

Regione Piemonte - Bando Scienze umane e sociali

TITLE:

Digital Library of Late-Antique Latin Texts

RESPONSIBLE

RAFFAELLA TABACCO, Università degli Studi del Piemonte Orientale, Dipartimento di Studi Umanistici

GENERAL INFORMATION

ACRONYM: digilibLT

SECTOR: Scienze dell'Antichità, Filologico-Letterarie e Storico-Artistiche

OTHER SECTORS: Valorizzazione e Tutela del Patrimonio Artistico e Culturale,

ABSTRACT:

The aim of the project is the creation of an open collection of textual data and the publication of such data on the Web. The project is designed to be open to the contribution of other scholars, not simply in the three years of its development, but also in the future.

The project digilibLT aims to continue a similar project by the Packard Humanities Institute (PHI): a digital library of Latin literary texts. The so-called PHI CD-ROM 5.3 offers a very important but selective collection of Latin texts and is the most widely used database of Latin texts for the earlier period (from the origins to the I century AD). In order to complete the database, our project is planning to digitize the Latin literary texts written in the period from the second to the fifth century AD. The texts will be tagged according to both state-of-the-art standards (XML), and to the standard adopted in the PHI CD-ROM 5.3. The digital library digilibLT will also include a number of important critical studies and editions which are now out of copyright.

The final product is going to be a rich, accessible website, which will offer: digitized latin texts; bio-bibliographical notes about authors and works;

digitized historical, critical, literary studies about the late-antique Latin authors and works; full text search in the digitized texts (also with an interface specifically designed for mobile devices);

selection and download of any text, by content criteria (word search) or by filing criteria (genre, period, and so on).

DURATION (months): 36

TECHNICAL SCIENTIFIC OBJECTIVES: The project digilibLT aims to continue a similar project by the Packard Humanities Institute (PHI): a digital library of Latin literary texts.

The so-called PHI CD-ROM 5.3 offers a very important but selective collection of Latin texts and is the most widely used database of Latin texts for the earlier period (from the origins to the I century AD). The Packard project has been closed and the digitization of Latin texts has stopped at the II century A.D. For late-antique Latin texts (II-V century AD) no comprehensive

textual database exists and the scholarly community must depend on incomplete and unreliable digital resources which have been sometimes prepared without critical study of the texts. Late antique culture in particular is understudied, even if it was especially important in transmitting the cultural, scientific, political, philosophical, literary and religious thought of antiquity to modern Europe.

The project DigilibLT will then digitize the texts of II-V century AD and will tag them according to both state-of-the-art standards (XML), and to the standard adopted in the PHI CD-ROM 5.3 in order to allow the scholars to continue using the existing PHI-oriented text analysis tools (such as Diogenes, Musaios, and others). The digital library will also include a number of important critical studies and editions which are now out of copyright.

The scholars who study the classical world, who make use of digital tools for their research and who are interested in the development of such tools constitute a large and very active community, and are especially attentive to the communication and the exchange of experiences and information. The scholarly community will see the continuation and completion of the collection as an essential tool for the study and the valorisation of Latin culture. Latin civilization originated in the geographical setting we now call 'Italy', but has a worldwide relevance for its cultural, literary, and political impact.

STATE OF THE ART:

The research team is aware that information technology is especially effective at collecting and analysing texts and at spreading them; and also that the worldwide impact is especially great in the case of classical texts (Latin, Greek). A very large international community of scholars works on these cultures and languages, and the interest on it is ever expanding.

Scholars and students interested in digital Latin texts, apart from the PHI CD-ROM 5.3 mentioned above, for texts written between the II and the V century AD don't find a single database in plain text format. They can find small and incomplete on-line collections dedicated to specific authors or genres; in alternative, they find digital collections which reproduce print collections (see in particular CSEL and Bibliotheca Latina Teubneriana) and which unfortunately present text in proprietary or crippled format. Scholars have often remarked on the need for an open-access database in text format; which will leave open the possibility of choosing the preferred software for textual searches, or even the possibility of creating new search tools.

As an example of this commonly acknowledged need, one can cite the P. Heslin's review of TLG, CD-ROM E (see Bryn Mawr Classical review - BMCR 2001.09.23: <http://ccat.sas.upenn.edu/bmcr/2001/2001-09-23.html>):

"the policy of distributing data only, and leaving users to create or purchase their own software for accessing it, is precisely the correct one for a serious scholarly project. Contrast many other databases, such as the Bibliotheca Teubneriana Latina, which can only be accessed via vendor-supplied software. This means that it cannot be peerreviewed; that you are restricted to using a certain type of computer and operating system to access the database; that you can only ask it the type of questions that have been envisioned by the designers and allowed by their software; and that you cannot easily and independently verify the correctness of the responses given by the software. In a recent BMCR review, Anne Mahoney made the point more generally: "Electronic texts are most valuable when they can be used by other scholars, not just their creators, perhaps for purposes the creators did not anticipate ..." (BMCR 01.07.10 (10 Jul 2001), Hockey, Electronic Texts in the Humanities, reviewed by Anne Mahoney). It is the

open architecture of the TLG that makes it possible for scholars to use it creatively, and this is one of its most important features.”

The most innovative and most recent teaching and research practices are based on the availability of on-line resources and on textual databases. A database which will include both primary sources (the texts of the Latin authors) and secondary literature (critical discussions on authors and texts) will be of prime importance in the field. This has several parallels in other fields (see e.g. <http://www.perseus.tufts.edu/hopper/>, a website mainly designed for teaching classical languages).

POTENTIAL IMPACT:

The project is potentially capable of influencing research and teaching in a very significant way at the international level, both for the importance of the texts that will be offered to the scholarly community and for the ease of use provided by the technology we are planning to use.

A very large worldwide community of scholars makes everyday use of Latin digitized texts. This is clear for instance from the very large debate on Latin text that can be found in the web-based discussion group 'Humanist'

(<http://www.digitalhumanities.org/humanist/>). The lively debate on the quality, the characteristics, and the coding standards of digitized texts witnesses the importance of the use of digital texts in everyday research and teaching. Conferences on these topics take place every year (Digital Humanities) or every two years (JADT); journals such as 'Literary and Linguistic Computing' demonstrate the centrality of this aspect of research on literary texts.

The availability of digital texts will open up a number of possibilities for research on late antiquity. A thorough linguistic analysis of late-antique Latin texts is at present lacking, and is made very difficult by the lack of a reliable textual database. Such an analysis is especially important because of the great changes the Latin language underwent in that period. Editors of late-antique Latin texts have often chosen to normalize the language, making late-antique Latin similar to the classical language. While it is true that several late-antique authors aimed at reproducing classical style and language, many did not conform to it. The existence of a database will make it easier for future editors to recognize linguistic change and to find linguistic parallels for non-classical language. The literary study of late-antique authors will also be affected: scholars will find it much easier to trace literary influences and intertextual relationships.

The availability of digital texts will also be important for higher-level undergraduate students (*Laurea specialistica*) and for PhD students. They will be able to prepare dissertations and theses making use of the scholarly texts that will be offered on line; they will also use the textual database in order to obtain a better knowledge of the Latin language, and of the peculiarities of the Latin language of late antiquity. They will also be able to use the XML texts, modifying them according to the results of their research, if needed. They can prepare new editions of the texts, study the style and language of the authors, and prepare indexes and concordances of the texts. Digital editions will be also very useful in the case of texts of disputed authorship: statistical linguistic analysis can be applied to make progress in the dispute.

The diffusion of texts and of the critical studies on them is the best form of promotion. It is also a form of conservation: worldwide diffusion of digital texts among scholars will preserve their survival.

Finally, the operative process, as defined in this digitization project, is potentially important

for other types of texts and other literary traditions. The method used for scanning and for preparing optical character recognition (OCR) of scanned texts has the potential of being fruitfully used for other texts. Our research team will perform the task of optical character recognition (OCR) twice, using two different techniques (starting from a single scanned image of each text). The two versions will be checked against each other. The machinery and the software can be used again in the future for other digitization projects, not just by our research team but also by other researchers.

DESCRIPTION OF THE RESOURCES:

Technical resources:

The team will make use of two scanners uniquely dedicated to speedy and accurate digitization of printed books. State-of-the-art scanners offer truly excellent performances and do not damage the physical integrity of printed books. The scanned images will be analysed using two different types of optical character recognition (OCR) software. The texts will be thus corrected using a method similar to the one called 'double-keying'. A specialized external firm will be contacted to work as a consultant on the XML and PHI encoding. The choice will fall on a very reliable external firm.

Library resources

We are aiming at digitizing philologically accurate editions of the texts. The team will start using the libraries of the two departments involved in the project. In addition, some texts will be digitized using books obtained through interlibrary loan. Finally, in several cases it will be necessary to acquire printed editions, both from publishing houses, in the case of books which are still in print, or on the antiquarian market, in the case of out-of-print publications.

Human resources

The project team will include, in addition to the proponents, three additional full-time people to be hired for three years in order to work in Vercelli (two people) and in Turin (one person); each of these additional people will work for three years, under fixed-term contracts. These additional people will work on the organization of book acquisition (interlibrary loans; purchase of editions, etc.), on the scanning process, on the optical character recognition (OCR), and on the XML and PHI encoding. Two further full-time people will work for three years in Vercelli on the correction of the digitized texts.

ORGANIZATION:

The proponents form a strictly connected team. Each of them is an expert in an area that is essential for the success of the project. The curricula attached demonstrate that the team includes scholars who are expert in the field of Latin literature and history (Tabacco, Lomanto, Malaspina, Piastri, Vanotti, Balbo), a necessary prerequisite for the project. Other team members are expert in Greek literature and history (Battezzato, Vanotti): expertise in this area is essential component for the project, since several late-antique Latin texts include words or passages in Greek, and most Latin authors are steeped in Greek literature and culture. The team also includes people who have a scholarly interest and have taken part in the creation of large textual databases in the field of Latin (Lomanto), and others who are expert in the field of Information Technology for Humanities (Lana) and in the fields of library studies, library management, and management of digital resources for scholars (Botto).

The proponents have been selected on the sole criterion of the competences that are necessary for a speedy and effective achievement of the project. Two institutions are involved: the Università del Piemonte Orientale at Vercelli (University of Eastern Piedmont,

Vercelli) and the Università di Torino (University of Turin). The research will take place in two departments: the Dipartimento di Studi Umanistici (Department of Humanities), in Vercelli, and the Dipartimento di Filologia, Linguistica e Tradizione Classica (Department of Philology, Linguistics, and Classical Tradition) in Turin.

The main operational centre is Vercelli; the research unit in Turin collaborates to the project contributing with scholars who specialize in several fields essential to the project, and will provide access to its large library collection. The role of both research units (Vercelli and Turin) will be stressed and promoted on the occasion of public announcements about the start, development and completion of the research, along with that of Regione Piemonte as main sponsor. The research units will monitor the development of the project and, with the collaboration of administrative offices at both institutions, will check every four months to maintain a balanced budget. The research units will prepare a reasonable timetable for the accomplishment of the goals of the project. Web publication of the results will offer an easy and public way to monitor the progress in the accomplishment of the objectives. The rigorous monitoring process is linked to the way the project budget has been formulated. The proponents have proceeded to an accurate evaluation of the income of each single participant (taking into account the career stage and the yearly increases in the salary of each participant). In this way, we have been able to determine with precision the financial amount of the contribution we are going to make to the project, and the corresponding amount of money we are asking from the Regione Piemonte. The online form does not allow us to enter the details about this large amount of data on each single participant. However, we keep records of all the documentation that we used to for our calculations, and, in case the external referees or the monitoring offices of Regione Piemonte wanted to examine them, we are ready to produce it.

Oracle Corp. – through its local branch in Vercelli – was asked to contribute to the funding of the project by providing the software necessary to the website development. Oracle Corp. main center in the US is now examining the request. The local responsible in Vercelli is Mr. Bosio of the Internal Sales branch.

DISSEMINATION:

The research team shares the opinion that a research project financed by public funds should allow open access to all results. This complies with the protocols signed in Berlin and Messina. Both the University of Eastern Piedmont (Universita' del Piemonte Orientale) and the University of Turin (Universita' di Torino) have subscribed to the Messina protocol, which stresses that the subscribing institutions should provide open-access resources, and scholars should have free access to research data. Our project addresses this ethical issue: offering free access to the results of the research will improve the diffusion of knowledge, will allow greater freedom for scholarly activities and will offer easier access to information.

DIFFUSION OF THE RESULTS:

The research team will display the results of its work on a rich, accessible website which will offer:

- digitized Latin texts;
- bio-bibliographical notes about authors and works;
- digitized historical, critical, literary studies (not covered by DRM) about the late-antique latin authors and works;
- full-text search (also with an interface specifically designed for mobile devices);
- download of any text, selecting it by content criteria (word search) or by filing criteria

(genre, period, and so on), or others.

The project's aim will be presented, along with its progress, at National and International conferences on Classics, Computing Humanities, and so on.

We are planning to organise two public events during the phase of the research funded by Regione Piemonte: a workshop will take place in the mid-project, and a conference at the end. These events will describe the attained results and the possibilities opened by the availability of digital late-antique Latin texts.

The workshop will illustrate the project and explain its intermediate results. It will offer the research team the opportunity to check their methods and to receive suggestions from Italian and international researchers in order to improve the last phase of the work.

The final conference, open to Italian and international scholars, will give opportunity to show and explain the results achieved by the project and to discuss them with other scholars, also with the presentation of original contributions based on the research methods developed in the project. The public event will stress both the improvement of knowledge of late-antique Latin authors guaranteed by the results of this project and the importance of the help received from the sponsor(s).

A volume of proceedings will be published containing the results of the research.

The research data will be spread also by means of other websites such as Tulliana (<http://www.tulliana.eu>) developed by SIAC (Société Internationale des Amis de Cicéron) and BGL - Bibliographie des Grammairiens Latins (<http://htl.linguist.univ-parisdiderot.fr/site%20bgl/bgl.htm>).

ISSUES ETHICS:

Offering a digital version of printed books involves some ethical questions. The Latin and Greek texts of classical authors can be freely quoted, copied and digitized. It is possible to use the Greek or Latin text as prepared by recent editions; it is however not allowed to quote, copy or digitise the critical apparatus of editions which are not out of copyright. The critical apparatus is property of the philologist who prepared the edition or of the publisher of the book; in consequence of these limits, the digitization phase will concern only the texts, not the critical apparatus.

[M.Lana. Traduzione L. Battezzato]