

# Networked Knowledge Organization Systems and Services (KOS) in Humanities Research and Cultural Heritage Collections

Johan Åhlfeldt (johan.ahlfeldt@ctr.lu.se)

Centre for theology and religious studies, Lund University, Sweden

## Background

Gazetteers and Prosopographical studies have long (since the 19<sup>th</sup> Century) played an important role in historical and archaeological studies and have constituted building blocks for classifying and organizing source material into knowledge systems and has left us huge printed source collections in different fields like Epigraphy, Numismatics and Diplomatics. They constitute important cultural heritage collections. In the digital age, many of these source collections have been digitized and become increasingly available online but not yet processed so that they can become machine actionable and aligned to each other. In other words, the potential of source material becoming digital has not yet been fully utilized. Gazetteers and prosopographies play a crucial role in this context to disambiguate places, organizations and people mentioned in the sources.

There are already many online resources and services regarding the sources of Greek and Roman civilizations, Early Medieval Europe and other time periods, although they don't all expose machine actionable data and speak to other systems, that is, there has been little authority control regarding, for example, places, organizations and people and other important concepts like time-periods, classification of things, etc. The last five years or so have seen many initiatives to build new kinds of digital gazetteers and prosopographies that overcome these problems. They utilize the concept of linked open data with shared ontologies and uniform resource identifiers.

## Aim

The aim has been to create a highly accurate and comprehensive digital gazetteer of ancient places attested in archaeological and textual sources belonging to the Greek and Roman civilizations, the *Digital Atlas of the Roman Empire* (DARE, <http://dare.ht.lu.se>). The aim was also to create a digital map that visualizes the content of the gazetteer. Both the gazetteer and the digital map have been constructed for easy and open access and reuse and are especially aimed at aligning distributed source collections and providing a contextual map for Ancient studies. One such source collection is the *Monastica* platform (<http://monastica.ht.lu.se>) which is a dynamic library and research tool for early monastic literature originating in Late antiquity. Source material in Monastica is compiled from both printed editions and scholarly work on manuscripts within the project. The author is responsible for the development of both systems. DARE has been an important part of the Pelagios collaboration, <http://commons.pelagios.org> from its start in 2012 providing both a gazetteer and the digital atlas.

## Methods

We have found it fruitful to organize and align data according to the principles of Linked Open Data representing digital objects with stable URI:s and shared ontologies to expose, share and connect data on the web. DARE and Monastica are both SQL-databases that map and serialize semantically defined data but also consume related and similarly constructed datasets from other providers in

real time to get an empirical context. LAWD is an ontology for Linked Ancient World Data that is utilized in both systems and to a large extent make both incoming and outgoing data exchange possible. The goal of LAWD is to fill in the gaps between the data used and published by projects with a focus on the ancient world and the standard Linked Data vocabulary schemes, like Dublin Core, the Open Annotation Collaboration, and CIDOC-CRM. LAWD accomplishes this by defining just enough new classes and properties to permit the semantic modeling of things like manuscript collation, relating provenance to geography, and prosopography. TEI (Text Encoding Initiative) represents a shared conceptualization of what exists in the domains relevant to text encoding and are utilized within the Monastica platform to describe both structure and content of text sources.

## Results and conclusion

Our work has resulted in two networked knowledge organization systems and services in the field of Ancient history. They are both networked in that they are connected to each other and to similar external systems that utilizes the same conceptual modelling. This enables the system to both produce and consume data within similar domains. They are both knowledge system, DARE is a historical gazetteer and Monastica is a thematic corpus of monastic literature and previous scholarly work. They both provide services that expose semantically defined data through persistent URIs and APIs and in the latter case define a network service, Canonical Text Services (CTS) for identifying texts and for retrieving fragments of texts by canonical reference expressed as CTS URNs.

From the two knowledge systems we can draw the conclusion that Humanities research needs its own specialized knowledge organization in the digital world to better coordinate dispersed online source collections. Digital gazetteers and prosopographies are important elements in aligning source collections to each other, that is, to disambiguate between places, organizations and people. There is a clear need for specialized authority files, classification schemes, thesauri, topic maps, ontologies etc. valid within a delimited spatial and temporal range, but also linked to more general knowledge organization systems. The author is involved in several initiatives that address these challenges: the *Digital Atlas of the Roman Empire* is one of several recognized gazetteer in the Pelagios Commons network, aligning source material to persistent and shared identifiers of place in ancient Greece and the Roman Empire. It has a special focus on bridging between cultural heritage collections and Humanities research. *Monastica - a dynamic library and research tool for early monastic literature* is a thematic collection of one of the oldest corpora of Early Christian literature originating in Late antiquity that utilizes digital gazetteers and prosopographies to reference places, organizations and people mentioned in the sources.