

The digital library DigilibLT (short paper)

Nadia Rosso¹

¹ Università del Piemonte Orientale, Via Galileo Ferraris 107-109, Vercelli, 13100, Italy

Abstract

The aim of this paper is to describe the main strengths of the digital library DigilibLT: an open work environment that serves first of all as a reference point for research ends and secondly as an interdisciplinary developmental platform for young students; it is a hive of ideas that contributed to the creation of new projects, such as TBL and Vertepε, that both share the winning choice of using the universal language XML-TEI. This project proves that the successful union between information technology and humanistic studies generates great scientific and didactic advantages in relation to the sharing of ideas, collaboration and communication.

Keywords

DigilibLT, Digital Libraries, Late Antiquity

1. Introduction

DigilibLT is an acronym for Digital Library of Late Antique Latin Texts.

The DigilibLT project, conceived by Raffaella Tabacco and developed in collaboration with Maurizio Lana², was officially founded on March 1th, 2010, thanks to the funds granted by Regione Piemonte in support of research in the area of Humanities and Social Sciences.

The digital library³ was born with the aim of producing a complete digital library of Late Antique Latin literary texts in prose (2nd-7th century AD) and today it is a reference point for Late Antiquity studies, representing a reference workspace for scholars. In particular, DigilibLT aims to continue the PHI, a database of Latin literary texts from the origin to the second century AD produced by the Packard Humanities Institute in the 1990s, issued in a series of CD-ROM releases. Late Antique Latin literary texts are sporadically available online up until 2010. DigilibLT texts, annotated according to XML-TEI standards and encoded according to Unicode, can be downloaded freely in different formats (TXT, HTML, PDF, EPUB). In addition, DigilibLT offers short encyclopedic entries on Late Antique authors and works, constantly updated bibliographies, a newsletter, free download of out-of-copyright essays or monographies whenever permission is granted by the legal owners of the publication rights, and a feedback section. However, one of the greatest advantages of DigilibLT is that it fills an important gap: on the one hand it offers a clear and accurate canon of Late Antique literature, including detailed information on the adopted critical editions, in response to the need to supplement and update *Thesaurus Linguae Latinae Index*, and on the other hand it indicates any editorial changes carried out, providing a frequent enrichment by the correction of misprints. DigilibLT allows users to narrow their research with the use of different filters: besides searching the entire collection, they can filter their research specifying a selection of authors, time period or type, or selecting individual authors and works. DigilibLT texts can be consulted by the users either through a simple search accessed on the homepage or through an advanced search of textual portions using asterisks and Boolean operators.

Proceedings TPDL2022: 26th International Conference on Theory and Practice of Digital Libraries, 20-23 September 2022, Padua, Italy

EMAIL: nadia.rosso@uniupo.com

ORCID: 0000-0003-0225-7427



© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

Figure 1: DigilibLT Advanced search

2. DigilibLT: an open work environment

DigilibLT is not only a database, it is also an environment for work and research that can be defined for all intents and purposes “open”: firstly because it offers scholars, students and enthusiasts the possibility to take part in the project through productive collaborations or simply by filling out the dedicated form in the ‘*Contact and feedback*’ page of the website in which they can make suggestions, offer collaborations, report mistakes found in the texts or suggest integrations to the *corpus*; in more than one occasion the digital library was enriched thanks to the network of national and international contacts that was created through this tool. Secondly, it is open because the whole *corpus*, complete with its explanatory lists of contents and correlated critical problems and its constantly updated bibliographic records, is freely accessible to the users thanks to the Creative Commons license.

The choice of the XML-TEI language was determined by the openness dimension of the library: this language is not subjected to a rapid obsolescence of technological standards, unlike the initial digital resources distributed on CD-ROM (such as GL-CD, the whole *corpus* of grammatical works collected in the early 1970s in a CD-ROM by Nino Marinone, a pioneer in the field of Digital Humanities); besides, it does not require the installation of any specific software, unlike other applications like Diogenes. Lastly, it is an extremely versatile universal language that allows a productive communication among different digital libraries, expanding the diachronic and synchronic research outlook. The extent of the *corpus*, made up of more than 600 texts, together with the heterogeneous nature of these works, determined the choice of a structural markup that takes into consideration the philological characteristics, the use of diacritics, the presence of symbols, numerals with multipliers, and unit of measurements, and the presence of the Greek language.

3. DigilibLT: a place of training

Sharing, collaboration and communication are the key words of the digital library, that since its creation in 2010 has been a platform for development and research. Choosing to manage internally all the processing stages, without delegating to experts or external companies, has been a successful move: the research group acquired specific skills in different fields, granting longevity to the project on one hand and enriching the research team with a solid and polyhedric experience marketable on different levels on the other hand. The full team is all equally involved in all the development stages of the project in accordance with the level of expertise of each member: digitization, OCR process, post OCR text proofing, markup, updating of bibliography and short entries, TEI and XML training events, conferences and seminars, problem solving, project management.

The DigilibLT training process initially involved the research team managers, who by attending and organizing specific publishing and digital philology workshops acquired the necessary skills to take care of the whole process, from the scanning stage to the online publication, and in the following years they were able to share their knowledge with the other team members and collaborators according to their interests. The encounter with the Digital Humanities field, even if it requires constant dedication and commitment, completed and enriched the traditional classicist knowledge of many of the project members, proving to be essential for the extension of the library itself and for the creation of new projects connected with it. Over the years DigilibLT invested a lot in young students, and it still does today. From 2015 the DigilibLT work environment initiated a collaboration with secondary schools. The most deserving students attending the last two years of the classical high school (*liceo classico*) have the possibility to choose DigilibLT among the activities for the school- work alternation (PCTO) for a negotiable amount of hours that usually ranges from 100 to 60, allowing the students to make their first experience in the field of the Digital Humanities. Every year about fifteen students take part in this activity working in the DigilibLT laboratory, mostly over the summer period, actively contributing to the publication online of the Latin texts of the *corpus*. Under constant supervision of a tutor, after witnessing the scanning stage and learning how to operate a professional planetary scanner, they devote their time to the reading and guided revision of one or more texts, according to the length of the document and to the amount of hours negotiated, drawn from their preferred literary genre – the wide variety of genres and topics allows to satisfy the specific interests of each student. More specifically, after learning how to use an OCR (Optical Character Recognition) software, trained for the Latin language, and converting the text in a RTF file, the student proceeds to the first reading of the document, paying attention to the text structure and eliminating the optical reading errors, whose quantity is variable according to the edition of the work – some works still refer to editions dating back to the late 18th century printed with stamp characters that greatly increase the amount of errors; after the first reading, the student proceeds with the second reading in which time is dedicated to the philological interpretation of the text, identifying and aligning the diacritics, and to the correction of potential misprints, indicating the amendment in the digital title page of the work, acquiring not only philological skills in support to the reading, but also linguistic-grammatical skills in the Late Antique Latin language. Lastly, the student can take a look at the markup used by the DigilibLT project for the examined text, learning the basic rudiments of the XML-TEI language. The active involvement in all the development stages of the digital library is also continued by the graduands of the Università del Piemonte Orientale and of other universities, that choose to carry out their internship in the DigilibLT laboratory – every year about 4 to 6 students choose to carry out their internship in the university laboratory, acquiring and developing skills in the field of Digital Humanities. This activity is also carried out in person under the supervision of a tutor, which in this case can be less attentive considering the experience accrued by the students. The amount of hours dedicated to the activity is greater – usually between 175 and 150 hours. In order to give an idea of the time required by the revision, it was calculated that during the first reading the students examine an average of 4 pages per hour, while during the second reading the average increases to 8 pages per hour. However, the collaboration with the graduands is not limited to the internship activity: along the years students have been assigned bachelor's and master's thesis investigating the works contained in the DigilibLT *corpus*, some with a purely philological-literary end and others intended to have a more interdisciplinary approach towards Digital Humanities, resulting in descriptive or experimental thesis. Over the years students from other universities have chosen to

analyze this digital library as the main focus of their thesis, which was a great opportunity for enrichment, not only for the students, but also for the University, thanks to the critical input provided by attentive users. The DigilibLT project allowed many young students to grow, co-financing PhD programs and through these educating PhD students that in some cases continued their academic path in the library project. This is mine and Alice's case, whose start of the academic path coincided with the founding of the library, a fruitful research environment that allowed us to grow joining a stimulating research group, thanks to which the work turns into pleasure, the coworkers into friends and the individual goals become shared objectives. From 2010 up until today about fifteen PhD students from our PhD course and from other courses took part in the project, contributing to the library for different periods of time according to their skills and interests. The development of young scholars, from the start of the project, invested in numerous research grants, expanding the research team and granting continuity to the young students that grew and keep growing within the library. First of all, the postdoctoral researchers, like myself, have been assigned the coordination of the different stages of the project development, from the scanning phase to the online publication, expanding their skills to the management field, assigning the tasks based on the specific skills of each of them. Secondly, they have been assigned the education of new young students. One of the strengths of the DigilibLT project is the choice of giving the postdoctoral researchers a 360 degrees education, in order for them to be able to complete all the activities connected with the project and to continuously devote themselves to the project with enthusiasm and dedication. Since its founding the DigilibLT project offers numerous Research Training scholarships (*borse di addestramento alla ricerca*) for specific activities of reading and revision, XML-TEI annotation and author and work records writing to the most deserving graduands and PhD students. Involving students in building the digital library has the great advantage of enriching the CV of a humanist graduate with significant technical competences, marketable in many different areas, such as the publishing industry.

4. DigilibLT: a hive of ideas

Coming up with a digital library as a development environment contributed to new ideas withing the research group, which thanks to the versatile skills accrued in the field of Digital Humanities have been translated into financed innovative and interdisciplinary projects. This allowed the research group to look ahead not only drawing from the already existing digital libraries and communicating with other projects, but also creating new types of approach to expand the study perspectives. A complete corpus of the Late Antique Latin works in prose, together with the complete development of the group, transformed the DigilibLT library into a promising and fruitful hive of ideas. As an example, two of the projects developed while working on the DigilibLT library are described below.

4.1. TBL

An example of this new study perspective is TBL⁴, acronym for Textual Bilingualism in Latin, a project born in 2015 at the Università del Piemonte Orientale, coordinated by Professor Maria Napoli, financed by the Fondazione Compagnia di San Paolo after a competitive tendering, which saw the creation of a database with the aim to investigate the phenomena of linguistic contact between Greek and Latin in the literary texts of the Late Antiquity with a strong interdisciplinary approach among linguistics, Latin literature, Greek literature and Digital Humanities. TBL's point of departure was J. N. Adams' landmark book *Bilingualism and the Latin Language* (Cambridge University Press, 2003), which offered the first large-scale treatment of bilingualism in the ancient Roman world. Adams provided an extensive survey of contact phenomena between Latin and other languages in the Roman period, from the early Republic to the Late Empire (approx. the 4th century). However, his *corpus* is primarily made up of epigraphic material (such us papyri, inscriptions and *ostraca*), and does not include literary texts (with a few exceptions). The term "textual bilingualism" indicates the totality of

the various manifestations of the use of the Greek language in literature, connected to the specific competence level of the different Late Antique Latin authors in the spoken and written Greek language, approximately from the 3rd to the 4th century AD, drawing from a variety of texts in the DigilibLT digital library, integrated with a specific XML-TEI annotation for bilingualism elaborated to classify the multiple contact phenomena between Greek and Latin – Grecisms, transliterations, hybrid terms or code-switching situations, scenarios in which there is a switch from the Latin to the Greek language, just to name a few. Based on the presence of these phenomena in a work, the competence level in the Greek language of the author can be evaluated and classified on three degrees: lexical, morphological and syntactic; the highest linguistic competence is recognized when the bilingual phenomena also involve the syntax.

Bilingual phenomena	Tags
Greek word in Greek characters	<foreign xml:lang="grc-Grek">...</foreign>
Greek word in Latin characters	<foreign xml:lang="grc-Latin">...</foreign>
Inter-sentential code-switching	<foreign ana="cdswinter">...</foreign>
Intra-sentential code-switching	<foreign ana="cdswintra">...</foreign>
Inter and Intra sentential code-switching	<foreign ana="cdswinterintra">...</foreign>

Figure 2: TBL List of the tags

4.2. Vertere

The consultation of bilingual texts, specifically the *Hermeneumata pseudodositheana*, about ten Late Antique texts that reached us through various reviews, that can be described as small travel dictionaries similar to the ones that are still used nowadays when visiting a foreign country, in 2018 originated the idea for another innovative project, *Le strategie del uertere in età tardoantica*, financed by the CRT after a competitive tendering. As is known, during the Late Antiquity the Greek language becomes harder to understand in the occidental part of the Empire, requiring translations even for the most educated social classes; some translations into Latin, specifically of works in the treatise genre, became fundamental for the cultural and scientific development of the Empire. The project investigates the translation from Greek into Latin in the Late Antiquity, with the aim of creating a parallel *corpus* of the antique translation from Greek into Latin annotated in XML-TEI. The digital text allows not only a precise classification and individuation of the phenomena defined by the translation studies, that can be searched both in the individual texts and in the *corpus*, but it also grants the full access through a screen view, free from the limitations of the paper form, and by placing the original version side by side with the translation the correspondences are highlighted. The same intent led to the development of a program for the automatization of the recognition of word Greek-Latin textual correspondences between hypotext and hypertext through parallel concordances, based on a pre-set inventory of correspondences between words and expressions in the two languages.

Vertepe Home Progetto




Quest'opera è distribuita con Licenza Creative Commons Attribuzione - Condividi allo stesso modo 3.0 Italia.

Le strategie del vertepε in epoca tardo antica

Vertepe è una banca dati elaborata con lo scopo di indagare il fenomeno della traduzione dal greco al latino nei testi prosastici della tarda latinità (approssimativamente dal III al VII sec. d.C.) secondo un nuovo metodo di analisi della traduzione antica che si avvale del fondamentale contributo delle Digital Humanities e degli Studi di traduzione.

Questa risorsa rappresenta il primo risultato del nostro progetto realizzato presso il Dipartimento di Studi Umanistici dell'Università del Piemonte Orientale e finanziato dalla Fondazione CRT.

La banca dati, che può considerarsi il primo corpus parallelo della traduzione tardoantica, offre ipotesti greci e ipertesti latini, annotati in XML-TEI, di libera e gratuita consultazione, con disposizione interlineare o parallela.

Sul sito Vertepe è consentita sia la ricerca semplice sia la ricerca avanzata delle strategie di traduzione individuate nei testi tardoantichi; inoltre, il sito offre a libera disposizione degli utenti le seguenti risorse: un canone completo delle traduzioni tardoantiche con l'indicazione dell'edizione di riferimento sia degli ipertesti latini sia degli ipotesti greci; le note di trascrizione e di codifica dell'annotazione XML-TEI elaborata specificamente per la traduzione antica; un programma di riconoscimento automatico delle corrispondenze greco-latine.

Pagina in costruzione

Figure 3: Vertepε Homepage

5. DigilibLT: a growing digital library

The digital library is constantly growing: the implementation concerns not only the number of texts already uploaded (almost 400), but also the variety of genres – in fact in 2017 DigilibLT added grammatical works and by 2019 it also started to include legal texts, thanks to the support of the Fondazione Cassa di Risparmio di Vercelli. As illustrated in the previous paragraphs, DigilibLT is a lively project in expansion based on sharing and communication.

6. References

- [1] A. Borgna, S. Musso, *Le sfide di una biblioteca digitale del latino tardoantico. Struttura, canone e questioni aperte di codifica e visualizzazione*, in: P. Mastandrea (Ed.), *Strumenti digitali e collaborativi per le Scienze dell'Antichità*, Edizioni Ca' Foscari, Venezia, 2017, pp. 69-94. doi: <http://doi.org/10.14277/6969-182-9/ANT-14-06>.
- [2] C. Fedriani, M. Napoli, N. Rosso, *Forms and functions of Greek words in Late Latin literary texts: a corpus-based approach*, in: N. Holmes, M. Ottink & J. Schrickx (Eds.), *Lemmata Linguistica Latina. Words and Sounds*, vol. I, De Gruyter (ISBN: 978-3-11-064518-7), 2019, pp. 294-311. doi: <https://doi.org/10.1515/9783110647587-019>.
- [3] C. Fedriani, M. Napoli, N. Rosso, *Un nuovo corpus per lo studio del bilinguismo testuale in latino tardo. Aspetti metodologici e primi risultati*, in: G. Haverling (Ed.), *Studies on Late and Vulgar Latin in the Early 21st Century: Acts of the 12th International Colloquium "Latin vulgare - Latin tardif"*, (Studia Latina Upsaliensia 37), Acta universitatis Upsaliensis, Uppsala 2022. To appear.
- [4] M. Lana, *Biblioteche digitali. Un'introduzione*, Bononia University Press, Bologna, 2012. ISBN: 9788873957942.

- [5] M. Lana, *Metodologie e problematiche per una biblioteca digitale. Il caso digilibLT*, *DIGItalia*, «Rivista del digitale nei beni culturali» 7 (2012) 40-64. URL: <https://digitalia.cultura.gov.it/article/view/516/355>.
- [6] M. Lana, *Da una digital library del latino tardo ad un corpus globale*, in: F. Ciotti, G. Crupi (Eds.), *Dall'informatica umanistica alle culture digitali, Atti del convegno di studi in memoria di Giuseppe Gigliozzi, Roma 27-28 ottobre 2011* (ISBN: 9788895814827), Digilab, Roma, 2012, pp. 134-150.
- [7] S. Musso, *La marcatura di testi latini tardoantichi. Un compromesso ragionevole*, «Aevum Antiquum» 11 (2014) 151-187. doi: 10.1400/225669.
- [8] T. Orlandi, *Informatica testuale. Teoria e prassi*, Laterza, Roma-Bari, 2010. ISBN: 8842093793.
- [9] N. Rosso, *Vertepe: una proposta di annotazione semantica XML-TEI della traduzione antica*, «FuturoClassico» 6 (2020) 136-152. doi: <https://doi.org/10.15162/2465-0951/1170>.
- [10] R. Tabacco, *La tarda antichità latina tra i codici e il web*, «Aevum Antiquum» 11 (2014) 3-17. ISSN: 1121-8932.
- [11] R. Tabacco, *La prosa latina pagana tardoantica e la biblioteca digitale digilibLT*, in: C. Marazzini, L. Maconi (Eds.), *L'italiano elettronico. Vocabolari, corpora, archivi testuali e sonori, Atti del Convegno "Piazza delle lingue", VIII Edizione, Firenze, 6-7 novembre 2014* (ISBN: 9788889369654), Edizioni dell'Accademia della Crusca, Firenze, 2016, pp. 125-141.