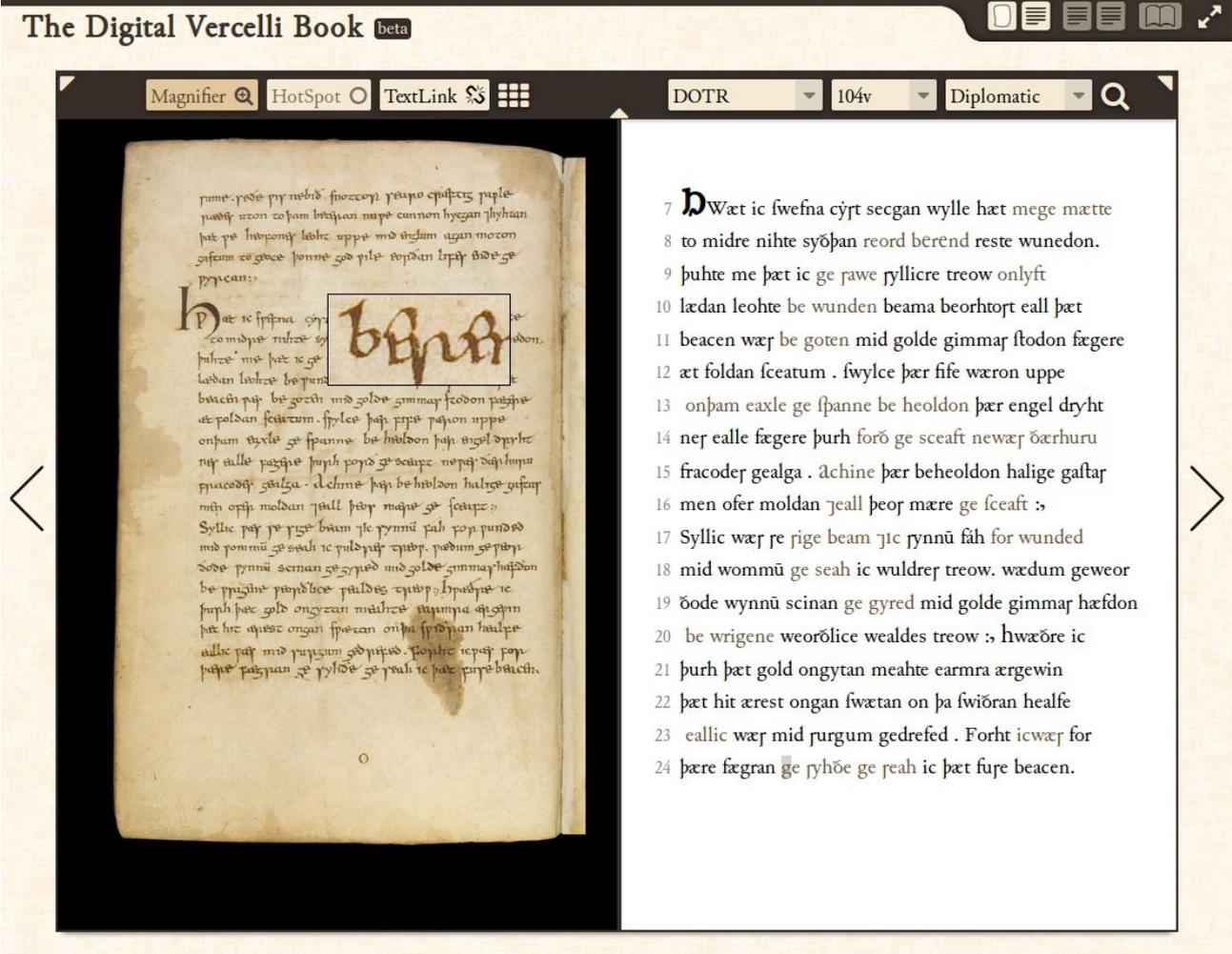


Europe's beginnings through the looking glass: publishing historical documents on the Web using EVT

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.

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.



The Digital Vercelli Book beta

Magnifier 🔍 HotSpot ○ TextLink ℹ️

DOTR 104v Diplomatic 🔍

7 Hwæt ic swefna cýrt secgan wylle hæt mege mætte
8 to midre nihte syðþan reord berend reste wunedon.
9 þuhte me þæt ic ge rawe syllicre treow onlyft
10 lædan leohte be wunden beama beorhtoft eall þæt
11 beacen wæx be goten mid golde gimmar stodon fægere
12 æt foldan sceatum . swylce þær fife wæron uppe
13 onþam eaxle ge spanne be heoldon þær engel dryht
14 nef ealle fægere þurh forð ge sceaft newæf ðærhuru
15 fracodef gealga . Achine þær beheoldon halige gaftar
16 men ofer moldan 7eall þeof mære ge sceaft :
17 Syllic wæx fe fige beam 7ic fynnu fah for wunded
18 mid wommū ge seah ic wuldref treow. wædum geweor
19 ðode wynnū scinan ge gyred mid golde gimmar hæfdon
20 be wrigene weorðlice wealdes treow : hwæðre ic
21 þurh þæt gold ongytan meahthe earmra ærgewin
22 þæt hit ærest ongan swætan on þa swiðran healfe
23 eallic wæx mid furgum gedrefed . Forht icwæx for
24 þære fægtran ge fyhðe ge feah ic þæt fupe beacen.

The Digital Vercelli Book edition using EVT

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.



The digital edition of the Codex 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).

Authors

Chiara Di Pietro (dipi.chiara@gmail.com) graduate student, University of Pisa

Raffaele Masotti (raffaele.masotti@gmail.com) graduate student, University of Pisa

Roberto Rosselli Del Turco (roberto.rosselidelturco@unito.it) Researcher/Assistant Professor, University of Turin

Florentina Armaselu (florentina.armaselu@cvce.eu) Researcher, Digital Humanities Lab, CVCE

Lars Wieneke (lars.wieneke@cvce.eu) Head of Information & Technology, CVCE

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/>