

Olof Karsvall, Swedish National Archives, olof.karsvall@riksarkivet.se
Leif-Jöran Olsson, Friprogramvarusyndikatet, info@friprogramvarusyndikatet.se

Medieval Sweden - a reference work transformed into XML

In the medieval and early modern research fields, there is in several countries a tradition of long-term publishing projects, the purpose of which is to compile, make available, and present difficult-to-access handwritten source materials. It could either be descriptions of manuscripts or paleographic editions, but also overview works and other types of chronological or thematic compilations.

The obvious way to publish these reference works has been and still is to print them as books. But as more and more source material is digitized, the need to also digitize the central standard works arises. It would simplify the use and increase awareness of the diversity of source materials and collections contained in the archives.

Reference works in digital format have several obvious advantages over printed versions. It will be easier to compile and find material when the content can be interpreted as data automatically. Basic information about people, places, institutions, and other phenomena can be identified, labeled, and linked to other resources that describe the same thing. A digital format can also stimulate broader collegial collaborations, where experts from different fields can collaborate to review, correct and continuously improve the content of the digital editions, which is not possible in printed editions.

The reference work Medieval Sweden - which is in focus in this article - is a printed book series of the medieval and 16th-century rural settlements in Sweden, excluding Finland (see <https://riksarkivet.se/medieval-sweden-dms>). Each published booklet covers a geographical area (a district or similar) and reports all the villages and hamlets that appear in the written source material from the 12th century until the middle of the 16th century. It contains the oldest place-name evidence, cadastral information on the number of farms etc. Moreover, the medieval letters concerning land transactions are extracted and summarized. The information is presented topographically (by district, parish, village) and alphabetically, which makes the content easy to overview.

As a reference work, it does not contain full texts or descriptions of source material per se, but excerpts (contextual extracts) that are systematically compiled for each settlement named in the sources. For example, a medieval letter describing a land purchase between two villages will be reported twice in Medieval Sweden, as an event linked to both the villages concerned. In other words, the focus is on the content, primarily place names, landowners, settlements, and events such as purchases, exchanges, and disputes on farmland. By collecting evidence from primary sources, Medieval Sweden serves as an entrance and base for further research and in-depth study. The target group is historians but it is also used by community planners, archaeologists, and others who analyze and describe the historical dimensions of the landscape, and for those who find the original sources too difficult to interpret.

This paper has two parts. To begin with, the conditions and possibilities of printed reference work in digital formats are discussed. What are the advantages of reference works, which follow a predetermined structure and accepted principles, compared to resources that are born digitally? What are the special challenges of digitizing printed works of this kind? How can they be represented in a digital format? For example, how can structural information and meaningful words and paragraphs become machine-readable elements?

The second part of this paper shows in practice how the digitization of Medieval Sweden has taken place, what choices and delimitations have been made and what techniques have been used. The demo application developed has two aims. It is a tool for the editors who create new booklets in the Medieval Sweden series, which will allow print on demand. It also has a public search interface where all information is presented and made searchable as a single data source. The digital format is based on XML with enumerations according to the Text Encoding Initiative, where each published booklet is its own XML file. The database and applications have been built with eXist-db. On top of this, application programming interfaces have been built that enable integration and linking to other digital resources.

If a research project such as Medieval Sweden were to start today, it would probably be designed from the beginning as a digital edition. However, there is much to learn from how established printed reference works are designed and structured. Not least methodologically, by providing a form and structure that is established and scientifically proven. They could therefore to a greater extent be used as role models and as inspiration in today's construction of digital resources and platforms in digital humanities.

Keywords: Medieval Sweden; Settlement history; Reference work; XML; Text Encoding Initiative