Historical Research meets Semantic Interoperability: The Documentation System SYNTHESIS and its Application in Art History Research

Digital Humanities 2022 (DH2022)

Presenter: Pavlos Fafalios (fafalios@ics.forth.gr)
Centre for Cultural Informatics, Institute of Computer Science, FORTH, Greece

Introduction and Motivation

Historical science is the field that describes, examines and questions a sequence of past events, and investigates patterns of cause and effect. Research in the field usually starts by first discovering, collecting, documenting and organizing historical sources, such as written documents or material artifacts. This often includes either the transcription (and then curation) of historical archival sources, like in Petrakis et al. (2020) for the case of Maritime History, or the detailed documentation of cultural artifacts and related evidence, like in Fafalios et al. (2021) for the case of Art History, with the latter being the focus of this presentation.

In this context, although computing in the field has developed enormously over the last years, data management problems still exist and are very varied. Common problems include: a) the difficulty for collaborative but controlled documentation by a large number of historians of different research groups; b) the lack of representation of the details from which the documented relations are inferred, important for the long-term validity of the research results; c) the difficulty to combine and integrate information extracted from multiple and diverse information sources; d) the difficulty of third parties to understand and re-use the documented data, resulting in the production of data with limited longevity.

The SYNTHESIS system

In an effort to cope with the aforementioned problems, we present the SYNTHESIS documentation system and its use by a large number of historians in the context of a large European research project of Art History, called RICONTRANS (ERC Consolidator Grant, No 818791). SYNTHESIS is Web-based, multilingual, configurable (for use in other digital humanities fields), and utilizes XML technology, offering flexibility in terms of versioning, workflow management and data model extension. It focuses on semantic interoperability (Ouksel and Sheth, 1999), enabling the exchange of data among computer systems with unambiguous/shared meaning, and achieves this by making use of standards for data modelling and publication, in particular the formal ontology CIDOC-CRM (ISO 21127:2014) and the data model RDF (W3C Recommendation). The aim is the production of data with high value, longevity and long-term validity that can be (re)used beyond a particular research activity.

SYNTHESIS offers a wide range of functionalities including i) interlinking of the documented entities (forming a network of interrelated entities), ii) management of static and dynamic vocabularies, iii) linking to thesauri of terms, iv) connection with geolocation services (TGN, Geonames), v) map visualization for certain types of entities, vi) support of comparable historical time expressions (e.g., ca. 1920, early 16th century), vii) management of digital files (images, etc.), viii) transformation of the documented information to a knowledge base of Linked Data (Heath and Bizer, 2011).

Figure 1: The user interface of SYNTHESIS displaying the supported entity types (on the left) and the table of documented entities of type 'Object' (on the right).

Figure 2: Viewing the documentation card of an entity of type 'Object Transfer' in SYNTHESIS.
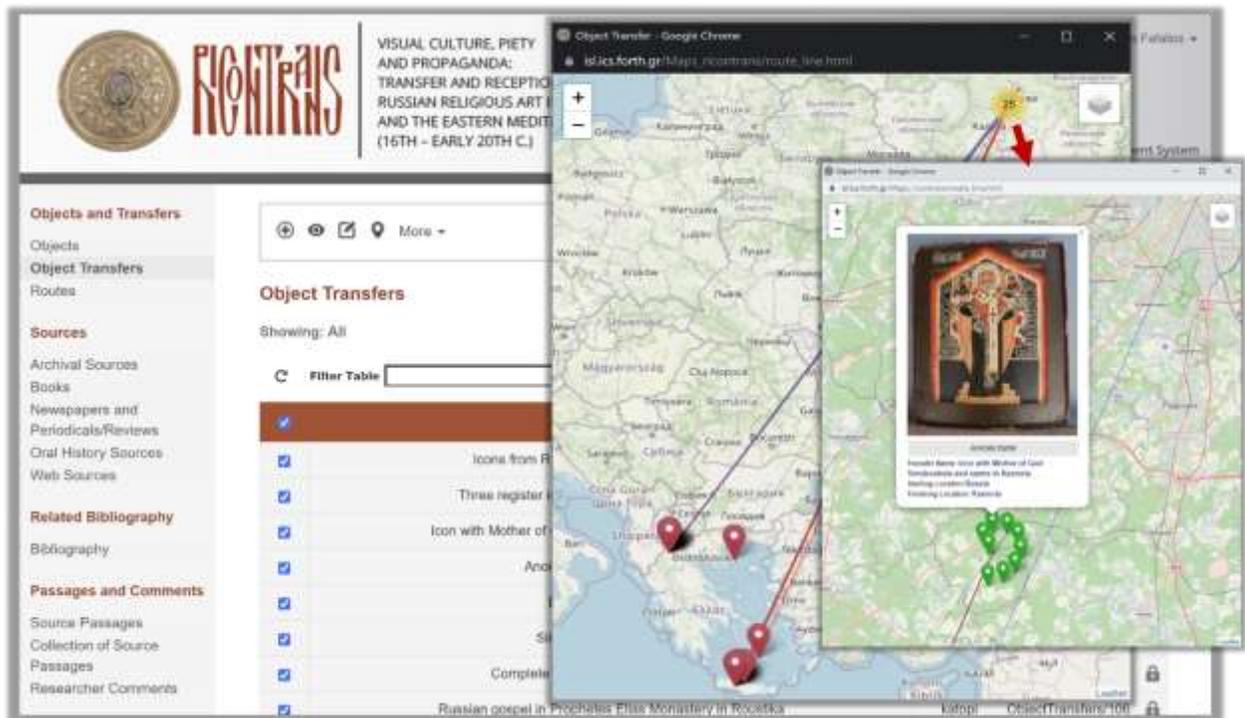
Figure 3: Visualizing a set of object transfers in a map.

Overall Philosophy and Novelty

SYNTHESIS provides full-fledged support for the complete knowledge production life cycle in historical research, enabling the inclusion of rich provenance information (metadata) about the documented data and providing embedded processes for transforming the data to a Knowledge Base (KB) of historical information that is compliant to CIDOC-CRM. Contrary, to approaches that support users in creating a KB from the beginning, such as ResearchSpace (Oldman and Tanase, 2018), SYNTHESIS considers a workflow that decouples data entry from the ontology-based integration and the creation of the KB. The main reasons behind our approach are two: (a) we regard as very different a KB of facts believed together as true, versus managing and consolidating the knowledge acquisition process of a large research team. The latter requires a document structure, for making local versioning, workflow management and provenance tracing easy; (b) we consider a KB as an ideal tool for integrating the 'latest stage of knowledge'; individual contributions, alternatives, corrections, etc., all in the same pool of valid knowledge, can hardly be regarded as a standard procedure.
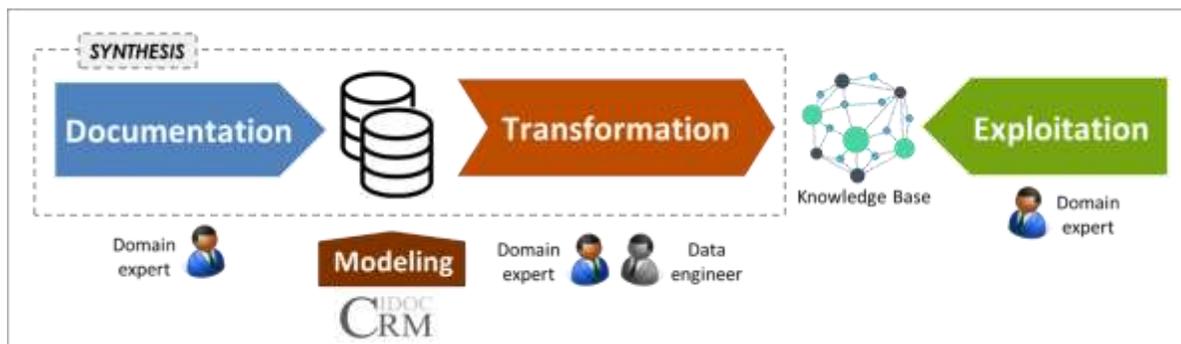

Figure 4: The SYNTHESIS workflow: documentation and transformation to a knowledge base of historical information for exploitation in historical research.
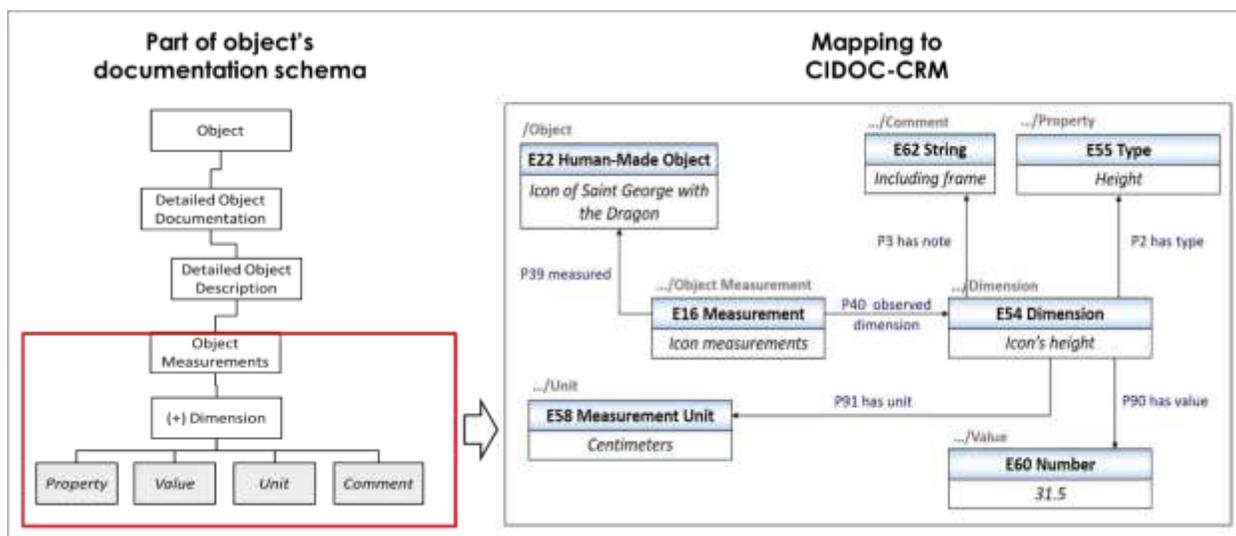
Figure 5: Mapping a part of object's documentation schema to CIDOC-CRM.

Usage in Historical Research

The system is currently used by around 40 users in five countries (mainly historians) in the context of the ongoing project RICONTRANS, whose research focus is the Russian religious artefacts brought from Russia to the Balkans after the 16th century and which are now preserved in churches, monasteries or museums. The system supports the documentation of entities belonging to totally 18 entity types, each one having its own documentation schema (data entry fields organized in a hierarchical, tree-like structure). Indicative entity types include: objects (like icons), object transfers (like donations), archival sources, oral history sources, source passages (like a paragraph in a newspaper that talks about a topic of interest, e.g., an icon donation), historical figures (like archbishops). Currently, more than 5,000 entities have been already documented and are used in historical research, including more than 1,700 objects, 550 object transfers, 200 archival sources, 850 source passages, and 230 historical figures. By exploiting the rich connections among the documented entities, historians can find answers to complex information needs, such as "finding the routes of icons transferred to Mount Athos before the 18th century as well as the purpose of these transfers".

Lessons Learned and Future Work

Finding the best trade-off between documentation richness and usability was a challenging problem that required extensive discussions between historians and data engineers, as well as many revisions and updates of the entity documentation schemas. On the one hand, researchers need to document information in a detailed and precise way, but on the other hand, this must be done in a quick and straightforward way. In addition, controlling the dynamic vocabularies, which allow the on-the-fly inclusion of new terms, is difficult when there is large number of editors. The problem is that we can end up with vocabularies containing multiple terms that refer to the same notion. Curation is then needed which though is laborious. A question for future work is how the system could support a better management of the dynamic vocabularies.

Acknowledgements

- The following members of FORTH have been contributed in the design and development of the presented system: Konstantina Konsolaki, Lida Charami, Kostas Petrakis, Manos Paterakis, Dimitris Angelakis, Chrysoula Bekiari, Pavlos Fafalios, Martin Doerr.
- The presented work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 890861 (ReKnow) and the European Research Council grant agreement No 818791 (RICONTRANS).

References

Fafalios, P., Konsolaki, K., Charami, L., Petrakis, K., Paterakis, M., Angelakis, D., Tzitzikas, Y., Bekiari, C., and Doerr, M. (2021). *Towards Semantic Interoperability in Historical Research: Documenting Research Data and Knowledge with Synthesis*. Proceedings of the 20th International Semantic Web Conference (ISWC 2021), pp. 682-698, Springer, Cham.

Heath, T., and Bizer, C. (2011). *Linked data: Evolving the web into a global data space*. Synthesis lectures on the semantic web: theory and technology, 1(1), pp. 1-136.

Oldman, D., and Tanase, D. (2018). *Reshaping the knowledge graph by connecting researchers, data and practices in ResearchSpace*. Proceedings of the 17th International Semantic Web Conference (ISWC 2018), pp. 325-340, Springer, Cham.

Ouksel, A. M., and Sheth, A. (1999). *Semantic interoperability in global information systems*. ACM Sigmod Record, 28(1), pp. 5-12.

Petrakis, K., Samaritakis, G., Kalesios, T., Domingo, E. G., Delis, A., Tzitzikas, Y., Doerr, M., and Fafalios, P. (2020). *Digitizing, Curating and Visualizing Archival Sources of Maritime History: the case of ship logbooks of the nineteenth and twentieth centuries*. Drassana: revista del Museu Marítim, (28), pp. 60-87.