Valeria Frazzica*, Francesca Galletti*, Massimo Orciani*

Gregoriano Cadastre: the Creation of a WebGis from Historical Cartography through the Techniques of Classification of Satellite Images

Keywords: Gregoriano Cadastre; historical data information system; sommarioni (property registers); classification technologies; WebGis portal

Summary: The project was developed with the intention of making available a long process of data acquisition, vector and not, derived from maps and brogliardi (property registers) of the Gregoriano Cadastre (census data of the Papal States from the early decades of the nineteenth century and preserved in the State Archive of Rome). This acquisition, in first instance, concerned the raster maps geo-referencing from State Archive and then, the manual digitizing of identified thematic topics (such as agricultural plots, main urban centers and scattered houses, waterways, roads and toponyms). Vector data have been associated, through the creation of a relational database, with crop production and land owners socio-economical information contained in the registers; this allowed the creation of a historical GIS, together with the reconstruction of the 3D model of the nineteenth-century landscape, concerning a significant portion (approximately 8% - 167 sq km) of the province of Ancona.

This first part of the work has been used as an instrument of knowledge and analysis of the best practices for acquisition, storage and management of data and information derived from this documentary source. The use of information technologies derived from the classification of satellite images, made possible a large scale vectorizing (about 35% - 700 sq km) of the themes present in historical maps. Next step was then to implement a WebGIS portal in order to make available, to a wide audience, raster data extended to the whole province, as well as the vectorized portion, taking care to propose a cartographic base suitable for a direct comparison with the current situation. The instruments provided for the interactive viewing of the portion already completed foreshadow an integral and complete historical data information system obtained from original maps.

The system also provides a direct link to the project Imago II of the State Archive of Rome, for the consultation of the brogliardi raster data for the territories for which is not yet available the specific database associated with the land parcels.

Introduction

This paper describes the various steps that led to the creation of a cartographic database associated with the maps of the Gregorian Cadastre. It is therefore necessary, to start with a description of the maps themselves and their technical characteristics which were closely linked to the use for which they were intended in order to introduce the successive stages of processing.

The Gregorian Cadastre is the first particle Land Register of the entire Papal States, promoted by Pius VII in 1816 and activated by Gregory XVI in 1835. It includes three related series: the maps (1:2000), the *mappette* in small scale (1:4000 or 1:8000) and the landowners' property registers (called *brogliardi* or *sommarioni*).

The most complete copy of these maps, unaffected by the changes that use necessarily entailed, is kept in the State Archives in Rome. These maps provide the most complete picture available of spatial and urban planning in the provinces of the Papal States

* Provincia di Ancona – Italy [v.frazzica@provincia.ancona.it] [f.galletti@provincia.ancona.it]

(www.asrm.archivi.beniculturali.it/).

Both *mappetta* and maps were created for each sheet. In a scale of 1:2000 and covering rather large portions of territory, these maps are often very large in size (up to about 4 x 2 meters). For this reason a series which was still in a scale of 1:2000, but divided into quadrants was also produced. This is mostly what is to be found currently at the local State Archives and referred to as 'maps' while the whole tablets in a scale of 1:2000 are often called *mapponi*.

Initially were purchased from the State Archive of Rome raster images of *mappette* in digital format (.tiff) that would ensure a coverage of the entire territory of the current province of Ancona. At the time of drafting up, the maps of this territory were divided between the delegations of Ancona, Pesaro and Macerata.

The method used to create a mapping database had, as its goal, the implementation of these points:

1. geo-referenced raster coverage, searchable and comparable to the whole map database already present;
2. geo-referenced vector coverage relating to the boundaries of individual sheets, such as an index map of all the *mappette*;
3. geo-referenced vector coverage relating to all the cartographic themes present in the *mappette*;
4. a relational database that contains all the alphanumeric data present in the *sommarioni* and also the data relating to each sheet.

For the vector coverage of the various cartographic themes, the acquisition of *mappette* in a scale of 1:4000 was chosen because these contain a range of very detailed information. Their size makes transferring them to digital format straightforward without the use of a scanner, which could damage the maps themselves and maintains orthogonality over the entire surface and a resolution that makes it easier to read the contents.

The State Archives of Rome created these images, one per sheet, and a database consisting in a table containing the basic information about each sheet.

The individual sheets, covering an average of about 735 hectares of land, have been geo-referenced in Gauss-ED50 East coordinates, using the Regional Technical Map of Marche in a scale of 1:10000 as a baseline. On average 40 points of correspondence for each sheet were identified approximately 1 point per 18 acres.

As often happens in historical registers, even in the case of the Gregoriano Cadastre, the boundaries of individual sheets, based on elements present in the area such as roads or rivers are not 'topologically accurate'. Once geo-referenced, they very often overlap. It was decided not to force this aspect, leaving the overlay, because in the case of dimensional analysis, such as the sum of sheet surfaces, it was important not to alter the value of individual sheets themselves in order to render the surface data verifiable with measurements incorporated in the *sommarioni*.

The situation in relation to the individual themes is a different matter.

The cartographic features, drawn in 'mappette', have been vectorized by means of a digitization process, separating them into five classes:

a) hydrographic network
b) road routes
c) place names (toponyms)
d) built areas
e) agricultural plots.

The criteria adopted in the digitalization were developed taking into account two main aspects: the first was to make the best mapping information and the second was to find the most easily used topology as well as the best typing data.

Information layers were kept separate in the first phase of the work, for an easier management of the acquisition and implementation phase of associated fields. Once the digitalization was completed by means of a field of typing level they were incorporated in order to obtain a single coverage.

This was possible for the first three classes and for a specific subset of the fourth, while the fifth, as we will see, followed a different procedure.

The digitization of information concerning **hydrography**, such as ditches, rivers and streams, was carried out mainly using 'linear elements' corresponding to the centerline of the cartographic element. A break was necessary for each branching point and for each sheet when the crossing was a transverse. In event of a water course at the border, the correct position was derived from the overlap mediation. Greater weight is given to the sheet where the river is actually drawn as compared to the sheet where the path bordering the river derives from the shape taken by the plots on the border. The most important waterways have been digitalized as 'polygonal entities' in only a few cases in order to recoup the riverbed in the middle of the nineteenth century. This choice proved to be useful for comparisons with the current riverbed of the watercourses themselves. The presence of puddles or ponds was reported as punctual entities associated with place names.

The same method was applied to **roads**. The main roads were digitalized along the centerline. The plots on the border do not come into contact with the tract. Once land plots were digitalized, the width of the carriageway became clear.

By analyzing the type of graphic sign and the different thickness of road chosen by the designer of the *mappetta*, it was possible to hypothesize a classification of such paths according to their importance. The presence of toponyms along the main road layout also made it possible to envisage a hierarchy (e.g.: departmental road, municipal road, local road, etc.). Local roads were digitalized as lines between properties. The presence of bridges was reported as a punctual entity associated with the corresponding toponym.

As for the **place names**, the *mappette* report the names of the main watershed (rivers, streams, ditches), morphological (surveys, hills, ridges) and anthropogenic (roads, bridges and urban centers) elements. Each name was entered as 'item on time'.

Built areas were treated using two levels both consisting of polygons as in the map itself. In fact, built areas belonging to villages and towns were inserted into the map on a scale of 1:4000 with graphic elements shown schematically respecting the overall shape of the area without any associated plot, while in the edge of the sheet the details of the plot division on a scale of 1:2000 have been reported with its cadastral maps.

Two coverages were thus made as a final result: the first relating to the whole built area with the urban areas shown in schematic 1:4000 version and a second with the built area in detail and, as will be seen later, the connection with plot data through *mappali* (plot numbers). This second level, also geo-referenced, describes the same portion of land, perfectly overlapping the first, but accompanied by more detailed information as it is connected to the plot and therefore to the *sommarioni* as can be seen below.

In the case of non urban buildings we have identified two cases:

1. buildings corresponding to a single plot of its own;
2. buildings associated with an attached plot of land.

In the first case the polygon was linked to the *mappale* of the plot, in the second case the *mappale* was linked to the land. Individual particles were digitalized as polygons following elements designed in the *mappette*. Each record was linked to the relative plot number (*mappale*).

Vectorization of extensive territories

The need for the information deriving from the Gregoriano Cadastre for the whole territory of the province of Ancona lead us to experiment with quicker acquisition procedures and ways of resolving difficulties relating to the size of the analysis area.
The work initially carried out had no variously 'automated' elements and therefore required a very high commitment of dedicated resources especially in terms of hours of work.
As it was a question of raster interpretation, however, we wanted to experiment on this historical cartography, with the application of the techniques developed for the classification of satellite images starting with the chromatic value of each pixel, expressed as a number, and continuing with classes of land use and the subsequent vectorizing of these.
Thus, starting from the classes already identified above, we worked with a commercial company which specializes in this type of analysis with the aim of obtaining a very similar product to that obtained for the territory which had already been worked 'by hand'.
From the outset it became clear that there was a reasonable chance of identifying issues the graphic themes but the calligraphic part, consisting of the plot number and its name, could not be done without the direct reading and interpretation of the operator.
Work therefore concentrated on the purely graphic themes.
The ultimate goal was to obtain classified maps in vector format with + / - 1 pxl accuracy in relation to the original raster with the following classes:

- Hydrography
- Roads
- built areas
- agricultural plots

using the work already done by the office team as a prototype.
On this prototype the company analyzed the results of the processes of the most commonly used classifications: classification by automatic segmentation of the area/plot, classification by semi-automatic and automatic tracking of the lines and spectral classification and object, integrating a range of techniques.
From these first verification it emerged that the quality of the scanning and of the original support, and deterioration such as blemishes, lines, partially interrupted hatches, etc. were such that appropriate results with completely automated processes could not be guaranteed. In order to reconstruct the limits of the plots a partial manual integration of the results was therefore necessary. To obtain a complete reconstruction of the plot boundaries, unidentified or incorrectly defined boundaries were worked on (partial editing), while the result obtained directly from the automatic process proved good for the other four classes defined.
These results were compared with a completely manual process of plot identification and classification editing performed with the help of SW which allowed us to facilitate the work with supervisors. These operate on a locally defined tract identified by the operator himself.
It thus emerged that the accuracy obtained from the manual process is in general superior to the automatic or semiautomatic process. However, a refinement of the map with an initial homogenization of support through appropriate intervention, allowed for improvement of the

results of the automatic process. An initial phase of 'cleaning' or homogenization of the substrate, a phase of automatic segmentation of the plots and subsequent refinement and manual completion was therefore chosen as a working hypothesis.

The technique used thus provides a mixed automatic and/or manual methodology assisted by specific line follower software, with an automation percentage forecast at process end of around 50%.

Work phases can thus be summarized as follows:

- Homogenization and filtering

Homogenization: preparing images by radiometric homogenization (shaded areas) and cancelling elements which distort the subsequent processing steps:

1. elimination of the elements of map legend and the cartography framework
2. elimination of the folds of the sheet
3. matrix elimination at the bottom and of light and dark effects
4. elimination of overlapping nomenclature on map elements

The results of homogenization and filtration are exemplified in the image below.

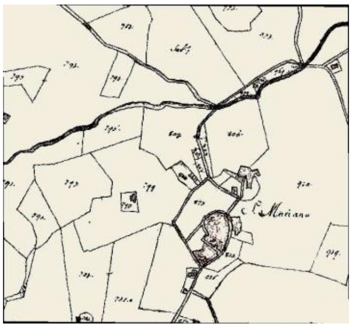

Figure 1: Object-oriented segmentation.

Creating a database of geographic vector polygons resulting from a multi-level segmentation process implemented through specific software.

Once you have recreated the hierarchical structure of the system of nomenclature, a first level of segmentation on the basis of the available images is provided. The segmentation, through a process of repeated trial and analysis of the results obtained, will be such as to produce the least number of polygons still homogeneous to the hierarchical level chosen.

- Allocation of legend entries

The polygons generated in the segmentation phase in this phase must be associated with the system of nomenclature adopted. To this end, a fuzzy approach has been adopted whereby each polygon will be associated with a chance (expressed in the range 0-1) linked to each of the classes of land-cover use adopted. Only in the cartographic data phase will a hardening operation be implemented associated with each polygon class use/land cover with the highest value possible.

Note that while the segmentation process starts with the data with the least detail, and proceeds with that in greater detail, the process of classification is based on the principle of aggregation: the most detailed level is classified thematic from which, thanks to the relations of the database multi-level, is derived the least detail.

- General parameters for the O-O classification

The parameters for object-oriented classification relate to the following information:

1. spectral: statistical data calculated for each polygon from the values of the digital number of the respective pixels belonging to each of the spectral bands available (three resulting from the

separation of the RGB channels, plus any bands derived from the filtering activity) such as mean, standard deviation, minimum, maximum, etc.;

2. geometrical information derived from the geometrical characteristics of the polygons originated with the segmentation process. The most common are area, perimeter, fractal dimension (perimeter area ratio), or other form factors. This category also includes the information about the spatial location of the polygons (e.g. further or less to the north, further or less to the east).

3. textural: information derived from the analysis of the spatial structure internal to the polygons obtained by segmentation. These may be derived from the analysis of the individual pixels contained in the polygons or polygons resulting from the analysis of a process of lower level segmentation (i.e., with a smaller scale factor).

4. hierarchical: statistics calculated for each polygon of the object of the classification level of segmentation on the basis of related polygons of a different hierarchical level (e.g. the number of polygons connected on the higher hierarchical level or lower).

5. thematic attributes: information from layers of information accessories, ancillary to the bands of the remote sensing image used.

- <u>Review and manual control</u>

The classification obtained by means of object-oriented logic should be handed over for careful scrutiny by experts in photo-interpretation. These will be responsible for validating each polygon and then accepting or declining the assignment automatically.

Photo-interpreters would also be able to refine the geometry of the polygons in order to make them perfectly congruent with the project specifications.

The shell definitive in vector format will then be exported to the required formats and features for final delivery according to design specifications.

- <u>Specific parameters for the classification of historical maps</u>

Specifically, segmentation and classification rules were drawn up according to a process called Process Tree as in the following list:

1. Multi-resolution Segmentation
- Extraction or Hydrography
2. Segmentation of other classes
- Built area extraction
- Road extraction
- Plot extraction

The following figures show the main steps of the process of automatic segmentation and classification.

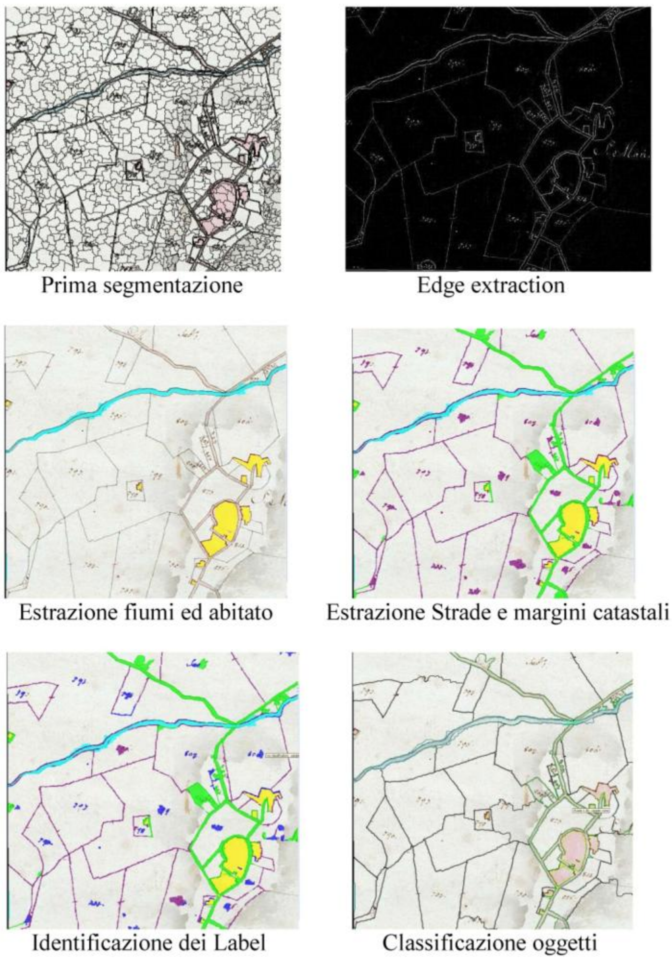

Figure 2. The main steps of the process of automatic segmentation and classification.

Database

A database was organized and structured that can handle all information levels, both cartographic and alphanumeric, derived from the Gregoriano Cadastre.

It is therefore a relational database with spatial data that can be managed by software which provides this function, such as Oracle. It can also be managed, though less immediately, with an alphanumeric database (such as ACCESS) joined to a software GIS able to relate to the former.

Information derived from the main issues related to mapping and *sommarioni* data, have been inserted. Specific attention was devoted to land uses and ownership in an attempt to make queries on the implementation of these features as smooth as possible.

The information relating to the sheets is contained in the alphanumeric part, while the cartographic part, substantially the perimeters of the sheets themselves, has been separated. This was done in order to make software management, spatial or otherwise, possible.

The following diagram summarizes the structure of the database. Tables that do not contain map data are in blue, those that contain them are in green.

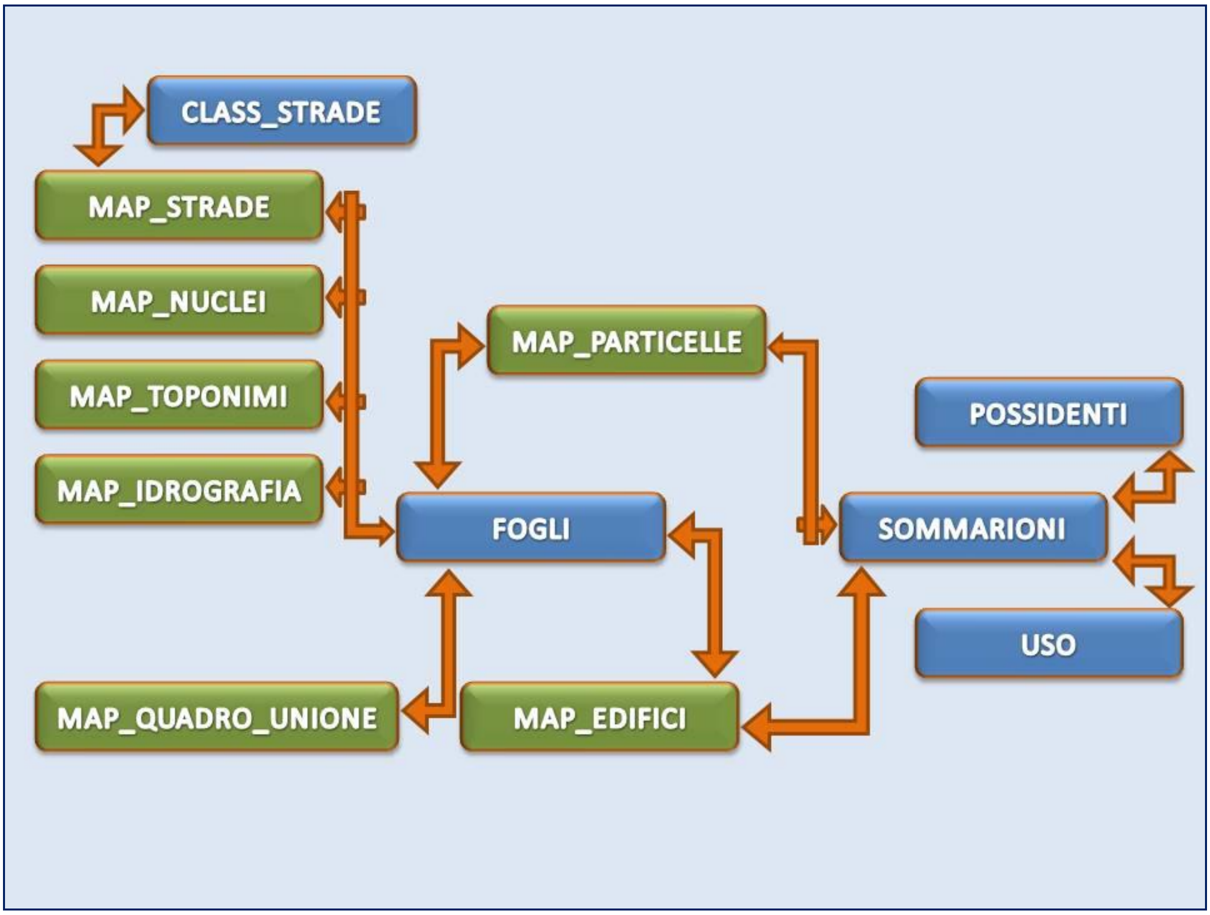

Figure 3: The structure of the database.

From the many possible choices relating to the relationships within the structure of the database, we decided to use the option which could guarantee the immediate recognition of themes and content in order to make the database easier to use and manage.

Regarding the type of relationships between the database elements, it was necessary to create a unique relationship between the plots contained in the MAP_PARTICELLE table whose records are related to each plot and describe their ground plan, and the *SOMMARIONI* table which contains the text and numerical information derived from *sommarioni*, such as, for example, ownership, use, surface, etc. This unique relationship can be obtained with a field code key speaker (e.g.: AAA_NNN_BBBB), but we preferred to make a multiple key which, in this case, is composed of fields: Sheet-Plot-Subsection-Fertility, which together make up a unique key, though, pending the full completion of the database, we decided to retain a key field in code talking, of course, deriving from the same fields as well.

Special consideration should be made for the theme of the built area, because this has special features arising from the type of map that has been used.

The database, as mentioned above, is based on the acquisition in digital raster format of the series of *mappette* in reduced scale (1:4.000). In these *mappette* the center is represented: as a single element relating directly to a *mappale* (plot number) or as a single or linked element which is not directly reported on a *mappale* but rather with a magnification, drawn on the board table, outside the perimeter of the sheet, where the plots are specified (Fig. 4).

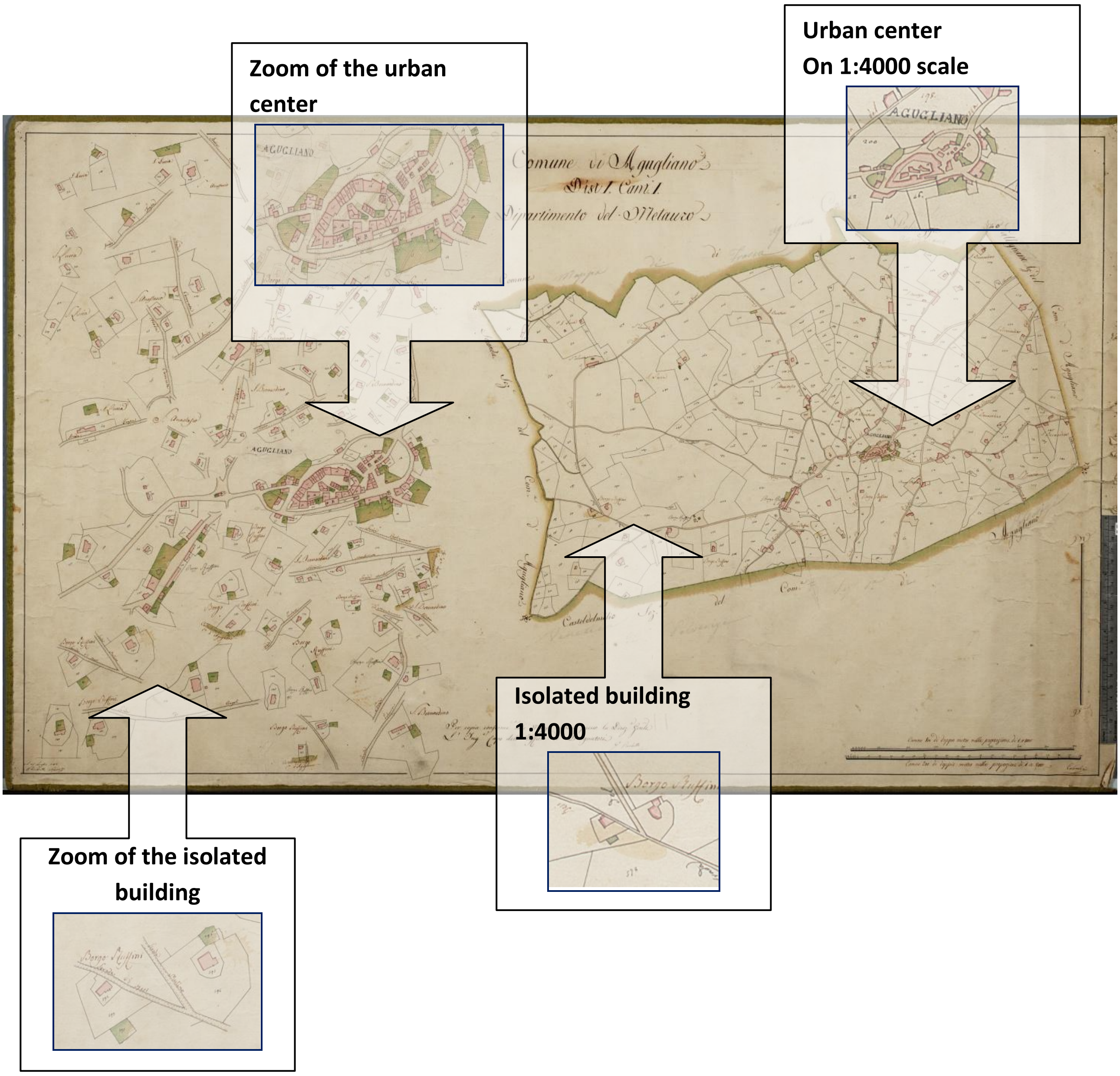

Figure 4: Different scales were used for cartographic representation.

This particular state of the cartographic representation, for which, inter alia, the magnifications described above could not be geo-referenced with the rest of the sheet, has forced us to divide the theme of the built area into two separate tables, one relating to the first case above (MAP_EDIFICI) and one relating to the second in the specific case of associated elements (MAP_NUCLEI).

For buildings that show magnification on the side of the table, we opted for two work phases: the first involving the insertion of the perimeter of the building indicated inside the sheet, and then on a scale of 1:4000, without compiling the number field of the particle, and a second phase, taking the data from the magnification, enabling the field for the number of plots to be compiled.

WebGis

Our decision to make the work done available to everyone was taken as a result of the considerable interest that the research has provoked and not only among insiders. A large number of requests have been received from academics, local authorities, institutions and professionals but also people interested in the history of the places they live, work and spend their leisure time in.

To publish these and other mapping and georeferenced data, the Geographic Information System has implemented new cartographic services based on proven WebGis technologies.

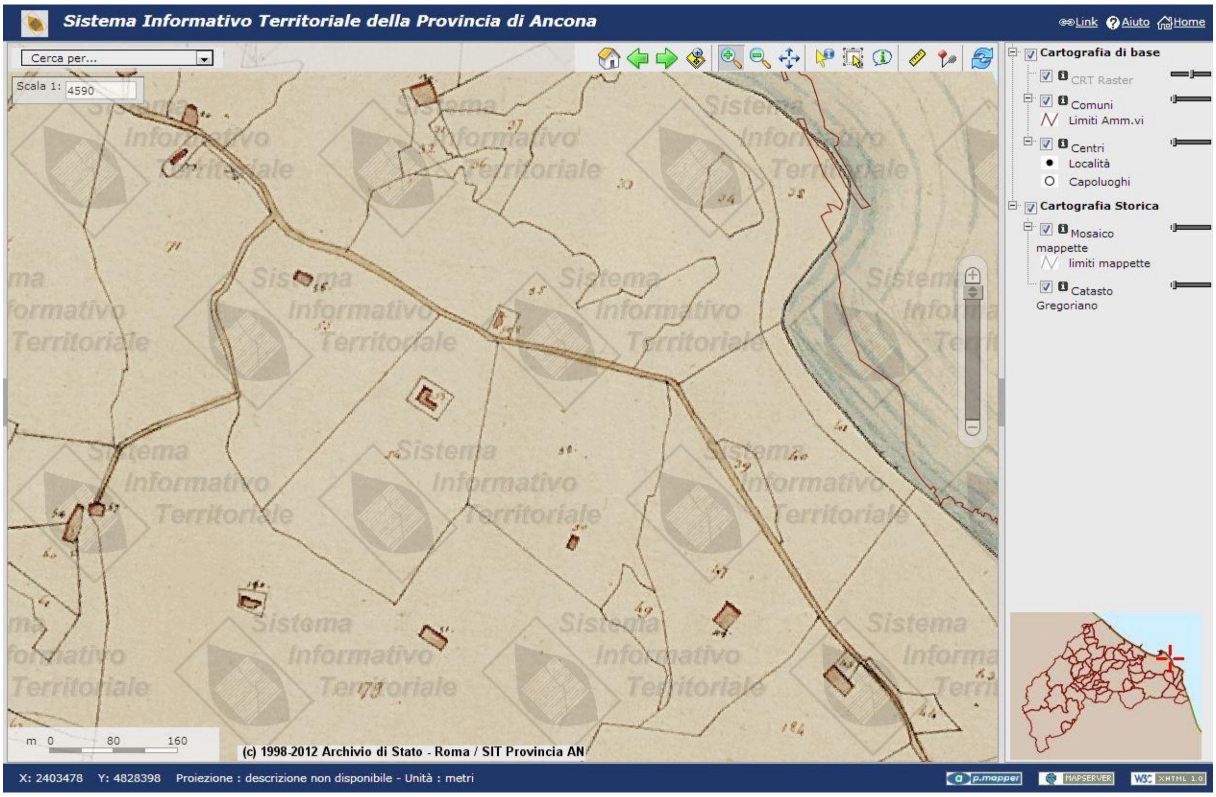

Figure 5: A specific map was designed to contain the Gregorian Cadastre information and - in future - all available historical maps.

This service is available on the ***internet*** on http://sit.provincia.ancona.it/sit/.

The Gregoriano Cadastre map currently contains the following information levels:

1. A raster base map of the whole province (original scale 1:25000) for comparison with the current situation;
2. The administrative borders of the current municipality;
3. Centres and current locations for easier search;
4. A mosaic of the boundaries of all the Gregoriano Cadastre *mappette* with the data characteristics of each *mappetta*;
5. A mosaic of geo-referenced raster *mappette*. This level of information has been modified by a light watermark because of image rights.
6. The geographical areas for which additional information is available, such as vector elements and associated data derived from the maps themselves or from the *sommarioni*;

7. Vector objects divided into classes: particles, buildings, urban centers, roads, hydrography, place names.

The system allows a high degree of customization of the map and the information visible for any part of the province. There are also specific information search specific tools:

- search by common or current location;
- selection of mappetta of interest;
- search by plot number and mappetta;
- plot search by owner's name;
- search by road, waterway and toponym name.

The service is totally accessible using a standard browser (Internet Explorer, Firefox, Chrome, etc.) and has been developed in-house by using exclusively Open Source free access tools hosted on their systems (Windows Server 2003 with IIS 6 web server).

The MapServer ver. 6.2.1 was used as a data rendering engine for web mapping. This software is now a project of OSGeo, and is maintained by a growing number of developers (around 20) from around the world. It is supported by a diverse group of organizations which fund enhancements and maintenance and administered within OSGeo by the MapServer Project Steering Committee made up of developers and other contributors. Functions included are:

- Advanced cartographic output:
- Support for popular scripting and development environments;
- Cross-platform support;
- Support of numerous Open Geospatial Consortium (OGC) standards;
- A multitude of raster and vector data formats;
- Map projection support.

MapServer is supported by ***p.mapper*** ver 4.2.1; the p.mapper framework is intended to offer broad functionality and multiple configurations in order to facilitate the setup of a MapServerapplication based on PHP/MapScript. Functions included are:

- DHTML (DOM) zoom/pan interface (no use of frames);
- Fully featured query functions (identify, select, search);
- Highly flexible configuration of functions, behavior and layout;
- Multilingual user interface;
- Plugin API to add custom functionality.

Future developments

The reading of *sommarioni* and manual compilation of the database linking the number (*mappale*) to the corresponding plot are extremely costly operations in terms of human resources and time. Therefore systems and semi-automatic methods for data acquisition should be pursued.

Figure 6: Example of human calligraphy in *sommarioni*.

The first hypothesis involves the possibility of using OCR systems for character recognition which are specially trained to recognize human handwriting. The main difficulty is the peculiarities of the nineteenth century spelling. However all *mappette* and *sommarioni* available are quite homogeneous from this point of view (they were written by a few people). This facilitates the training phase of the recognition software. Semi-automatic grammar correction is also made difficult by the many linguistic conventions and abbreviations adopted at the time. It makes the prior creation of a special dictionary essential. Another interesting possibility is the new collaborative arrangements that have developed on the net, of which we will cite only the most famous such as Wikipedia and - remarkably similar to our activities - OpenStreetMap.

This would take advantage of the potential offered by academics, students, technicians and enthusiasts who, needing specific data for their activities, can make their work available to all. Or those who simply wish to contribute freely to an ambitious project. The example of the OpenStreetMap (OSM) project started in 2004 with the aim of creating a free map of the world can provide a good reference basis on how to involve users (the organization of events such as Mapping Weekends or Mapping Parties) and for the innovative software tools it makes available for editing maps.

References

A.A. (1982). Atlante storico del territorio marchigiano. Cassa di Risparmio. Ancona.

Bitelli G., Gatta G. (2011). Experiences on georeferencing of maps from the XIX century Gregorian Cadastre of Bologna (Italy). e-Perimetron, Vol. 6, No. 4.

Buonora P. (2004). Digitalizzazione e accesso on-line della cartografia storica. In digital form: www.storiaurbana.it/biennale/Relazioni/b7.buonora.doc

Buonora P. (2009). Digitization, online utilization and preservation of cadastral very large format cartography, e-Perimetron, Vol. 4, No. 3.

Orciani M., Frazzica V., Colosi L., Galletti F. (2005). Analisi GIS integrate multisettoriali nella pianificazione paesaggistica di area vasta della Provincia di Ancona. Mondo GIS, 50, Roma.

Orciani M., Frazzica V., Colosi L., Galletti F. (2007). Gregoriano Cadastre: transformation of old maps into Geographical Information System and their contribution in terms of acquisition, processing and communication of historical data. e-Perimetron, Vol. 2, No. 2.

Orciani M., Frazzica V., Colosi L., Galletti F., Cartaro A. (2007) "From the Gregoriano Cadastre to a three-dimensional model of nineteenth-century landscape through the creation of historical spatial data", 10th AGILE International Conference on Geographic Information Science "The European Information Society: Leading the way with geo-information", Aalborg University, Denmark.

Orciani M., Frazzica V., Colosi L., Galletti F., Cartaro A. (2009). "Gregoriano Cadastre (1818-35) from old maps to a GIS of historical landscape data", EGU General Assembly. Wien.