielefeld); MOHAMMADI, Maryam (Universität Bielefeld); POLITT, Katja (Universität Bielefeld); BUSCHMEIER, Hendrik (Universität Bielefeld)

Session Classification: Poster Session