

# Towards a Union Catalogue of Encoded Manuscript Descriptions

*Eric Haswell<sup>1</sup>, Matthew Driscoll<sup>2</sup>*

<sup>1</sup>Humanities Computing and Media Centre, University of Victoria  
P.O. Box 1700 STN CSC, Victoria BC V8W 2Y2, Canada  
e-mail: ehaswell@gmail.com

<sup>2</sup>Department of Scandinavian Research, University of Copenhagen  
Njalsgade 136, 2300 Copenhagen, Denmark  
e-mail: mjd@hum.ku.dk

## Demonstration

The recent work of the Text Encoding Initiative (TEI) in developing an XML standard for manuscript description has provided an opportunity to explore methods of making such information more widely available through a web-based system. While traditional web database applications built on relational systems such as MySQL are well-suited for web delivery of data-centric, tabular information, they are not appropriate for heavily marked-up, document-centric information such as manuscript descriptions. The emergence of native XML databases as a viable storage mechanism has made web delivery of XML-encoded material easier and more efficient.

The presentation discusses a prototype system developed as part of a research project carried out at the University of Copenhagen's Arnamagnæan Institute in the summer of 2005. The Arnamagnæan Manuscript Collection, which derives its name from the Icelandic scholar and antiquarian Árni Magnússon (1663-1730), comprises the single most important collection of early Nordic manuscripts in the world, in all some 3000 items, the bulk of them Icelandic. Following independence from Denmark in 1944, Iceland requested the return of the Icelandic manuscripts in Danish collections, and after much heated debate the Danish parliament agreed to transfer a significant number of manuscripts from the Arnamagnæan Institute -- in the end slightly over half -- to a newly established sister institute in Reykjavík. The process of transference took over 25 years, with the last two manuscripts handed over in June 1997.

At about that same time representatives of the two Arnamagnæan institutes began looking into the possibility of reuniting the two halves of the collection virtually, in particular through the development of a searchable web-based catalogue. The Arnamagnæan Institute in Copenhagen participated in the MASTER project, an EU-funded project whose goal was to define and implement a general purpose standard for the description of manuscript materials using XML. The project period ran from January 1999 through June 2001, during which some 500 records were produced in Copenhagen. In Iceland basic cataloguing began in the year 2000. Since then work has proceeded apace at both institutes, if at times only slowly, and there are now at least minimal records for the entire collection. There has not, however, until now, been a suitable system for searching and browsing these records, and the goal of the research project was to create such a system.

Using PHP and the eXist native XML database, a three-tier web database application was developed. Users are provided with a facility for executing queries on the database through a web form designed to allow complex query formulation involving many different criteria. User input from a search form submission is processed by PHP into an XQuery expression which is then passed on to the database. Results are returned to the user after being processed by an XSLT engine on the server. The web system is multi-lingual and places an emphasis on usability, standards-compliance and the use of open-source software.

The realisation of this resource demonstrates a method by which other institutions may undertake similar projects involving XML-encoded source material. Because the manuscript records are encoded in standardised TEI XML and have a known and consistent structure, the potential exists for integrating records from this project with those from other collections, thereby creating a larger and more complete catalogue. Indeed, this could easily be the first step toward a (virtual) union catalogue of manuscripts in European repositories. The large number of available search criteria, and the ability to combine these criteria in complex ways, allows researchers to assemble datasets which may otherwise have been difficult to gather. Given that there are clear limitations on the number of researchers able to physically view a manuscript due to constraints of time, funding and manuscript fragility, providing electronic access benefits researchers significantly. There also exists the

possibility that researchers may discover useful and interesting information which they had previously not even considered. There is, therefore, a potential to change the way people do research into manuscripts.

The goal of creating a virtually unified catalogue of European medieval manuscripts may demand some measure of a standardised approach to encoding, as the ability to program query functionality is dependent on data that is structured in a similar manner across all documents being queried. Despite the availability in TEI P5 of a general tagset for encoding manuscript descriptions, the number of possible combinations of elements and different stylistic approaches to encoding present some obstacles to total integration. Surmounting the challenges imposed by encoding irregularities may be possible, however, through the use of query techniques which accommodate these differences. More work is needed to assess the implications of such an approach. It is also uncertain at this stage how feasible it is to expect the system to smoothly scale upwards as the number of documents increases. Initial indicators are positive, but a more rigorous case-study is required.

The development of a union catalogue is therefore dependent on the availability of a technical infrastructure of sufficient flexibility and reliability. The work done in developing the prototype, and the positive results from it thus far, suggest that this is indeed possible and within reach. The eXist XML database is a viable option for storage and document management. XML documents need only to be uploaded to the database in their complete form to be added to the collection, greatly simplifying management and allowing for participating repositories to be widely dispersed geographically. A fully-functioning deployment would require a centralised server and some direct coordination of the system and its various collections, but these are logistical matters that can be readily addressed with the provision of adequate funding and, more importantly, the enthusiastic participation of manuscript repositories.

The presentation will comprise a demonstration of the web resource, including examples of XML source material, the eXist database system, PHP code used to build the application, and the web interface.