

Creating XML-Based Tools for Academic Departments: Lessons from the Early Modern Spain Project

Paul Spence (paul.spence@kcl.ac.uk)
King's College London (London, United Kingdom)

PAPER

In recent years XML has been used extensively across text-focused humanities computing projects that embrace a wide variety of academic aspirations, ranging from digital edition to linguistic analysis and beyond. In particular, the Text Encoding Initiative (TEI) guidelines have become a key point of reference for any modern humanities-focused scholar who plans to digitise text, and there is a very active community around TEI that is not only keen to engage with new markup challenges but is increasingly in the forefront in creating and sharing new tools that can actually leverage the potential benefits of deep scholarly encoding in XML.

Humanities computing scholarship has been influential in the development of XML while benefiting from its emergence as a universal standard for data storage and from the transformational possibilities of its sister technology XSLT, but to what extent has it really taken advantage of XML's enormous potential for data interchange? Many XML-based projects achieve interesting results within their initial research parameters, but how often is their data integrated more widely into their immediate academic context or indeed made more widely available?

Three years ago, the Centre for Computing in the Humanities at King's College London began a pilot project to explore the extent to which some of the classic scholarly activities of an academic department could be represented using an XML-based architecture. This project, called *Early Modern Spain*, focused on one of the major research areas within the Spanish and Spanish American department at King's College London, namely the literature, culture and history of the Spanish Golden Age. It aimed to integrate the work of King's scholars, their collaborative projects, publications, teaching materials and electronic editions of primary source texts from the period.

The first phase of the website at <http://www.ems.kcl.ac.uk/> includes twenty electronic versions of primary texts, over two hundred bibliographical entries relating to the participating scholars (with many publications available in digital form in Spanish and/or English) and information on five core research areas. There are some experimental uses of XML to generate a text-analysis interface for some texts. The project also drew widely on our broader research into a generic XML-based publishing tool called xMod (<http://www.cch.kcl.ac.uk/xmod/>).

Inevitably, in an ambitious project such as this, there has been more progress in some areas than in others, but the use of XML has significantly enhanced the integration of, and navigation between, data at both 'document' and 'intra-document' levels. At 'document' level, users are able to move effortlessly from one area to another, and to view the same source data re-arranged in such a way that the focus can be on any one of the main themes of the site: scholar, publication, teaching programme, research project theme or digital version of a primary text.

At a deeper level, in one experiment, we carried out some text analysis of two texts that both describe a disastrous voyage to Florida in the years 1527-1537, led by Pánfilo de Narváez. Using XML markup on *Historia general y natural de las Indias* by Gonzalo Fernández de Oviedo y Valdés and *Naufraios* by Alvar Núñez Cabeza, we produced some comparative results for themes of scholarly interest such as religious terminology, cognates of hunger and forms of the verb 'comer'. This research fits neatly within the overall architecture of the site and uses an approach tested on the *Aphrodisias in Late Antiquity* project which takes advantage of deeper encoding across project resources. There is also the potential to provide broader access to research themes within the project documents as a whole in the future.

Using EMS as a model, I will assess the benefits and drawbacks of using an XML-based approach in creating integrated departmental resources. How do the flexibility and deep encoding possibilities of XML compare to the ease of use and collaborative facilities of a Content Management System, for example? Are the tools for XML robust enough to rely on XML primarily as the data source? How feasible, or indeed useful, is it to integrate research projects that require deep encoding with more general information about the scholarly activities of a department? What effect do multilinguistic resources have on the implementation?

Drawing more broadly on experience from some of CCH's recent involvement in over 30 projects, I will also describe attempts to ensure that data may be shared more broadly with other projects of a similar nature, in such a way that a project that uses relational database technology can exchange data with a TEI XML-focused project.

Finally, I will appraise the potential for creating a truly generic departmental tool, placing the discussion within the context of recent XML-related developments (XSLT 2.0, xQuery), the next release of the TEI guidelines (P5) and broader developments (such as Topic maps, web services and metadata standards) that will increasingly play a part in text markup projects.

References

Centre for Computing in the Humanities. <<http://www.kcl.ac.uk/humanities/cch/>> (24 April 2006)

xMod. <<http://www.cch.kcl.ac.uk/xmod/>> (26 October 2005)

Early Modern Spain website. <<http://www.ems.kcl.ac.uk/>> (28 July 2005)

Text Encoding Initiative. <<http://www.tei-c.org/>>

TEI P5. <<http://www.tei-c.org/P5/>>

Aphrodisias in Late Antiquity: The Late Roman and Byzantine Inscriptions. <<http://www.insaph.kcl.ac.uk/a1a2004/>> (2 November 2005)