

Digitizing Croatian Latin Writers

Neven Jovanović

Department of Classical Philology
Faculty of Humanities and Social Sciences
University of Zagreb
Lučićeva 3, HR-10000 Zagreb, Hrvatska / Croatia
neven.jovanovic@ffzg.hr

Abstract

Croatian Latin literature is usually taken to mean the body of works written in Latin both in the region that is today Croatia, and by authors of Croatian origin. The corpus of such literature comprises — according to a recent bibliographical survey (Katalog, 2008) — 3685 authors and 5994 (printed) titles, written from Middle Ages until modern times. The *Croatiae auctores Latini*, a digital collection in the making, aims to make accessible as many of these authors and works as possible (Jovanović, 2006). But how should it be done?

A collection wants to be used; therefore, it has to attract its users, enabling them to do something not possible without the collection. Moreover, if a collection which wants to be used offers material that is exotic, almost unknown, the collection also has to provide as many background information as possible. Finally, if a collection is built with limited resources, it has to rely on its main asset: the community of its users, along the lines of commons-based peer production (Wikipedia, 2008).

So the *Croatiae auctores Latini* digital collection should include not only texts in its three “aggregate states” of digitization (as images, as text only, and encoded in different levels). The collection should also include search facilities and secondary literature. To engage its user community, the collection should be open — making it easy to correct, enhance, or add texts and other resources, giving due credit to each contributor — and flexible enough to be used even for the purposes the designers have

not envisaged.

1 Introduction

The body of literature written in Latin by Croatian authors, from about 9th to 20th century, is today little known abroad, and studied by small number of people in Croatia. Here I will present, first, an *in nuce* survey of this literature and the state of research on it (2). Then I will show how scholars active in the field would profit from having a digital collection of Croatian Latin Writers, and why: how this collection can meet their current needs (3). The field is so small that — at least in Croatia — we know well both the people involved and their *modus operandi*. So, in building a collection for such a specialized research area, the Digitizing Croatian Latin Writers Research Project¹ has another agenda: we intend to build — or, more modestly, strengthen — a scholarly community around this collection (4). This community would, ideally, bring together scholars from various disciplines and various places, enabling exchanges which would otherwise not happen (5). To illustrate, I will give a short overview of what we intend to offer to the users (6), and how we plan to experiment with the approach used by the open source software developing communities — the so-called commons-based peer production — encouraging users to be-

¹The Digitization of Croatian Latin Writers Research Project at the Faculty of Humanities and Social Sciences, University of Zagreb, and funded by Croatian Ministry of Science, Education and Sports, started in 2007; its main researcher is Neven Jovanović. The project is part of larger research programme *Mediævalia et Neolatina Croatica*, also at the University of Zagreb, led by Darko Novaković (Jovanović, 2006).

come collaborators, to add further content to the collection (7). This approach differs somewhat from scholars', especially philologists', usual standards of publication; therefore, caution is necessary (8). To conclude, I will sketch what we currently have and what we do not have, pointing also to the initiatives and institutions — Croatian and abroad — which we see as natural allies of the *Croatiae auctores Latini* collection (9, 10).

2 Croatian Latin Literature

According to a recent bibliographical survey (Katalog, 2008), before the year 1860, in the region that is today Croatia and some parts of neighboring countries, at least 3685 authors have written and published at least 5994 books in Latin. I insist on numbers because we have got them only recently, having migrated bibliographic and catalogue records into a computer database. But I insist on numbers also because they are large: the Croatian literature written and published in *Croatian* comprises, until the year 1860, just half as much titles (exact ratio is 5994:3000 in favor of Latin texts).²

In modern Croatia, literature in Latin is proudly perceived as a proof of the nation's long-lasting participation in European culture. But the works and authors of this Latin literature share the fate of all Neo-Latin literature: people generally do not know them, have difficulties understanding them (especially because of the vigorous dialogue with classical antiquity in many such works), and, in general, read the texts only in translations. There is a canon of Croatian Latin authors.³ In Croatia, perhaps the best known among them is Marko Marulić from Split (1450–1524), but we could mention also Janus Pannonius (1434–1472), Matija Vlačić Ilirik (1520–1575), Marko Antun de Dominis (1560–1624), and Ruđer Bošković (1711–1787). However, even the authors from the canon are *nomina tantum* for the general educated public. Admittedly, among these authors there is no Dante, Petrarch, or Erasmus. Also, many of authors which the Croatians consider

²The count of Latin works is not final; searches in Italian databases (EDIT 16, 2000) find authors and titles not included in a recent Croatian bibliographical list (Katalog, 2008).

³This canon, as established by an influential anthology from the 1970's (Gortan and Vratović, 1969–70), consists of 37 writers from the Middle Ages until the year 1848.

their own belonged to more than one culture, so some of them are claimed also by Italians, others by Hungarians, others again by Serbs.

3 Desiderata: Scholars' Needs

After some 140 years of research on Croatian Latin literature,⁴ we still do not have modern editions of all texts which merit publication. We do not have two basic reference works: a dictionary and a secondary bibliography of Croatian humanist Latin. Also, we do not have a national catalogue of manuscripts (even those written in Croatian language), so we have to rely on international bibliographies such as Kristeller's (Kristeller, 1963–1997), and on personal knowledge of individual libraries and collections.

These desiderata cause a noticeable disproportion in research: there are many monographic studies, but little work on genres and periods, and virtually no explorations of styles in different periods.

The access to materials needed for such studies is difficult. Some are rare, precious, and fragile (manuscripts, early printed books). Other are simply too dispersed. Once you know what title you are looking for, it can turn out to be over a hundred years old, published in an exotic scholarly journal, in a country that does not exist any more (the Habsburg Monarchy, Yugoslavia), or by some private or local initiative, for whom keeping a title in print, or distributing it, is not a task of prime importance.⁵ The same goes for both the texts of Croatian Latin authors and for secondary literature on them.

4 A Small Community

A small and little known field of research has its merits. Apart from ample possibility for original work and dramatic discoveries,⁶ little time is needed

⁴First 130 years of this research are surveyed in a publication translated into Italian, and accessible online (Novaković, 1997 b).

⁵An important project of Croatian neo-Latin philology, publication of the *Opera omnia* of Marko Marulić, began in 1984, and comprising 15 volumes so far, is being carried out not by a scholarly or university press, but by the Književni krug Split, which is, legally, a citizens' association; however commendable the initiative, the fact remains that some of the early volumes are long out of print.

⁶A good example is provided by the research on, again, Marko Marulić: in the last sixteen years — from 1992 — there

to get to know the scholars involved, their special interests, the way they do research, and the degree of their openness to change.

Digitizing Croatian Latin writers is a meaningful enterprise if it does not remain a goal in itself, but if it results in a resource that will be used. Thinking this way, we came to realize that planning should begin from the actual scholars' needs, somehow offering the actual scholars what they do not have or cannot easily get. Only afterwards — only after the existing community has, so to speak, grown into the collection — should we start to explore *new* ways to do research in the field. This line of reasoning is similar to the one that led John Bradley to develop his *Pliny* software for annotation and notetaking: “it tries to model some aspects of established scholarly practice... [supports by the computer] the act of developing an interpretation... [builds a tool for] scholarship [that] draws on many different resources” (Bradley, 2007).

Furthermore, a collection built for an existing community naturally supports another aspect of scholarship: research (and teaching of it) as a social activity. If carried out successfully, the digital collection of Croatian Latin writers could become a focal point, a center of the community researching these writers. This community is at the moment too dispersed, too atomized, to really function together.

5 The User Profile

So, who are those potential users of the *Croatiae auctores Latini* collection? Several generalizations can be made.

There are scholars, and there are those who are not (yet) scholars: university and highschool students, highschool teachers. Those scholars and learners belong to different areas of study: to philology, linguistics (not many of them), history, history

were findings not only of new manuscripts and autographs, but of completely unknown works by Marulić, such as the 141 Latin poems in a manuscript in the University Library, Glasgow (Novaković, 1997). Startling connections were made and proved, such as the fact that Marulić temporarily owned, annotated, and wrote in the manuscript codex that we today know as *codex Parisiensis lat. 7989 olim Traguriensis*, our main source for Petronius' *Cena Trimalchionis*: on empty pages of this codex the author from Split copied Claudian's *De Phoenice*; and Catullus' text in the same codex bears ample annotations by Marulić (Lučin, 2005; Lučin, 2007).

of literature, history of art (quite an active research group), theology. They differ in language skills: some are at home with Latin, others not; many of them read and write Croatian, but we hope to attract scholars from abroad too. Finally, all users can be supposed to have some rudimentary computing skills (writing texts, using the internet, searching the catalogues) — but there certainly are not many “power users”; e. g. we cannot expect our users to know what regular expressions are.

Because of all these differences, we have to aim at designing a collection that would be general enough, and open enough; a collection that would actually be stronger in infrastructure (tools, standards, methods, interfaces, documentation) than in texts themselves; a collection whose main goal would at one point become receiving content and metadata *from its users*, and stimulating their collaboration.

6 What Can We Offer?

Ideally, the *Croatiae auctores Latini* digital collection would offer to its users the following:

- Content
 - digitized Latin texts in different formats (as plain-text files, as images of scanned pages with or without uncorrected OCR text, as XML encoded editions), in different phases of editing process
 - digitized translations of these texts, also in different formats and phases of editing process
 - metadata to help users orientate, both in the collection and within a particular text
 - biographic and bibliographic information on authors and works (including the resources available over the internet)
 - secondary literature on authors and works
- Tools
 - clear, attractive user interface
 - flexible utilities for searching and browsing
 - utilities for researching allusion and intertextuality (for identifying quotations and citations, by querying important source texts such as the Vulgate)

- utilities for creating one’s own collections from the available material, for describing and sharing it
 - utilities for adding markup (and tags, and bookmarks, and annotations) to existing content
 - utilities for adding new content (both primary texts and secondary literature)
- Good integration of data and utilities

The texts included in such a collection would be not only literary — in the broadest sense of the word, as used in classical antiquity — but also documentary (charters, statutes, inscriptions); not only primary sources, but also secondary literature; finally, there would be not only “our” texts — texts of Croatian Latin writers — but also texts with which our texts are entering into a dialogue — by quotation, citation, allusion, intertextuality.

7 The Road to Imperfection

Two further points should be noted regarding the lists in section 6. In building the *Croatiae auctores Latini* digital collection we want to embrace the “release early, release often” principle, as proposed by the Open Source software initiatives (Raymond, 2002): offer something as soon as possible, however imperfect it is (e. g. “photo albums” of pages from a book, an unreliable OCR from a scanned page, a collection with little integration), but, at the same time, be ready for constant updating, for supporting the users and keeping in lively contact with them.

We choose this strategy because we believe that something is always much better than nothing; at the same time, we think that the release of an imperfect resource can be seen and presented as an *invitation* to our colleagues to become not only users, but also collaborators; if you do not like what you see, if you think something is missing, help us improve what we have!⁷

⁷Such an approach would follow partly in footsteps of the Perseus Project and the impressive CAMENA / MATEO neo-Latin text collection (Kühlmann, 1999) — editors of the latter especially stress that they present something imperfect, and they rely heavily on page images of books, adding XML markup to transcriptions as they go. But we want to give even more room to the approach practiced by Wikipedia and similar Web 2.0 projects (corrected by the high standards of a scholarly community).

The lists in section 6 show also that we see the content and the tools for using it as equally important, as deserving of equal care. This is quite a challenge, as it requires juggling with additional balls in the air. But recent research on why digital resources become neglected by the user community (Warwick et al., 2008) confirms the importance of components like the interface. Though creators of digital resources sometimes neglect such components, choosing to concentrate on contents, what is at stake here is more than just “look and feel” of a resource. The interface is main helper or main obstacle to users; it determines how the collected material will be accessed and manipulated. An example: for someone working on a synthetic study of Croatian Latin epic poetry it is obviously vital to have access to texts in reliable editions, but this someone will also profit greatly from creating (or finding) a tailor-made corpus of epic poems inside the general collection, and — furthermore — from being able to return to that corpus again and again.⁸

8 Excursus: a Caveat

Fate of a digital resource in the humanities depends not only on ease of use, but also on the quality of content. If, when building a collection, we stress speed in releasing the texts, if we open the gate of collaboration to anybody interested — do we not risk loss of scholarly credibility?

Editions of texts which offer scanned page images, unreliable OCR, proofread plain text files, or even a combination of all three, certainly cannot meet the standards for an electronic critical edition as have been proposed e. g. by the “Textual Scholarship Research Guide” (CTS, 2007). What we intend to provide — at least in early phases of the collection — is some sort of a “practical edition”, as described in the same research guide:

1. base the text on a historically important source

⁸This could be done with a tool similar to “Filtro autori/opere” in the Poeti d’Italia in lingua latina database, or to the “Bibliographic Search” in the Perseus under PhiloLogic collection (note that the latter has a “Genre” field, while Poeti d’Italia rely on users’ knowledge); another excellent example is the search interface of the digital Biblioteca Italiana (Biblioteca, 2007). Still, none of these resources offers an option of *saving* a particular private corpus, which can be done, for example, in Peter Heslin’s *Diogenes* (Heslin, 1999–2007).

text

2. accurately identify the text source
3. proofread the text, describe proofreading methods employed
4. add one or more introductions or associated essays and annotations describing the composition and production of the work, the critical reception (in its own time and subsequently), the historical backgrounds that make the work more accessible to readers from more modern times or other cultures
5. give indication of where additional scholarship can be found

Choosing wisely, and providing ample information about the quality, provenance, and comprehensiveness of materials presented, we would certainly meet the expectations our users have from their contact with traditional scholarship. Even more important, this information would create additional metadata, and thus enrich the collection.

9 The Allies

If the *Croatiae auctores Latini* collection is to function on principles of commons-based peer production — as a “coordinated, (chiefly) internet-based effort whereby volunteers contribute project components, and there exists some process to combine them to produce a unified intellectual work” (Wikipedia, 2008) — the collection must be well visible. People have to be aware of the collection, have to be able to find it. This means that it has to be “advertised” in places where potential users and collaborators usually seek information, and in places one would turn to if one wanted to find digitized Croatian cultural heritage. This is where our natural allies are; these are primarily libraries (university and national, online and offline) and research institutes, but also other on-line digital collections and digitization projects of similar nature (i. e. connected with Latin language, with Croatia, or with Europe from 9th to 20th century), both national and international.

Following this idea, contacts have been made with Croatian National and University Library in Za-

greb, and with the new Croatian Cultural Heritage project.⁹

10 What We Have, What We Need

At the moment, the Digitization of Croatian Latin Writers project has amassed sufficient material to offer a sample of its intended variety. This means that we have available both carefully edited texts and texts as simple photo-albums or PDF images of scanned pages; there is both prose and poetry, short and long texts, works of literature and historical documents. When it goes online, the *Croatiae auctores Latini* collection will comprise also an in-depth presentation of one important author (Marko Marulić) and one representative literary genre (epic poetry). Some bibliographic data have also been compiled, both for older literature and for texts already available on the internet. A backbone of our collection will be a bio-bibliographic database of some 230 Croatian Latin writers, culled from a recent lexicon of Croatian writers (Leksikon, 2000); to this backbone we will be gradually adding texts and secondary literature.

There is also a “scriptorium”, a group of interested students, both undergraduates and graduates, competent in Latin, ready to transcribe, proofread, and — with some training — also encode texts for the collection.

We have easy access to practically all recent editions of Croatian Latin writers prepared in Croatia, many of them obtainable in some electronic format. Indeed, one of aims and incentives of the Digitization of Croatian Latin Writers project is to serve as an additional scholarly publishing channel, an electronic printing press both for other components of the *Mediaevalia* and *Neolatina Croatica* research programme, and for other scholars editing Croatian Latin texts.

Also, according to Croatian copyright law, all works published in Croatia before the year 1939 are now in the public domain, freely available for digitization.

What we do not have?

⁹For information about the Croatian National and University Library in Zagreb, see <http://www.nsk.hr>; the Croatian Cultural Heritage website is at the address <http://www.kultura.hr>.

We do not know yet what software tools will we use for doing what we want to do. The resources — both human and financial — for developing the Croatiae auctores collection Latini are severely limited, and probably always will be. Therefore, we are strongly inclined towards tools already created and used by others (and made available as open source software). There are several open source tools for building a digital collection that look promising; in the next year we will be testing them with our material.

These tools are:

- PhiloLogic (developed by the ARTFL Project at the University of Chicago in collaboration with The University of Chicago Library), for searching large encoded databases on the World Wide Web (PhiloLogic, 2008)¹⁰
- the eXtensible Text Framework (XTF) application for search, browse and display of heterogeneous digital content, by the California Digital Library (XTF, 2008)¹¹
- Omeka web platform for publishing and annotating collections and exhibitions, by the Center for History and New Media and the Minnesota Historical Society (Omeka, 2007)
- Collex, a set of tools designed to aid students and scholars working in networked archives and federated repositories of humanities materials, by the Applied Research in Patacriticism project (Collex, 2008)

We have a scriptorium, but we need also a “programmatorium”, a group of people interested in humanities research and well versed in programming and computing.

¹⁰Used in Perseus under PhiloLogic collection at the University of Chicago: <http://www.lib.uchicago.edu/efts/PERSEUS/>, and in several other interesting resources; cf. <http://philologic.uchicago.edu/samples.php>, and especially the Montaigne project, <http://www.lib.uchicago.edu/efts/ARTFL/projects/montaigne/>.

¹¹The XTF system, together with the Exist XML database, provides the infrastructure for the Biblioteca Italiana digital collection (Biblioteca, 2007), which is — in my opinion — a representative example of static digital collection (“static” only because it provides no means for users’ contribution).

The resources need to be integrated in a logical and easy-to-follow way; connections have to be made not only inside our own collection, but also from the collection towards the outside world (as the collection wants to be open to the users, it wants to be open to other collections too). It would be a great achievement to find a way not only to point from the Croatiae auctores Latini to complementary resources on the web, but also to query them from our site (for example, a digital Vulgate edition we need for identifying Bible quotations does not necessarily have to sit *among* the Croatiae auctores Latini — we would be happy to send queries to somebody else’s service).

Furthermore, caution is called for when speaking about notes, collections and tags: experience shows that tasks which seem trivial in the real world, with tools such as pen and paper (adding one’s own annotations, deciphering citations), turn out to be quite complicated — and complicated in unexpected ways — in the world of computing.

Finally, I think we would have to work hard to persuade our colleagues to *actively* contribute to the collection: a good deal of encouragement, support, and especially demonstrations of usefulness — showing why it is better to use the collection than not to use it — would be necessary. But it is worth the effort: a working digital collection with an active user base (as international as possible, at that) could bring about consolidation of research in Croatian Latin literature, something that this field has not yet achieved.

References

- Applied Research in Patacriticism. 2008. *Collex, the social software and knowledge discovery tool*. <http://patacriticism.org/collex/> [accessed April 28, 2008].
- ARTFL Project and the Digital Library Development Center (DLDC) at the University of Chicago. 2008. *PhiloLogic full-text search, retrieval and analysis tool: 3.1*. <http://philologic.uchicago.edu/> [accessed April 28, 2008].
- Bradley, John. 2007. Pliny and traditional scholarly practice. *DRHA Dartington / Doing Digital: Using Digital Resources In The Arts And Humanities, 9 - 12 September 2007*. <http://www.dartington.ac.uk/drha07/>

- papers/abstract.asp?uid=35 [accessed April 28, 2008].
- Centro interuniversitario Biblioteca italiana telematica (CiBit), Università di Roma "La Sapienza". Copyright 2007. *Biblioteca Italiana (BibIt)*. <http://www.bibliotecaitaliana.it/> [accessed April 28, 2008].
- The California Digital Library (CDL). 2008. *The CDL eXtensible Text Framework (XTF) version 2.1*. <http://www.cdlib.org/inside/projects/xtf/> [accessed April 28, 2008].
- Center for History and New Media, George Mason University. 2007. *Omeka web platform: 0.9.1.1*. <http://omeka.org/> [accessed April 28, 2008].
- Centre for Textual Scholarship collaborators. Copyright 2007. *Textual Scholarship Research Guide*. <http://www.cts.dmu.ac.uk/index.php?q=researchguide.html> [accessed April 28, 2008].
- Istituto Centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche — Laboratorio per la bibliografia retrospettiva. 2000. *EDIT 16: Censimento nazionale delle edizioni italiane del XVI secolo*. <http://edit16.iccu.sbn.it> [accessed May 7, 2008].
- Gortan, Veljko. and Vratović, Vladimir. 1969–1970. *Hrvatski latinisti = Croatici auctores qui Latine scripserunt*. Pet stoljeća hrvatske književnosti, knj. 2–3. Matica hrvatska, Zora: Zagreb.
- Heslin, Peter. 1999–2007. *Diogenes, a tool for searching and browsing the databases of ancient texts: Version 3.1*. <http://www.dur.ac.uk/p.j.heslin/Software/Diogenes/> [accessed April 28, 2008].
- Jovanović, Neven. 2006. *Digitization of Croatian Latin writers: a research project proposal*. <http://www.ffzg.hr/klafil/dokuwiki/fetch.php?cache=cache&media=z%3A060321digi.html> [accessed April 28, 2008].
- Kristeller, Paul Oskar. 1963–1997. *Iter Italicum: a finding list of uncatalogued or incompletely catalogued humanistic manuscripts of the Renaissance in Italian and other libraries*. Warburg Institute, London.
- Kühlmann, Wilhelm. 1999. *DFG-Projekt CAMENA, Heidelberg-Mannheim*. <http://www.uni-mannheim.de/mateo/camenahtdocs/camena.html> [accessed April 28, 2008].
- Lučin, Bratislav. 2007. Marul, Katul i Trogirski kodeks Petronija (Codex Parisiensis Lat. 7989 olim Traguriensis) / Marulus, Catullus and the Codex Traguriensis (Paris, Bibliothèque nationale, Lat. 7989). *Colloquia Maruliana*, 16:5–44. http://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=18260 [accessed April 28, 2008].
- Lučin, Bratislav. 2005. Marulićeva ruka na trogirskom kodeksu Petronija (Codex Parisiensis lat. 7989 olim Traguriensis) / Marulić's Hand on the Codex Traguriensis (Paris, Bibliothèque Nationale, lat. Parisiensis lat. 7989). *Colloquia Maruliana*, 14:315–320. http://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=13648 [accessed April 28, 2008].
- National and University Library Zagreb contributors. s. a. *Katalog 045: Hrvatski latinisti (= Catalogue 45: Croatian Latin Writers)*. <http://161.53.3.18/liste/045/> [accessed April 28, 2008].
- Nemec, Krešimir, Dunja Fališevac, and Darko Novaković (eds.). 2000. *Leksikon hrvatskih pisaca*. Školska knjiga, Zagreb.
- Novaković, Darko. 1997. *La filologia neolatina in Croazia (breve bilancio degli ultimi 130 anni)*. Hrvatsko društvo klasičnih filologa, Zagreb. <http://www.ffzg.hr/klafil/dokuwiki/fetch.php?cache=cache&media=z%3A1997-novakovic-filologia-neolatina.pdf> [accessed April 28, 2008].
- Novaković, Darko. 1997. Dva nepoznata Marulićeva rukopisa u Velikoj Britaniji: MS. ADD. A. 25 u oxfordskoj Bodleiani i Hunter 334 u Sveučilišnoj knjižnici u Glasgowu *Colloquia Maruliana* / Two Recently Discovered Manuscripts of Marko Marulić in Great Britain: MS. ADD. A. 25 in the Bodleian Library, Oxford and Hunter 334 in the University Library, Glasgow. *Colloquia Maruliana*, 6:5–31. <http://www.ffzg.hr/klafil/dokuwiki/fetch.php?cache=cache&media=z%3A060321digi.html> [accessed April 28, 2008].
- Raymond, Eric Steven. 2002 (copyright 2000). *The Cathedral and the Bazaar*. <http://catb.org/~esr/writings/cathedral-bazaar/cathedral-bazaar/> [accessed April 28, 2008].
- Warwick, Claire and Melissa Terras and Paul Huntington and Nikoleta Pappa. 2008. If You Build It Will They Come? The LAIRAH Study: Quantifying the Use of Online Resources in the Arts and Humanities through Statistical Analysis of User Log Data. *Lit Linguist Computing*, 23(1): 85-102.
- Wikipedia contributors. 2008. Commons-based peer production. *Wikipedia, The Free Encyclopedia*. http://en.wikipedia.org/w/index.php?title=Commons-based_peer_production&oldid=193586738 [accessed April 28, 2008].