

Željka Salopek
Faculty of Humanities and Social Sciences
University of Zagreb
Ivana Lucica 3, 10000 Zagreb, Croatia
e. zsalopek@ffzg.hr

COST Action IS1310 Reassembling the Republic of Letters, 1500–1800
Short term scientific mission scientific report

Title of the STSM: Croatian neo-Latin correspondence and the EMLO catalogue :
collaboration, consultation and training

Time and venue: 20th to 31st of March 2017. University of Oxford.

1. Purpose

The purpose of this STSM was to work on the standardization of metadata and the further development of a catalogue of correspondence, to learn new digital techniques and methods in developing and managing a digital catalogue, and to collaborate with and learn from the team at the Early Modern Letters Online (EMLO) database, a part of the project *Cultures of Knowledge - Networking the Republic of Letters, 1550-1750*, which is based at the University of Oxford. I was fortunate to be working with the digital editor of the EMLO catalogue, Ms. Miranda Lewis.

The aim was to contribute the collection of the Croatian neo-Latin correspondence into the EMLO union catalogue with the focus on the method of metadata standardization. The letter records in the catalogue would be connected to the Croatiae Auctores Latini (CroALa) database¹ via a link. The other part of the STSM was to try to create a guide or a set of instructions in Croatian for potential new contributors from Croatia and the region. The goal was to enable researchers to prepare their collections and the metadata for contributing to the union catalogue and thus facilitate scholarly collaboration as well.

2. Corpus of letters

The corpus of letters proposed for the STSM is a part of a bigger collection, Croatiae Auctores Latini (CroALa) developed at the University of Zagreb Faculty of Humanities and Social Sciences. CroALa comprises digital editions of Latin works of Croatian authors or authors strongly connected with Croatia. It encompasses texts written between the Middle Ages and modern times. The epistolographical collection was created within the research project *Croatica et Tyrolensia - a digital project of Croatian and Tyrolean neo-Latin*

¹ CroALa <http://croala.ffzg.unizg.hr/>.

*literature*². The project of Faculty of Humanities and Social Sciences in Zagreb and Ludwig Boltzmann Institute for Neo-Latin Studies in Innsbruck compared neo-Latin literature of these two regions bibliographically, textually and methodologically. The texts are of interest from a historical and cultural point of view as Latin was used in Croatia until the late 19th century, and the authors were a part of cultural, literary, and administrative life in Croatia as well as abroad.

The digital editions of letters are in XML format encoded by Text Encoding Initiative (TEI) standards. Each letter has its own unique identifier. The author, recipient, place of origin and destination, and date of the letter are marked according to TEI standards. Due to various reasons in the course of the project some letter files contain more than one letter. Every file contains the bibliographic information for the resource the letter (or letters) was digitized from. At the time of the proposal we had the list of 194 files³ that was made during this COST action. After the extraction of the metadata, there was a collection of 299 letters ready to be processed.

3. Short term scientific mission in Oxford

During my stay in Oxford I collaborated with the team of Cultures of Knowledge Project, in particular with Miranda Lewis. At the beginning of my stay we discussed the CroALa collection and the best way to incorporate the letters into the union catalogue. Miranda explained the EMLO workflow and the details concerning the metadata for the letters, persons and places as well as the ways of uploading the metadata to the catalogue. At the moment it is possible to upload a letter collection into EMLO in two ways: via Excel spreadsheets and via a webform. We decided that the Excel spreadsheets were the better option for this set of data. There were several reasons for choosing this method. Currently the webform requires each letter to be added one at a time. In the case of the CroALa collection, and because we wanted to consider the metadata standardization process, the spreadsheet seemed better because it was possible to export the dataset from CroALa into csv format, to prepare all the entities together, and to make global changes before the upload. It also reduced the possibility of making mistakes when copying and pasting across individual letter metadata from the TEI. In the webform it is easy to create a letter record and to add people and places connected to the letter by choosing them from a drop down menu. In order to do this, however, the people and places have to exist already within the EMLO catalogue. As the majority of people and place names from CroALa collection were not in the catalogue, it would be necessary to add them. At the moment the webform does not contain enough fields to enter all the necessary information for new people and the default method is to enter this additional information into the *Editor's notes* field (for example synonyms, roles, resources etc.) and then the EMLO editorial team reassign this metadata to the relevant field at a later stage. We decided that in this case spreadsheets were the quicker option. During the last days of my stay I was able to try the webform upload in the test base.

2 For more information on the project see <http://www.ffzg.unizg.hr/klafil/dokuwiki/doku.php/z:croatia-et-tyrolensia>.

3 The list of the letters: <http://solr.ffzg.hr/basex/croala-epist-cat>.

We exported the metadata from the CroALa XML, but encountered some problems during the export because of which the characters with diacritics were not rendered in the right way. This would have taken a lot of time for someone not familiar with the collection to process, but I was able to correct them. By entering people and places from Croatia and other countries which contain diacritics in their names, we were able to look into the possibilities of the search functions of the catalogue, and to begin to solve the problems with the great help of the developer Mat Wilcoxson.

The people and places records were also entered in bulk via bespoke Excel spreadsheets. As Miranda explained, this is a relatively new step in the EMLO workflow. In addition to the basic information for the people (names, dates), I set out into this spreadsheet synonyms, CroALa IDs, and URLs to VIAF, Wikipedia, and Croatian bibliographic lexicon where applicable. In place records I added links to Geonames and geo-coordinates.

The possibility of collaborating with the team in Oxford and working closely with Miranda saved time in preparing the collection for upload and enabled the problems we encountered to be solved more quickly and to add changes which were necessary due to the characteristics of the data and language. It also brought to our attention the information we did not mark in the annotations, in particular the degrees of certainty concerning people, places, and dates.

4. Outcomes

During my stay at Oxford I prepared the metadata of the letters, people and places connected to them, and we uploaded them into the back end of the EMLO catalogue for review in the area called EMLO-Collect. Because of the multiple letters in some files or more than one bibliographic record noted for the resource, it was not always possible to determine with certainty the exact manifestation of the CroALa letter in the course of the STSM. We decided that adding this information would be the next phase of the work. I started adding the data while I was in Oxford, and continued in Zagreb. Upon completion of this work, the collection will be made available to the end users.

The export from XML resulted in some letters which for the moment do not fit into the early modern letters catalogue, but which are a part of the CroALa collection. These include several letters from prefaces to printed editions, verse letters that were written with the intention to be published, and one letter dated in 1913. We have decided to enter these into the catalogue with a note specifying their type. For the moment they will not be visible to the end user, but rather would serve as a test base for including them in this or a similar catalogue in the future. The verse letters also raised the question of typology of the letters which would be useful to define for the needs of the Action.

In several of our talks Miranda and I discussed the guide for the contributors. The EMLO team has an excellent guide in English and keeps it up to date with all the additions and changes to the catalogue, as well as new methods that are being tested. We concluded that because of the different users and the diversity of correspondence projects, it would be optimal to have the guide in different formats: a text guide, a video tutorial, a hybrid one that would combine the screenshots of the process with a short description etc.

5. Further work

The second phase of the work outlined above is to define with certainty and enter the information on the manifestations, i.e. the sources of the digital editions of the letters. At the CroALa collection there are more letters that need to be annotated which we would add to the catalogue. The experience of working in Oxford enabled us to improve the work in CroALa considering the annotations and the standardisation of the metadata. We learned it was optimal to have each letter in a separate file, and to take into account the levels of certainty for a letter's author, recipient, origin, and destination. In the EMLO catalogue these can be noted as: *inferred*, *uncertain*, *as marked* in the letter. Dates may be flagged as *inferred*, *approximate*, or *uncertain*. Furthermore there is the possibility to add into the catalogue people and places mentioned in the letters. In CroALa we have marked these only on a small portion of letters. Through the Excel spreadsheets it will now be easier to add them after the annotation of the letters.