Europe's Beginnings through the Looking Glass: Publishing Historical Documents on the Web Using EVT

Roberto Rosselli Del Turco
Università di Torino
roberto.rossellidelturco@unito.it

Florentina Armaselu
CVCE
florentina.armaselu@cvce.eu

Chiara Di Pietro
Università di Pisa
dipi.chiara@gmail.com

Lars Wieneke
CVCE
lars.wieneke@cvce.eu

Raffaele Masotti
Università di Pisa
raffaele.masotti@gmail.com

EVT (Edition Visualization Technology) [1] is a software aimed at the creation of image-based web editions of TEI P5 encoded texts. It is a light-weight, open source tool specifically designed to simplify the production of digital editions, freeing the scholar from the burden of web programming and enabling the user to browse, explore and study digital editions by means of a user-friendly interface, providing a set of tools (zoom, magnifier and hot-spots for manuscript images, an internal search engine for the edition texts) for research purposes.
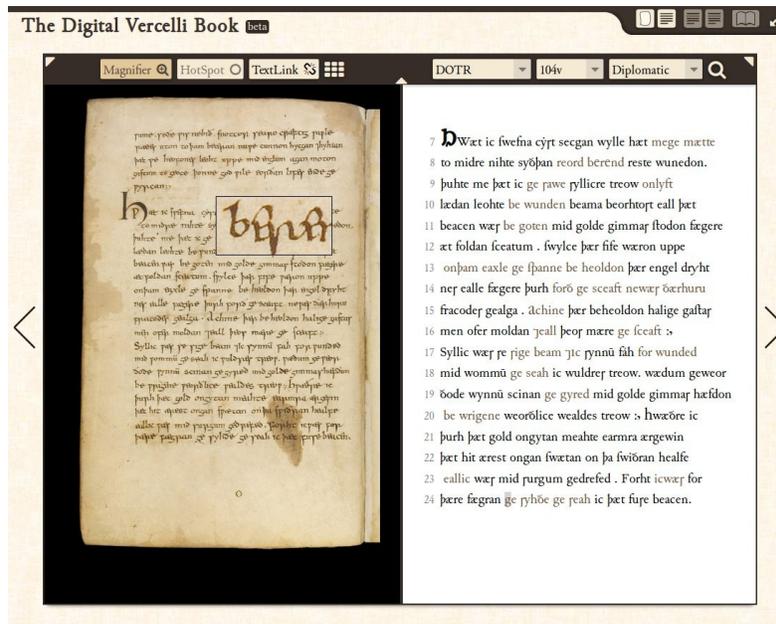

Figure 1: The Digital Vercelli Book edition using EVT

Everything is created around the data and the encoded text itself: by applying a single style sheet to the TEI XML file that contains the whole transcription of a document, an XSLT 2.0 transformation chain is started that results in a web based application – a mix of HTML 5, CSS3 and JavaScript – that can be easily shared on the Web. Besides presenting the digital scans of the original manuscript (if available) linked to the corresponding text of the edition, the software provides a bookreader visualization mode if double side images are supplied.

EVT was born in the context of a specific use case (the Digital Vercelli Book project, whose first version has been available online for about a year [2]), but it is now being used to publish another digital edition, that of the Codice Pelavicino manuscript [3]. The need to adapt it to different types of documents has led to a revision and expansion of the underlying code to make it more flexible and suitable for many different types of TEI-encoded texts. With this proposal we want to demonstrate the flexibility and re-use of EVT by applying it to an edition of diplomatic documents to be published on the CVCE's Web site [4]. We will discuss both the technical implications of its application as well as methodological consequences.
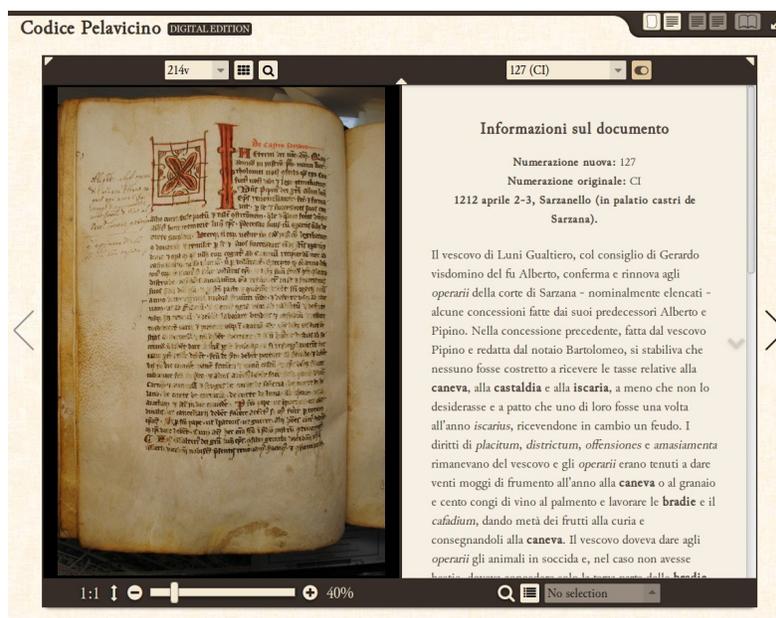

Figure 2: The digital edition of the Codice Pelavicino manuscript

The intended digital edition we plan to build using EVT is based on XML-TEI P5 bilingual (French, English) documents of the W.E.U. (Western European Union), concerning armament production, standardisation and control in the period 1954-1982. The corpus was selected from the Luxembourg National Archives, W.E.U. collection, and implied: OCR processing with ABBYY FineReader (one image file per typewritten page), Microsoft Word styling and OxGarage [5] conversion to XML-TEI

P5, as well as semi-automatic enrichment by XSLT 2.0, Named Entity Recognition with GATE [6] and manual annotation.

Several types of documents compose the corpus: meeting minutes, notes from the Secretary-General, memoranda and studies. The encoding includes metadata (title, author, document reference, copy number, version, date of availability, confidentiality status); structural markup for headers/footers, sections/subsections, paragraphs, line breaks; content-related annotations (discourse of institutional/country representatives, names of persons, organisations, etc.).

A digital edition will imply features for browsing and searching the collection, as well as side-by-side visualisation of the transcribed pages and facsimile images. The presentation would address technical issues (adjustment of the EVT framework to generate the Web edition from the XML-TEI corpus) and more general questions (to what extent the XML-TEI encoding and associated technologies may support enhanced "mirror-like" digital representations of the original documents).

The presentation will end with a short demo of the digital edition, based on the documents that have been digitized and encoded.

References
[1] EVT home page: https://sourceforge.net/projects/evt-project/
[2] Digital Vercelli Book beta edition: http://vbd.humnet.unipi.it/beta/
[3] Codice Pelavicino digital edition: http://labcd.humnet.unipi.it/
[4] CVCE (Centre Virtuel de la Connaissance sur l'Europe): http://www.cvce.eu/
[5] OxGarage: http://www.tei-c.org/oxgarage
[6] GATE (General Architecture for Text Engineering): https://gate.ac.uk/