

# The Structured Assertion Record (STAR) Model for Event-based Representation of Historical Information

## Background

Among the many challenges posed by the study of the world of the Christian communities of the eleventh century is the scattered and arbitrary nature of the surviving sources; following from this, the information we derive from our sources varies wildly in its precision and its reliability. Moreover, our sources represent a wide variety of viewpoints that are represented therein and give rise to an equally wide variety of scholarly interpretations, which can be very difficult to represent usefully in collections of digital data.

One of the goals of the RELEVEN project has been to re-think the way in which we model data about the past. A straightforward Linked Open Data (LOD)-based model tends to result in positivistic data collections, where the pieces of information become separated from their provenance and context and thus difficult or impossible to verify (Kahn and Simon 2020); on the other hand, models such as factoid prosopography (Bradley and Pasin 2015) or even Wikidata ('Wikidata' 2023) emphasize the retention of provenance (source) information, at the expense of providing coherent and non-contradictory accounts of past events (Baillie 2021).

## Methodology

One of the key problems we face for eleventh-century history, which informs our approach to a solution, is that the historical claims that are made are often not traceable in their entirety to specific primary source material. A suitable information system must represent claims that arise from a preponderance of evidence, claims that arise from absence of evidence, and even claims that arise from conjecture.

This constraint leads to the core of our solution, which is to extend the linked open data model from a triple (subject, predicate, object) to a five-pointed structure (thus "STAR model") which includes subject, predicate, object, source, and authority. This model allows for the possibility of adopting and enriching existing LOD collections with source and authority data when it exists, since the subject, predicate, and object can come from any defined ontology.



With the STAR model we have been able to ingest data from the factoid-based Prosopography of the Byzantine World, and have begun a similar exercise to ingest data from the Pleiades gazetteer. Future work will include data-mining experiments for the automated detection of conflicts in information, and methods for visualization of the conflicts and ambiguities that are detected.

## Acknowledgments

This work arises from a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101002357).

## References

- Baillie, James. 2021. 'Alternative Database Structures for Prosopographical Research'. *International Journal of Humanities and Arts Computing* 15 (1-2): 117-32. <https://doi.org/10.3366/ijhac.2021.0265>.
- Bekiari, Chryssoula, George Bruseker, Martin Doerr, Christian-Emil Ore, Stephen Stead, and Athanasios Velios, eds. 2021. 'Definition of the CIDOC Conceptual Reference Model'. <https://www.cidoc-crm.org/Version/version-7.2>.
- Bradley, John, and Michele Pasin. 2015. 'Factoid-Based Prosopography and Computer Ontologies: Towards an Integrated Approach'. *Digital Scholarship in the Humanities* 30 (1): 86-97.
- Kahn, Rebecca, and Rainer Simon. 2020. 'Feast and Famine: The Problem of Sources for Linked Data Creation'. In *Graph Technologies in the Humanities - Proceedings 2020*, edited by Tara Andrews, Franziska Diehr, Thomas Efer, Andreas Kuczera, and Joris van Zundert, 3110:86. CEUR Workshop Proceedings. Vienna, Austria: CEUR. <https://ceur-ws.org/Vol-3110/#paper5>.
- 'Wikidata'. 2023. In *Wikipedia*. <https://en.wikipedia.org/w/index.php?title=Wikidata&oldid=1139146582>.