

CroALa index locorum: a gazetteer of place names in Croatian Latin texts

Neven Jovanović, Department of Classical Philology, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia (neven.jovanovic@ffzg.hr)

Summary

A team of tenured professors, postdocs and PhD students offers to create, over the next four months, a gazetteer of 10,000 occurrences of place names in 5.7 million words of Croatian Latin texts, written mainly during the Early Modern period and published as a digital collection *Croatiae auctores Latini*. The gazetteer, which we consider an *index locorum*, will use stable URIs to refer to locations in the collection and to specify which place in which period is being referred to, and which lexical form of a Latin word contains the reference. Published under a CC-BY license, the gazetteer will be made available as a public repository, as a set of web pages on our own server, and as a dump file ready to be included into Pelagios. Long-term preservation of the dataset will be ensured by the University of Zagreb.

Documentation for both the dataset and the principles of its creation will be part of the repository. The *Index locorum* will offer nuanced and peer-reviewed interpretation of literary references to places (for example: personified places, places at specific points in time) which will make Pelagios an initiative more open to philologists, literary scholars, and historians of ideas and culture. The index may serve as inspiration for functionalities in the next version of Recogito as well.

CroALa – a growing collection of Neo-Latin texts connected with Croatia

Croatian Neo-Latin is one of the main research areas at the Department of Classical Philology, Faculty of Humanities and Social Sciences, University of Zagreb. It engages five of its twelve tenured/tenure-track faculty members, as well as colleagues from the Juraj Dobrila University of Pula and the Marulianum – Split Literary Circle. For this research, the *Croatiae auctores Latini* collection (CroALa, croala.ffzg.unizg.hr; editor-in-chief is Neven Jovanović) serves as both an exploratory tool and a publishing outlet.

CroALa is a collection of Latin texts written by Croatian authors between 976 and 1984; its first version was published in 2009. Infrastructurally, it is supported by the Faculty of Humanities and Social Sciences, University of Zagreb.

The collection was developed during 2006-2014, in the framework of the project

“Digitisation of Croatian Latin Writers”, funded by the Croatian Ministry of Science. A bibliographic component ([CroALaBib](#)), containing TEI XML records of primary and secondary works, was added to the collection in 2010 thanks to a Google European Digital Humanities Award. In 2013-2015, the project [Croatica et Tyrolensia](#), funded by the Croatian Unity through Knowledge Fund (and a collaboration of our Faculty with the Ludwig Boltzmann Institute for Neo-Latin Studies, University of Innsbruck), made possible further work on deploying and analysing the collection as an XML database (testing deployment is at [solr.ffzg.hr/baseX/](#); a list of queries is [part of the Bitbucket repository](#)). Thanks to the various smaller grants and personal enthusiasm, further texts are being added to CroALa all the time, slowly but steadily, usually transcribed and TEI-encoded by graduate and undergraduate students of Latin under the supervision of the professors of the Department. A new batch of editions will be added to CroALa thanks to the project TeMrežaH (Textual Networks of Early Modern Croatia, 2014-2018), funded by the Croatian Science Foundation.

Currently, CroALa comprises some 5.7 million words. The collection exists in three forms: as a full-text searchable PhiloLogic database (unfortunately, the PhiloLogic 3 version, our main deployment, is becoming hard to maintain because of its outdated dependencies, while the PhiloLogic 4 is still in Beta – there is a [testing deployment of CroALa in this version](#) too), as a cluster of XML databases to be queried through BaseX with tailor-made XQuery scripts (with better and easier control over the data, though without a public user-friendly and universal interface), and as a repository on [Github](#) and [Bitbucket](#) (the bibliographic component has [its own repository](#)). The editions are encoded in TEI XML, and published under an Open License (CC-BY); it is very important to us that the texts are reused in as many different ways as possible.

At the moment we are considering ways to present CroALa as an interesting resource both to specialist colleagues (national and foreign Neo-Latinists and Early Modern historians), and to researchers from other disciplines (linguists, digital humanists). It may seem that the Latin language is a barrier in using such a resource – but there are still more people learning and reading Latin than there are academic readers of Croatian. Our experiences suggest that a more important obstacle is the fact that most of the authors and works in CroALa are little known. This is why we want to provide a set of familiar entry points from which the prospective users might start exploring, reusing, or remixing our texts.

Adding an Index locorum to CroALa

One familiar entry point to approach the unknown texts in CroALa would be a detailed index of place names (*index locorum*). As an addition to a digital collection, this index would be designed in such a way to be exportable – that is, to be included in other digital resources, such as Pelagios – and interoperable, linking scholarly data published in different contexts.

Careful planning of such an index (in another words, an opportunity to consider the index as a scholarly resource in itself, and not just a welcome addition to something else, as is usually the case with the indices in books) will make it possible to distinguish – and thus, later, reflect upon – meanings which the place names have at various locations in various texts. For example, here are some occurrences of the nominative case of the Latin word “Roma” from CroALa:¹

(1) nam Roma tam diu mundi tenuit monarchiam, quam diu uiguit iusticia in Romanis

(2) Multum Roma suo debet reparata Camillo, / Sed plus Guarino lingua Latina suo.

(3) IN TRES CARDINALES CAMERARIOS / Exstincto vitulo, vivo bove cornua praesunt, / Roma, tibi: en facta es altera Pasiphaë.

(4) Tibur oppidum est ab urbe Roma ad XVI lapidem distans

(5, of st Jerome:) Sic enim ad uirgines deo dicatas de ipsarum institutione scribens ait: Me ante quam uicesimum etatis annum attingerem urbs Roma in summum elegerat magistrum in omnibus pene liberalibus artibus.

While it may be said that occurrences (1), (2), and (5) refer to “the capital of the Roman Republic and Empire”, as defined at the URI <http://pleiades.stoa.org/places/423025>, we may also assert with more nuance that occurrences (2) and (5) refer to Rome at specific periods of Antiquity (<http://n2t.net/ark:/99152/p0qhb66qj4c>) and Late Antiquity (<http://n2t.net/ark:/99152/p06v8w42jtx>). Moreover, (4) may be said to refer Rome from 750BC to the time of writing of the *In epigrammata priscorum commentarius* by Marko Marulić (ca. 1500), while (3), written by Janus Pannonius in the 1440s, refers to a very specific moment in the history of the Italian Renaissance (<http://n2t.net/ark:/99152/p0qhb66pk6z>).

1 Cf. also occurrences of “Roma” in our [PhiloLogic 3](#), [PhiloLogic 4](#), and [BaseX](#) deployments.

All assertions stated above may be modeled semantically using existing RDF ontologies; we may also say *who* is making the assertion, *when*, and who reviewed the assertion. Once expressed as RDF, this set of assertions may be transformed into different machine-actionable formats, either to be published as a separate HTML page providing to readers links both to locations in CroALa and to definitions of periods and places being referred to, or as an output that can be ingested into other digital resources.

A further layer of complexity is introduced by the fact that “Roma” in occurrences (1), (2), (3) and (5) above is noticeably personified – it is at the same time a place and a person (it “holds the rule over the world”, it “owes much to Camillus”, it “becomes another Pasiphae”, it has chosen Jerome as the “summus magister”). Clearly, we are dealing with literary texts here. This layer of complexity could and should be expressed in our semantic modelling as well – especially because nothing stops us from saying (in RDF) that an occurrence refers to a place and a person *at the same time*. This will bring our project near not only to Pelagios, but also to SNAP:DRGN (which have the ambition to somehow include fictional persons into their prosopographical meta-collection).

Not all references to “Roma” will be in the nominative case. Again, building on existing ontologies, we can add the following simple assertion to our note (this is just a mock-up example):

```
<http://croala.ffzg.unizg.hr/verba/0003248.htm> a lemon:LexicalEntry ;
```

```
lemon:Form [ lemon:writtenRep "Roma"@la ] .
```

The assertion says “this address is a lexical entry (description of a word) which has the form “Roma” (in Latin), which is represented in writing as “Roma” (the list of forms and written representations can, of course, be extended).

Finally, because of the interpretative nature of our assertions, we think that it is important to state who made the assertion (and when), as well as who checked the assertion (and when). This guarantees scholarly quality and controllability of the resource

CroALa index locorum and Pelagios

We believe that such an index of places in digital editions of Croatian Latin texts would benefit not only CroALa, but Pelagios as well. First, Pelagios is currently strongly oriented at “real” historical places, but a lot of historical texts refer to

places whose reality is, at least, open to interpretation; providing a way to tackle such references (instead of either skipping them over or “flattening” them into “references to a historical place”) may widen the circle of Pelagios users beyond archaeologists and “hard-core” historians – though the index will provide more nuanced data to these groups too – to literary scholars and historians of ideas and culture in general.

A similar reasoning applies to the often encountered situation of treating a place name as a person. Again, we can choose to ignore the ambiguity (in which case we lose something), or we can work to express it (and then choose to disregard it in one analysis and take it into account in another).

Furthermore, we feel that an index such as the one proposed here cannot be produced with Recogito. Admittedly, our experiences with Recogito are very limited – but they are limited precisely because Recogito seems best adapted to places which can be “shown on the map”. Creation of the CroALa index locorum could provide a conceptual help for a Recogito with wider and more nuanced coverage of the ways places are referred to in historical texts – again attracting to Pelagios students of historical ideas, mentalities, and cultures in the widest possible sense. Finally, Recogito is a tool for “annotating place references in [...] documents”. As philologists, we feel that such a tool should be able to take into account that verbal references to places can take different linguistic forms, and that such capability could make the tool more useful, or more familiar, to our fellow philologists and literary scholars (active in initiatives such as the [Digital Roman Heritage](#), led by Susanna De Beer, or in the [Ludwig Boltzmann Institute for Neo-Latin Studies](#) at the University of Innsbruck, and in larger philological infrastructures such as the [Open Greek and Latin Project](#) at the University of Leipzig).

Pelagios development resource grant and CroALa index locorum

In the past decade, people connected to CroALa have gained a certain amount of experience with analysing, producing, and transforming many variants of XML-encoded texts and bibliographic records (via XSLT and XQuery), with the use of version control for teamwork, and with reuse of curated digital data from other sources. CroALa also has provided opportunities to build a team of experienced TEI encoders with excellent knowledge of Latin language and a thorough understanding of Croatian Latin literature. The current team consists of postdocs and graduate students, working under supervision of the tenured staff at the Department of Classical Philology. At the moment, however, the Department does

not have the means to either employ members of this team or to finance their internships.

Therefore, we propose to use the Pelagios development resource grant in the amount of 5,000 GBP to produce a pilot *Index locorum* for 10,000 occurrences of place names in CroALa texts as described above. The funds from the grant will be used to finance personnel costs: 500 hours of work of seven encoders/reviewers (a record created by one person will be reviewed by another) over four months, from the moment of acceptance of the proposal until 30th November 2016.

A record in the CroALa *Index locorum* will contain seven statements:

1. a stable URI referring to the occurrence of a place name in CroALa
2. a reference to a URI which identifies the place outside CroALa
3. a reference to a Latin lexical entry whose written representation occurs in the text
4. an assertion that a special historical period of the existence of the place is being referred to (or that this cannot be claimed)
5. an assertion that a place is mentioned unambiguously or ambiguously as a place (whether it is personified or not)
6. a note on who made the record and when
7. a note on who reviewed the record and when

The *Index locorum* will be published under an Open License, and in three formats: as a version-controlled repository (through Open Science Framework or Zenodo – both have Github integration capability), as a set of web pages with live links on the main CroALa server (croala.ffzg.unizg.hr), and, on the same server, as a gazetteer summarized and formatted according to the Pelagios Cookbook (“dump file”).

Additionally, both because we will be working as a team, and because experience shows that lack of documentation is often an issue with digital humanities projects, documenting all stages of the process of creating and reviewing the *Index locorum* is an essential part of this proposal.

Formats, structures, and principles of the index will be documented on the repository wiki; they will be formulated as a series of “user stories” (as has been done on the Pelagios project pages, for example), describing how other creators / annotators and researchers can use and reuse the index and its structures, to refer to it from their own text (for example, through Pelagios / Peripleo), to study the

CroALa collection, or to create their own indices using the CroALa *Index locorum* as a template.

The index will be documented in parallel with its development. Most of the documentation will be ready and published by the November 2016; corrections and additions will be performed as necessary.

Supported by the IT infrastructure of the University of Zagreb, both resources (the *Index locorum* and its documentation) will remain accessible for at least 10 years; long-term archiving will be done by [Dabar – Digital Academic Archives and Repositories](#) service of the University of Zagreb.

We will also use the Pelagios Commons to share information with the Pelagios users and keep them up to date with what we are doing.

Team members and responsibilities

The CroALa *Index locorum* team will be coordinated by Neven Jovanović (Associate Professor at the Department of Classical Philology, Faculty of Humanities and Social Sciences, University of Zagreb, editor-in-chief of the CroALa collection). Jovanović will also set up the necessary repositories and databases, provide templates and forms for the index and instructions for annotation, and provide automatic formatting and validation of the final results.

The index entries will be reviewed by Irena Bratičević, University Docent at the Department of Classical Philology, and by Petra Šoštarić, Assistant Professor at the same department.

Entries will be created (and also reviewed) by PhDs Nina Čengiđ and Luka Špoljarić (both currently unemployed, previously postdoc researchers at the Croatica et Tyrolensia project of the Department), and by the doctoral students Anamarija Źugić and Źeljka Salopek.

Čengiđ will also coordinate documentation writing and copy-edit the English.

The Pelagios funds will be used to pay for the work of Čengiđ, Špoljarić, Źugić, and Salopek.