Open Research Practices with the OntoME-Geovistory environment

Alamercery, Vincent
vincent.alamercery@ens-lyon.fr
ENS de Lyon, LARHRA

Beretta, Francesco
francesco.beretta@cnrs.fr
CNRS, LARHRA

Favey, François-Joseph
francois-joseph.favey@unifr.ch
KleioLab

Ferhod, Djamel
djamel.ferhod@cnrs.fr
CNRS, LARHRA

Knecht, David
david.knecht@kleiolab.ch
KleioLab

Muck, Gaétan
gaetan.muck@kleiolab.ch
KleioLab

Perraud, Alexandre
alexandre.perraud@cnrs.fr
CNRS, LARHRA

Pica, Morgane
morgane.pica@ens-lyon.fr
ENS de Lyon, LARHRA

Schneider, Jonas
jonas.schneider@kleiolab.ch
KleioLab

Stebler, Andreas
info@astebler.ch
KleioLab

Academic research in the Humanities and Social Sciences (HSS) collects considerable amounts of relevant, curated, high-quality information in form of digital data, in order to produce new knowledge about different aspects of the past and present human societies. Whereas natural sciences have developed best practices and workflows for research data publication and re-use, there is a limited practice in this regard in the HSS. In this domain, a lot of valuable data is collected but there is a lack of (a) standardized knowledge representation languages (ontologies) to document the meaning of data and (b) virtual research environments (VRE) facilitating collaborative data production, publication and reuse. Instead, spread-sheets and more or less sophisticated information systems are employed, which often leads to data loss at the end of the project or results in a long-term archiving of data whose semantics is not made explicit and/or the model strictly focused on the specific research approach and therefore hardly reusable. Hence, current established practices in HSS for data production are mostly not in line with the FAIR principles and the used infrastructure often not fully sustainable. This leads to limited open research data (ORD) practices in this area.

In the meantime, digitization of research becomes a more and more important topic in scientific disciplines. Agencies like SNSF, ANR and Horizon Europe ask funded research projects to detail how the produced data respond to the FAIR principles. Furthermore, universities increasingly include data management courses in their curricula. This trend increases the awareness of the academic domain for the need of standardization and best practices in producing and reusing research data.

To cope with these issues, the Digital History Research Team of “Laboratoire de Recherche Historique Rhône-Alpes” (LARHRA), building on its several-year-long experience with the symogih.org project, and the IT-startup KleioLab based in Basel initiated a new approach, which is further developed by an academic consortium around the University of Bern, supported by swissuniversities.

This approach, on the one side, consists of OntoME (https://ontome.net), a collaborative ontology management environment focusing on semantic interoperability, developed by LARHRA in order to foster distributed data management and data reuse using semantic web technologies. OntoME supports the development of an ecosystem of domain related extensions of the CIDOC CRM ontology (ISO 21127). Of particular relevance in OntoME is the ‘Semantic Data for Humanities and Social Sciences’ ontology ecosystem (sdhss.org), which ensures coherence between research models from different projects in HSS, at different levels of abstraction (Beretta 2021, 2022). In parallel, the Data for History initiative was launched, with a first international conference held in 2021 [1] (Beretta/Alamercery 2020).

On the other side, a new VRE and data publication platform, Geovistory (https://www.geovistory.org), has been created in the context of a public-private joint venture between LARHRA and KleioLab. Thus, Geovistory builds on LARHRA’s longstanding experience in collaborative modelling and ORD management but is structured as a public-private initiative and therefore more independent and can easily respond to researchers’ needs from all over Europe.

Geovistory combines different features that make it an attractive tool to strengthen ORD practices. To begin with, it has been developed according to ‘user experience methodologies’. Also, each research project works on its own data perspective but at the same time automatically contributes to the joint knowledge graph. A particular strength of Geovistory is its handling of the challenges of scientific information in HSS: The context-sensitive nature of information and its relation to different research agendas, the wide variations in meaning for the same terms and vocabulary complexities, competing views or gaps and fragmentation of available information. This is what makes the Geovistory graph data model and semantic enrichment capabilities so interesting for HSS research.

Importantly, the data model is collaboratively managed in OntoME and takes advantage of the sdhss.org project and the avail-

able, extensible application profiles for different HSS domains. Already existing profiles for basic biographical data, social activities, concepts of laws and customs, archeology, etc. can be extended with application profiles such as the treatment of dramatic works and literature, vocabularies for sociology and political science, concepts used in linguistics, etc. The possibility of extending the ontology as needed thanks to a robust and modular modelling method allows Geovistory to be proposed as VRE to researchers in different HSS disciplines, thus facilitating the collection of information about past and present societies.

Finally, Geovistory allows publishing each project's data on project-specific webpage and dedicated sparql-endpoint, making it an interesting tool not only for data production, but also publication.

The poster highlights the workflow proposed in the OntoME-Geovistory environment.

Notes

1. See: http://dataforhistory.org/meetings

Bibliography

Beretta, Francesco (2022): "Interopérabilité des données de la recherche et ontologies fondationnelles : un écosystème d'extensions du CIDOC CRM pour les sciences humaines et sociales" in: Lasolle, Nicolas / Bruneau, Olivier / Lieber, Jean (eds.): *Actes des journées humanités numériques et Web sémantique.* Nancy 2-22. DOI: 10.5281/zenodo.7014341.

Beretta, Francesco (2021): "A challenge for historical research: making data FAIR using a collaborative ontology management environment (OntoME)", in: Semantic Web 12, 2: 279–294. DOI: 10.3233/SW-200416.

Beretta, Francesco / **Alamercery, Vincent** (2020): "Du projet symogih.org au consortium Data for History - La modélisation collaborative de l'information au service de la production de données géo-historiques et de l'interopérabilité dans le web sémantique", in: Revue ouverte d'ingénierie des systèmes d'information 3, 1. DOI: 10.21494/ISTE.OP.2020.0532.