# MaRa v2.0

*Alfredo Cosco*

*alfredo[dot]cosco[at]gmail[dot]com*

https://github.com/orazionelson

**MaRa** stands for: **Marcatore dei Registri Angioini**

This software is part of a project of **University of Napoli "Federico II"**, that aims to build an XML archive, a database and a search tool for the **Southern Italian Angevine Canchery Papers**.

## A short historical intro

During the World War II, at the end of September 1943, a German army troop retreating from southern Italy burned the *Angevine Canchery Archive of Naples Kingdom*. The papers were hosted, for safety reasons, in a village, **San Paolo Belsito**, about 30km far from their building in Naples (then almost daily bombed by the Allies). The German retaliation destroyed thousands of folders containing parchments from the 13th to the 15th century.

After the war, the archivists of **Archivio di Stato di Napoli**, leaded by Riccardo Filangieri, went in search of the transcripts of the lost acts, administrative and political, which had been made over the centuries by archivists, scholars and historians. Then, they started to rebuild the archive in printed volumes. The reconstruction work continues today, it reached the number of 50 volumes and it covers a range of years from 1265 to 1434, under the main title: **I Registri della Cancelleria Angioina**.

## Software Dev for Angevine Canchery Papers in XML

Since 2004, Prof. Roberto Delle Donne, from *University of Napoli "Federico II"*, managed for a digital edition of the **Angevine Canchery sources**, with the aim to realize a research tool accessible through a web interface.

I was involved, as a student in Arts, at the early stages of the project.

Focusing on available options XML appeared as the most suitable format to store the data, then a DTD was defined and early transcripts where made by random documents.

A series of difficulties arised after this first trial:

- slowness to mark-up the documents,
- different tools used by editors,
- non-separation of concerns.

Moreover, the frustration of editing by hands full sets documents repeating tags hundreds of times, or the time spent to debug them when transcribed by other editors, brought to a different approach:

> **let's start from raw *.txt files and automate mark-up as much as possible.**

Then this target became the object of my degree thesis.

**MaRa v1** was a set of unordered PHP scripts with a minimal web interface. it was born to transform text transcrips (made in a standard format easier than XML) in well-formed XML.

**MaRa v2** is its evolution and was used to realize a first heavy draft of our XML repository.

All the scripts were rewritten or nested in CodeIgniter, a good option to add features with little effort, extended by an HMVC CMF called No-CMS, really useful and comfortable to build a user interface around the already written functions.

All the functions now are in *models/* and *controllers/* of *mara2/* module.

The new interface integrates an highly customizable text editor like CodeMirror in the framework in order to manage everything inside the browser.

There is a dedicated set of regex written for CodeMirror to highlight the text files with different colors: useful to check transcript formal errors. It's possible to see them in the file:

> registridev/modules/mara2/assets/scripts/angevin.js
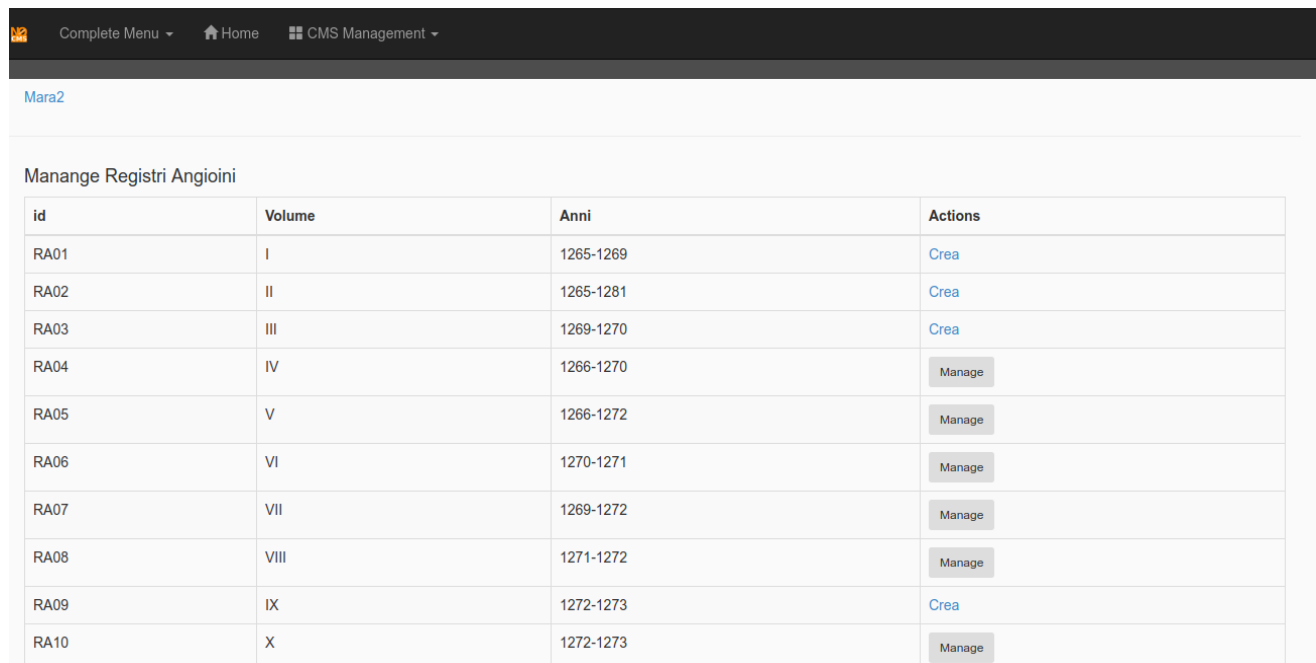
And, that's the most important thing, the job is designed in a flow:

> from the transcript import to the export in a set of xml valid documents, eventually packed in a zip file.

# MaRa2 workflow

The starting point is the panel in:

[localhostPath]/*mara2/mara2* controller, here you have a TOC based on printed editions from which the flow can be managed.



You can create a registry and import the text file.

Then you can **Manage** the registry in 5 steps.

## Few notes about the transcript standard

Each *registry* transcript is made by one *.txt* file, it contains different *partitions*, each partition contains from 1 to n *documents*.

Each partition is separated by the tag:

```
#PARTITION#
```

The first line after the tag is the **partition title**.

Then the documents.

Each document must have this structure:

```
[id]. - [title]
[text][primary-sources]
FONTI:[secondary-sources]
```

The pagination follows the printed text, like this example:

```
#PARTITION#
[partition title]
[page number]
[document]
[document]
[document]
[page number]
[document]
[document]
[document]
...
```

# The STEPS

## Step1a

This **STEP** splits the entire registry in partitions using the tag **#PARTITION#**.



## Step1b

Same as before but it adds page numbers to each document
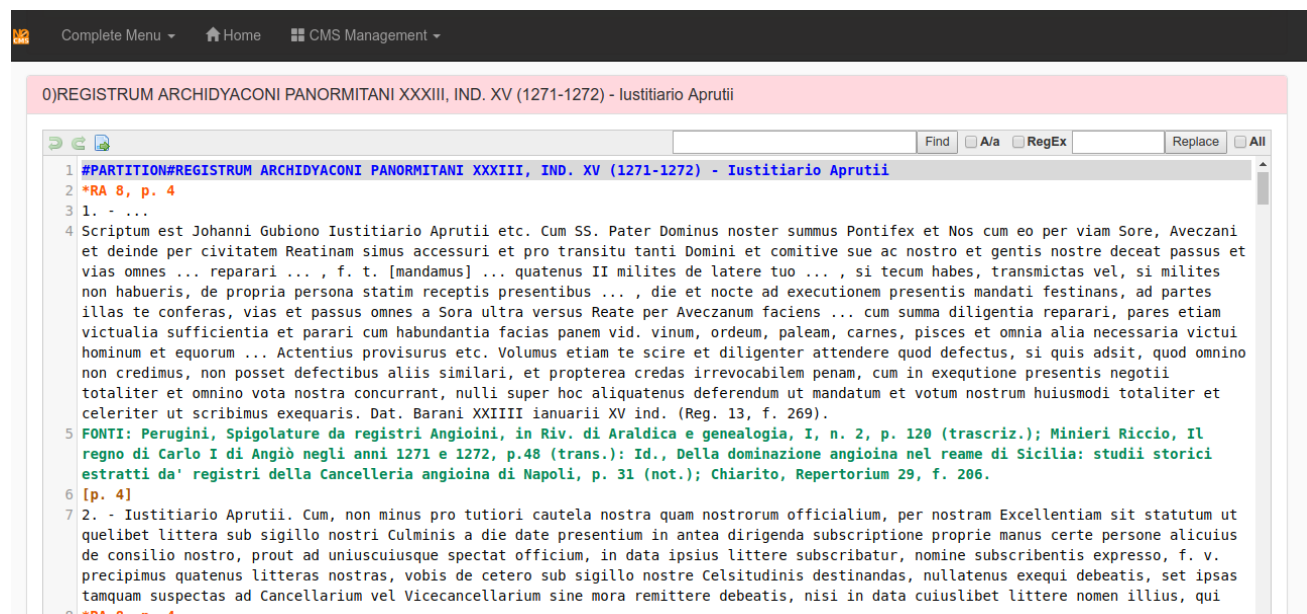


## Step2a

This is the core of the application, it inherits most of **MaRa1** functions. The software splits the partitions in single documents and import them in 4 db tables:

- ra_mara_RXX_index: and index of all registry partitions with their status: true (for the currrent editing partitions), false (for the partition to edit), approved (for the edited partitions);

- ra_mara_RXX_document: for the document data: Title, Text, Pagenumber;
- ra_mara_RXX_regorig: for the document primary sources;
- ra_mara_RXX_source: for the document secondary sources.

*(where XX is the registry number)*

Each table is analyzed in the tabs in the page.



The software normalizes the most common errors appeared in editing using **regular expressions**. Sometimes it correct them, in other cases an errors control shows a message.

**Sample error:** in the next image the document title has not beed separated by the text.



Is possible correct the error in the integrated editor.

**Sample error:** in the next image the software alert for a a leak of page number.



**Sample error correction:** in the next image the software advise for an autocorrection.



**Sample Corrections**: in the next image the software advise you and correct the notation *ibidem* taking the info from the previous record.

The tab for the secondary sources, called **Fonti**, will split and normalize them.

See in the next image that the notation *Ms. Bibl. Naz* changes to *Biblioteca Nazionale di Napoli, ms.*



It is possible to edit the transformation db from the sub-tab **Casistica di trasformazione**

Sometimes the splitted vision of the sources shows you that something is wrong, see the next image:
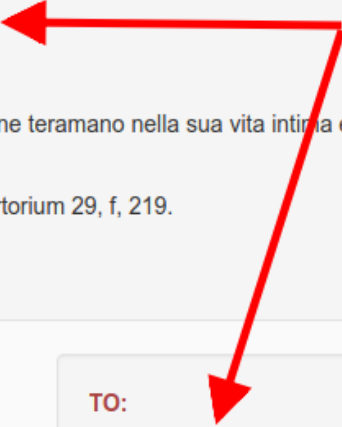


Once there are no more errors it is possible to click on the button **Approva** to release the partition.

This will lock the partition and will unlock the next one.

When all partitions are completed on the page appears the **Crea XML** that leads you to the next step.



## Step3a

This step will transform the records in XML files.

Converti in XML

In questa interfaccia trasformeremo i db con le trascrizioni in file XML

Database trasformati

R01  R02  R04  R05  R06  R07  R10  R11  R13  R14  R17  R18  R20  R40  R44  tmp

Database da trasformare

RA08

☐ Seleziona tutti

☐ RA08.01) Registrum Archidyaconi Panormitani XXXIII, Ind. XV (1271-1272) - Iustitiario Aprutii
☐ RA08.02) Registrum Archidyaconi Panormitani XXXIII, Ind. XV (1271-1272) - Iustitiario Terre Laboris et Comitatus Molisii
☐ RA08.03) Registrum Archidyaconi Panormitani XXXIII, Ind. XV (1271-1272) - Iustitiario Principatus
☐ RA08.04) Registrum Archidyaconi Panormitani XXXIII, Ind. XV (1271-1272) - Iustitiario Basilicate
☐ RA08.05) Registrum Archidyaconi Panormitani XXXIII, Ind. XV (1271-1272) - Iustitiario Vallis Gratis
☐ RA08.06) Registrum Archidyaconi Panormitani XXXIII, Ind. XV (1271-1272) - Extravagantes
☐ RA08.07) Registrum Iustitiariorum XV Indictionis XXXIV, Ind. XV (1271-1272) - Iustitiario Principatus
☐ RA08.08) Registrum Iustitiariorum XV Indictionis XXXIV, Ind. XV (1271-1272) - Iustitiario Sicilie
☐ RA08.09) Registrum Iustitiariorum XV Indictionis XXXIV, Ind. XV (1271-1272) - Documenti senza citazione di fonte
☐ RA08.10) Registrum Secretorium XV Indictionis XXXV, Ind. XV (1271-1272) - Sceretis Principatus Terre Laboris et Aprutii
☐ RA08.11) Registrum Secretorium XV Indictionis XXXV, Ind. XV (1271-1272) - Secretis Apulie
☐ RA08.12) Registrum Secretorium XV Indictionis XXXV, Ind. XV (1271-1272) - Secreto Calabrie
☐ RA08.13) Registrum Secretorium XV Indictionis XXXV, Ind. XV (1271-1272) - Secreto Sicilie
☐ RA08.14) Registrum Secretorium XV Indictionis XXXV, Ind. XV (1271-1272) - Magistris Portulanis
☐ RA08.15) Registrum Secretorium XV Indictionis XXXV, Ind. XV (1271-1272) - Magistris Massariis
☐ RA08.16) Registrum Secretorium XV Indictionis XXXV, Ind. XV (1271-1272) - Siclariis, Distributoribus nove monete et Magistris tarsienatuum.
☐ RA08.17) Registrum Extravagantium XXXVI, Ind. XV (1271-1272) - (Extravagantes Infra Regnum)
☐ RA08.18) Registrum Cancellarie XXXVII, Ind. XV (1271-1272) - Extravagantes
☐ RA08.19) Registrum Cancellarie XXXVII, Ind. XV (1271-1272) - De matrimoniis
☐ RA08.20) Registrum Cancellarie XXXVII, Ind. XV (1271-1272) - Privilegia
☐ RA08.21) Registrum Cancellarie XXXVII, Ind. XV (1271-1272) - Apodixarium
☐ RA08.22) Registrum Cancellarie XXXVII, Ind. XV (1271-1272) - Quaternus extractorum victualium.
☐ RA08.23) Registrum Cancellarie XXXVII, Ind. XV (1271-1272) - Vicario et Camerario Urbis, Vicario Tuscie, Senescallo Lombardie, Ballivo Andegavie, Senescallo Provincie et aliis extra Regnum.
☐ RA08.24) Registrum Cancellarie XXXVII, Ind. XV (1271-1272) - Extravagantes extra Regnum
☐ RA08.25) Registrum Cancellarie XXXVII, Ind. XV (1271-1272) - Lictere de creatione officialium sub mag. Simone de Parisius Regni Sicilie Cancellario.
☐ RA08.26) Registrum Cancellarie XXXVII, Ind. XV (1271-1272) - Documenti senza citazione di fonte
☐ RA08.27) Quaternus Iustitiariorum Capitanie Karoli Primogeniti XXXVIII, Ind. XV (1271-1272) - Iustitiario Principatus
☐ RA08.28) Quaternus Iustitiariorum Capitanie Karoli Primogeniti XXXVIII, Ind. XV (1271-1272) - Iustitiario Basilicate
☐ RA08.29) Quaternus Iustitiariorum Capitanie Karoli Primogeniti XXXVIII, Ind. XV (1271-1272) - Iustitiario Capitanate
☐ RA08.30) Quaternus Iustitiariorum Capitanie Karoli Primogeniti XXXVIII, Ind. XV (1271-1272) - Iustitiario Terre Bari
☐ RA08.31) Quaternus Iustitiariorum Capitanie Karoli Primogeniti XXXVIII, Ind. XV (1271-1272) - Iustitiario Terre Idronti
☐ RA08.32) Quaternus Iustitiariorum Capitanie Karoli Primogeniti XXXVIII, Ind. XV (1271-1272) - Iustitiario Vallis Gratis et Terre Iordane
☐ RA08.33) Quaternus Secretorum Capitanie Karoli Primogeniti XXXIX, Ind. XV (1271-1272) - Secreto Calabrie
☐ RA08.34) Quaternus Extravagantium Capitanie Karoli Primogeniti XL, Ind. XV (1271-1272) - Extravagantes de Curia
☐ RA08.35) Quaternus Extravagantium Capitanie Karoli Primogeniti XL, Ind. XV (1271-1272) - Extravagantes de privatis
☐ RA08.36) Quaternus Extravagantium Capitanie Karoli Primogeniti XL, Ind. XV (1271-1272) - Apodixarius

Genera XML

Back to top

## Step3b

The last step is dedicated to validate the XML. It is possible validate the entire transcription or the single document.
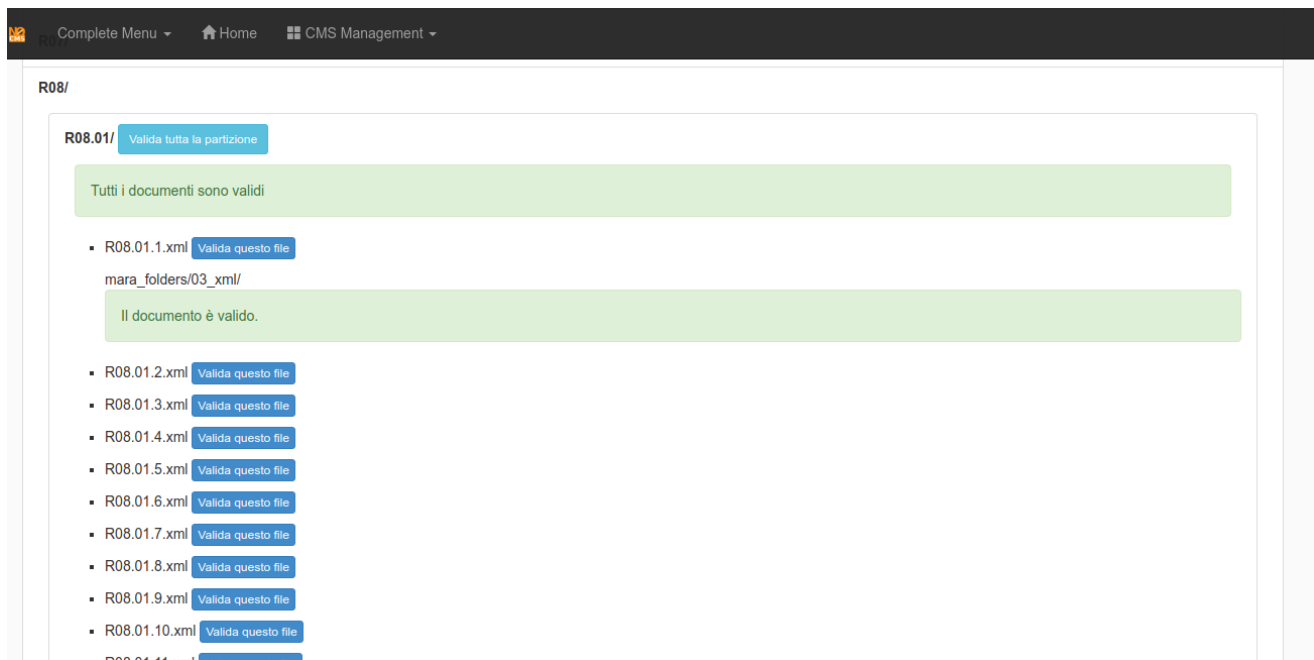
Then at the bottom of this interface is possible to pack each generate regisin a zip file or create a TOC of them.
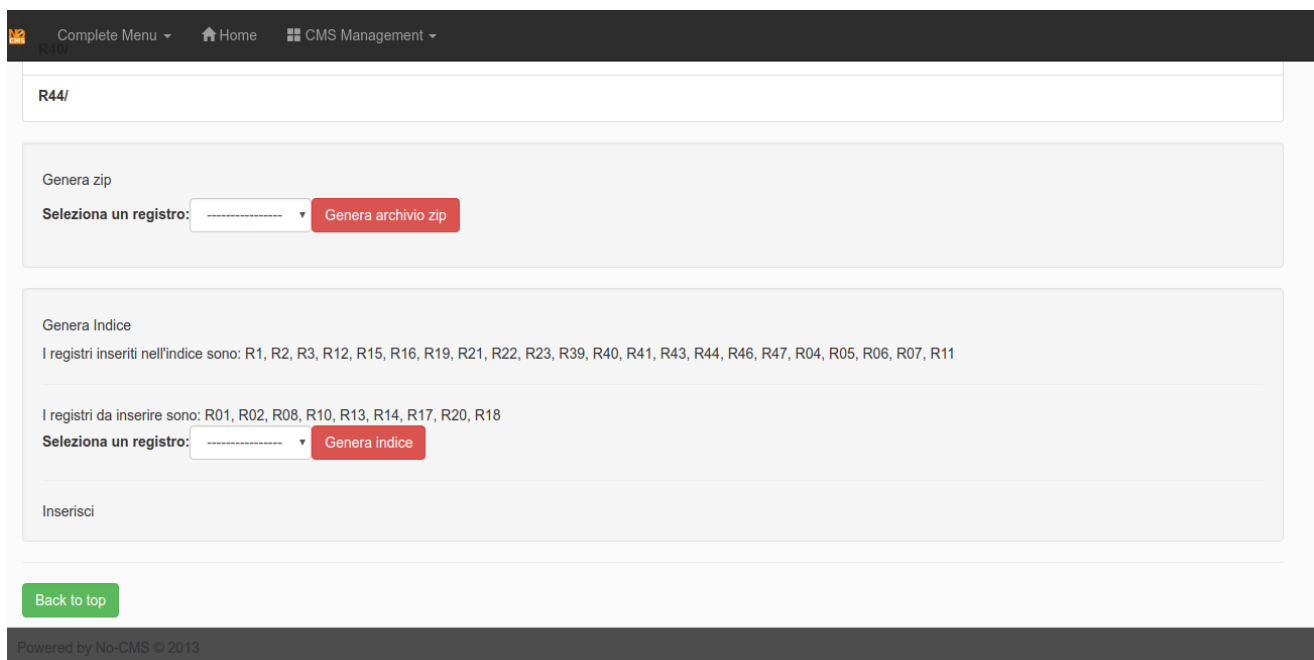


The workflow is strictly connected with the filesystem organization inside the directory: try

- *mara_folders/*

the sub-directories:

- *00_originali/* Contains the original transcripts
- *01_trascrizioni/* Contains the transcript at step1
- *02_tocpic/* Contains the images with the TOC of each transcript
- *03_xml/* Contains the generated XML documents
- *04_zip/* Contains the zipped folders with documents
- *05_rollback_files/* Transcript txt files regenerated from the DB during the flow if needed by the editors for further corrections

- *06_toc/* Contains a useful toc of generated documents
- *maradb/* Contains a text-db used for the sources normalization process

##Targets

Almost the half of the entire documents set edited in **I Registri della Cancelleria Angioina** has been tranformed to XML using MaRa v1 and v2.

# Notes about the software installation for testing

**MaRa2** is designed for a localhost use, its specifical features make its use for a different projects a *nonsense*.

Nevertheless, its installation and use comes as common CMS:

- unzip the package to your apache web-server, so your path will be: http://localhost/registridev
- import the **registridev.sql.gz** file in a MySql table called: *registridev*
- set the CodeIgniter config files as any CI application with you server and mySql data
- set the *registridev/* directory (and everyting inside it) permissions to 777
- log-in on **registridev/main/login** as admin/admin
- test the application, go to **registridev/mara2/mara2**

If you try an installation you will find a sample registry (R11) ready for the **STEP2A**, you can try to find and correct some errors then transform it in XML files.

# References

*I registri della Cancelleria Angioina*, Napoli, since 1950

Riccardo Filangieri, *L'Archivio di Stato di Napoli Durante la Seconda Guerra Mondiale*, Napoli, 1946

Jenkinson, Hilary, Sir & Bell, H. E. (Henry Esmond) & British Committee on the Preservation and Restitution of Works of Art, Archives, and other Material in Enemy Hands (1947). *Italian archives during the war and at its close. H.M.S.O*, London

PHP: http://www.php.com

CodeIgniter: http://www.codeigniter.com

jQuery: https://jquery.com/

No-CMS: https://github.com/goFrendiAsgard/No-CMS

CodeMirror: https://codemirror.net/