Vasile Crăciunescu*, Ştefan Constantinescu**, Ionuţ Ovejanu***, Ioan Rus****

Project eHarta: a collaborative initiative to digitally preserve and freely share old cartographic documents in Romania

Keywords: geo-spatial.org; georeferencing; open-source; web services

Summary
Historical maps represent a valuable resource in reconstructing the natural and built environment evolution. In order to integrate such maps with recent geospatial datasets, in GIS environment, some preparatory actions need to be done. This includes scanning, color enhancement, metadata creation, georeferencing and reprojection. From our experience, when it comes to collections with a large number of map sheets, these preparatory actions can take a very long time to accomplish.

With this disadvantages in mind, geo-spatial.org, a collaborative online project aiming to facilitate the sharing of geospatial knowledge and the discovery and publishing of free geographic datasets and maps for the Romanian community, started the eHarta initiative. Within eHarta, Web 2.0 geospatial enabled interfaces were carefully designed, using free and open source software, to allow the community to get involved in old maps publishing. Through these friendly interfaces, the users were able to do basic map preprocessing operations, fill standard metadata forms and enter georeference control points for thousands of old maps. The results were published as georeferenced files (GeoTiff) and standard web services (WMS, WMTS, KML) using a public domain license.

Introduction

There is an enormous amount of old maps and other historical documents in the collections of libraries and archives all over the world (Přidal and Žabička 2008: 10). In the last years, an important number of such data holders started projects to digitally preserve and disseminate the old cartographic information. A good example in this sense is offered by ARCANUM's Digitized Maps of the Habsburg Military Surveys Project (Timár et al. 2010: 273).

The eHarta Project, started by geo-spatial.org – a collaborative online platform aiming to facilitate the sharing of geospatial knowledge and the discovery and publishing of free geographic datasets and maps for the Romanian community – has many similarities with this kind of projects but also some remarkable differences. The most important features of the project are described bellow:

- eHarta aims to offer free online access to many thousand of old map sheets and atlases. All the maps are made available for download as georeferenced files and accessible through a number of web services that fit a broad range of users (e.g. Zoomify tiles or KML files for ordinary users; standard geospatial services like WMS/WMTS for users with advanced GIS skills).
- eHarta has no financial resources available. It is totally depending on community contributions (e.g. metadata creation, data processing, etc.), institutional donations (e.g. Internet bandwidth, free access to map archives, access to high performance scanners and other digital preservation instruments) and free and open source software usage.

* National Meteorological Administration, Bucharest, Romania [vasile.craciunescu@meteoromania.ro]
** Faculty of Geography - University of Bucharest, Romania [stefanc@geo.unibuc.ro]
*** Faculty of Geography - University of Bucharest, Romania
**** Faculty of Geography - University Babes Bolyai, Cluj Napoca, Romania [nelurus@geografie.ubbcluj.ro]

Community involvement

Community powered projects are no longer a novelty. Online initiatives like Wikipedia or OpenStreetMap proved that is possible to achieve incredible results, both in quality and quantity, only with volunteers, open data/knowledge model and an adequate set of web tools. eHarta employs the same principles and methods like other online collaborative projects:

- Friendly web interfaces for old map related operations (e.g. metadata creation, add control points for georeference, image enhancements, etc.);
- Feedback channels;
- Open data license for the final products;
- Public recognition of each member contribution;
- User content control mechanisms.

A couple of months after project start, more than 200 unique accounts were registered into the system. We consider the number significant when talking about a niche community like Romanian historical cartography passionate. More important, thousand of map sheets were processed in just two-three days after release.

System architecture

eHarta online system was designed following an distributed architecture, entirely based on free and open source software. The content is managed by Textpattern, a powerful and flexible open source content management (CMS) application. For supplementary, specific functionality, custom modules were built. Other free applications are providing server-side functionality: MySQL (relational database management system), PHP, Python, Java (server-side scripting languages), Apache (webserver), Tomcat (servlet container), phpMyAdmin, phpPgAdmin (web clients for database management).

For geospatial data management, top open source applications were also integrated in the website: PostGIS (vector geospatial data storage), GeoNetwork (geospatial data catalog and metadata editor), GeoServer (standard geospatial server for serving data via WMS), GeoWebCache (Java based WMTS/TMS server, with pluggable caching mechanisms and rendering backends), OpenLayers and GeoExt (client webmapping application), GDAL (data processing).

The information flow between the various server side applications and the front-end graphical interface is determined by the interaction with the portal users and their requests (Figure 1).
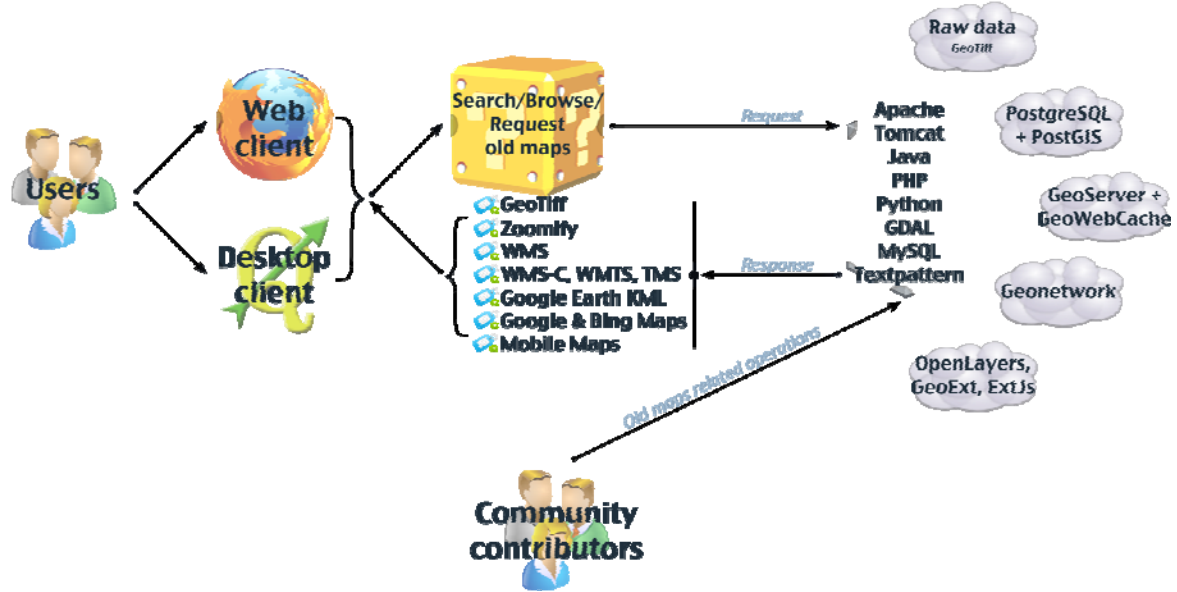
Figure 1: eHarta system architecture.

Case study: 1: 20,000 Romanian maps under 'Lambert-Cholesky' (1916-1959) projection system

As well as many other European countries, Romania has an important archive of historical cartographic documents. In the beginning of the eHarta project we decided to involve the community in the publication of the 1: 20,000 Romanian map series under 'Lambert-Cholesky' projection system. The maps were created after the World War I, for military proposes (especially artillery). The basic map, called „Plan Director de Tragere" was drafted under 1:20000 scale in 2118 drawings, covering all the Romanian territory. Under graphical aspect, such map sheet had a 75 cm width (the equivalent of 15 km of land), respectively 50 cm height (the equivalent of 10 km of land). Usually, at the upper part of the map, frequently to the left side, less frequently to the right side, the drawings nomenclature appeared, made following the principle: the first two letter meant the columns number divided by 15 and the last two digits represented the row number. So, the drawing whose southwest corner had the Cartesian coordinate of 450,000 m Eastings and 20,000 m Northings would have received the codification 3020 (Rus et al. 2010: 106). More than 1800 map sheets from this collection were identified in archives belonging to faculties of Geography in Bucharest and Cluj Napoca and the National Museum of Romanian History. The maps were scanned at 300 dpi and saved as uncompressed Tiff files. Zoomify versions were later derived for metadata creation and identification of control points for georeference. Several web user interfaces were created for community actions.

Metadata creation

First step was to add metadata information for each map sheet. After login into the system, the users were able to select and reserve a number of map sheets through an ExtJs grid interface (Figure 2).
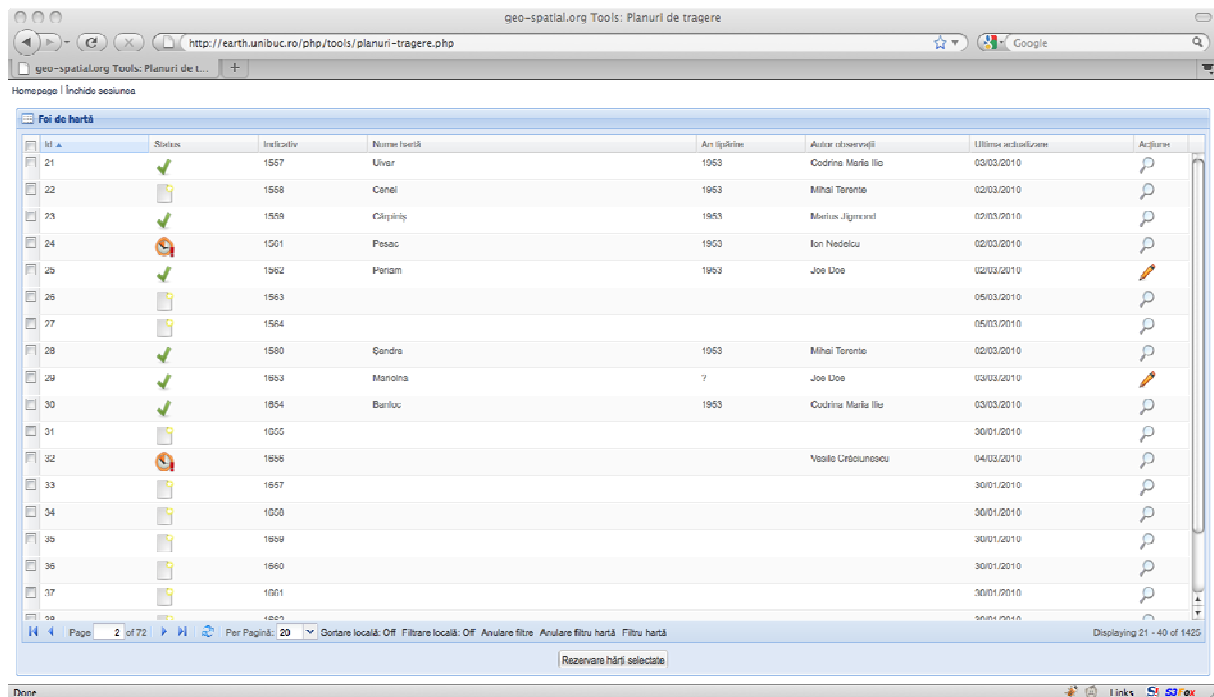
Figure 2: Map selection and reservation interface.

The reservation system was put in place to avoid work duplication. The interface also offers tools to sort and filter the map sheets using tabular and/or spatial criteria. The metadata status of each map sheet (not started, in progress, finished) is displayed with appropriate graphical symbols. Once a map sheet is reserved, the user is able to advance to a new screen where the actual meta-data information can be filled (Figure 3).
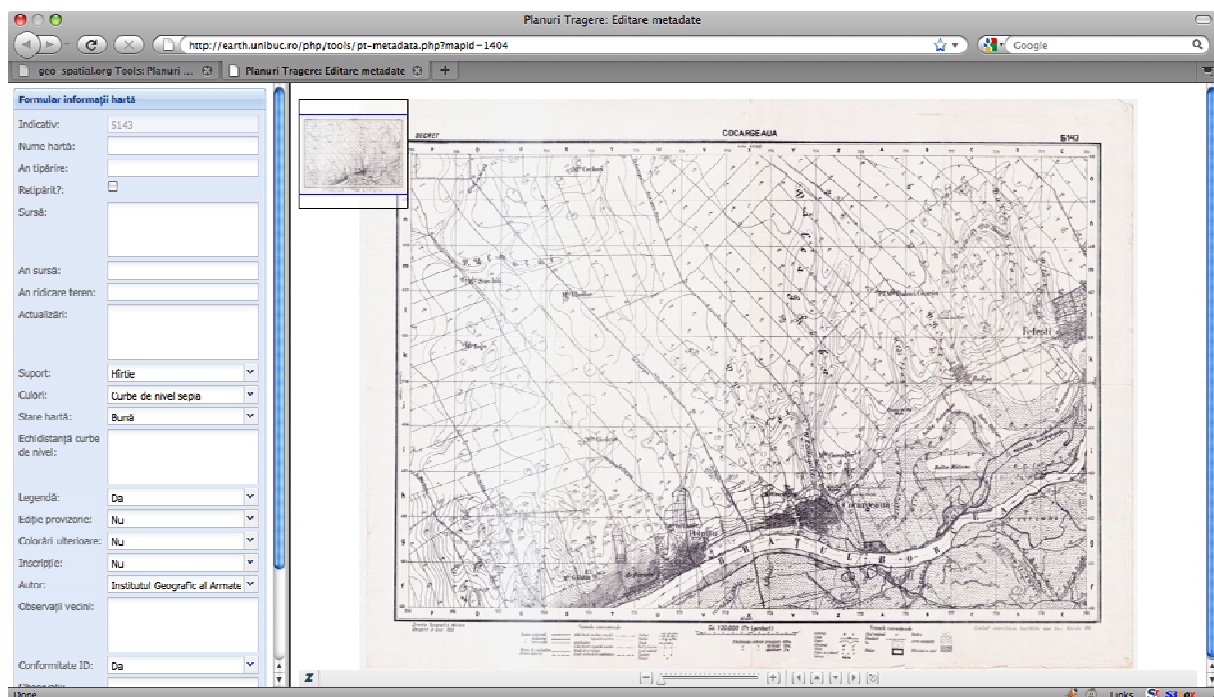

Figure 3: Metadata editor.

The metadata editor interface is composed from two functional areas:

- The actual metadata form on the left side, containing fields for information like map sheet name, year of production, author, map quality, data sources, survey period, contour lines interval, etc.;
- A Zoomify based interface for map preview and navigation (zoom and pan). The information collected in this manner was later compiled in XML files, using ISO 19139 template, and imported to GeoNetwork.

Georeferencing of scanned maps

Represent the second major step carried out with help from the community. This time the users were asked to use the same reservation system to further mark the four corners of each map. For this, a simple webmapping interface, based on the well know OpenLayers library, was constructed. Fortunately, OpenLayers is capable to read and display Zoomify tiles, allowing us to re-use the same tiles used for metadata purpose. The webmapping application (Figure 4) includes controls for two types of actions:

- Interaction with the map (pan, zoom in, zoom out, zoom to maximum extent, previous zoom, next zoom, overview map);
- Handle the control points (mark new point, modify position of an existing point, delete point).
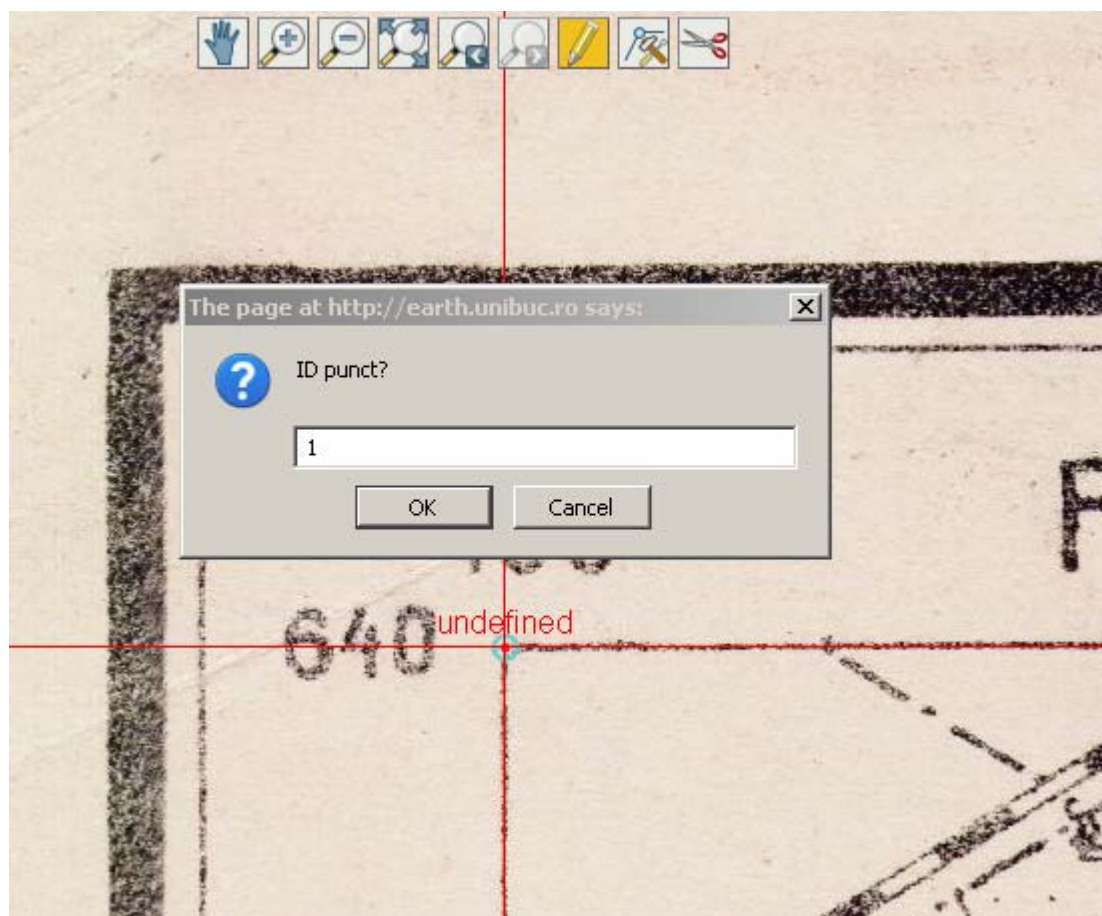

Figure 4: Control point editor.

When marking a new point, the user is prompted to also add a unique ID using the following simple schema:

- 1 for Northwest corner;
- 2 for Northeast corner;
- 3 for Southeast corner;
- 4 for Southwest corner.

A small PHP REST service was setup to save/update the control points (IDs and pixel coordinates) in a MySQL database. GeoJSON format was used to transport the data from map to database and back. Again, the response from the community was great, more than 1800 map sheets being processed in less than two days. A few maps required special attention because one or more of the corners were physically missing.

The actual georeferencing, and some additional processing steps, were done automatically using GDAL scripts. The main commands are presented bellow:

1. Georeferencing: gdal_translate map_sheet_name.tif -gcp PIXEL_1_X PIXEL_1_Y COORD_1_X COORD_1_Y -gcp PIXEL_2_X PIXEL_2_Y COORD_2_X COORD_2_Y -gcp PIXEL_3_X PIXEL_3_Y COORD_3_X COORD_3_Y -gcp PIXEL_4_X PIXEL_4_Y COORD_4_X COORD_4_Y -of Gtiff -a_srs EPSG:31700 map_sheet_name_georef.tif

2. Clip map collar: gdalwarp -r lanczos -cutline map_sheet_mask.shp -crop_to_cutline -dstnodata 255 map_sheet_name_georef.tif map_sheet_name_temp.tif

3. Create the final file with transparency, JPEG compression and tile TIFF encoding: gdal_translate -b 1 -b 2 -b 3 -a_nodata 255 -co COMPRESS=JPEG -co TILED=YES map_sheet_name_temp.tif map_sheet_name_clip.tif

4. Create overviews for fast display: gdaladdo -r gauss --config COMPRESS_OVERVIEW JPEG map_sheet_name_clip.tif 2 4 8 16 32 64 128

The runtime exceeded ten days on a high end machine due to the Lanczos resampling method used at step two, which produce a high quality output but is quite slow in computation. The script processed approximately 500GB of uncompressed Tiffs and produced and output of around 120GB. The size reduction was due to clipping and JPEG compression used inside final GeoTiff files.

Web publication

The final step of this first eHarta project was to publish the maps as raw files and standard web service. The step was more challenging than we expected due to the relative large size of the data pack. One of the biggest problems was caused by GeoServer inability to serve GeoTiff files larger than 2GB. The use of other special compressed file formats like ECW (ERDAS Compress Wavelets) was not an option due to license restrictions. Finally we opted to use a combination of GeoServer, GDAL and two GeoServer plugins: ImageMosaic and Image Pyramid. The steps are presented bellow:

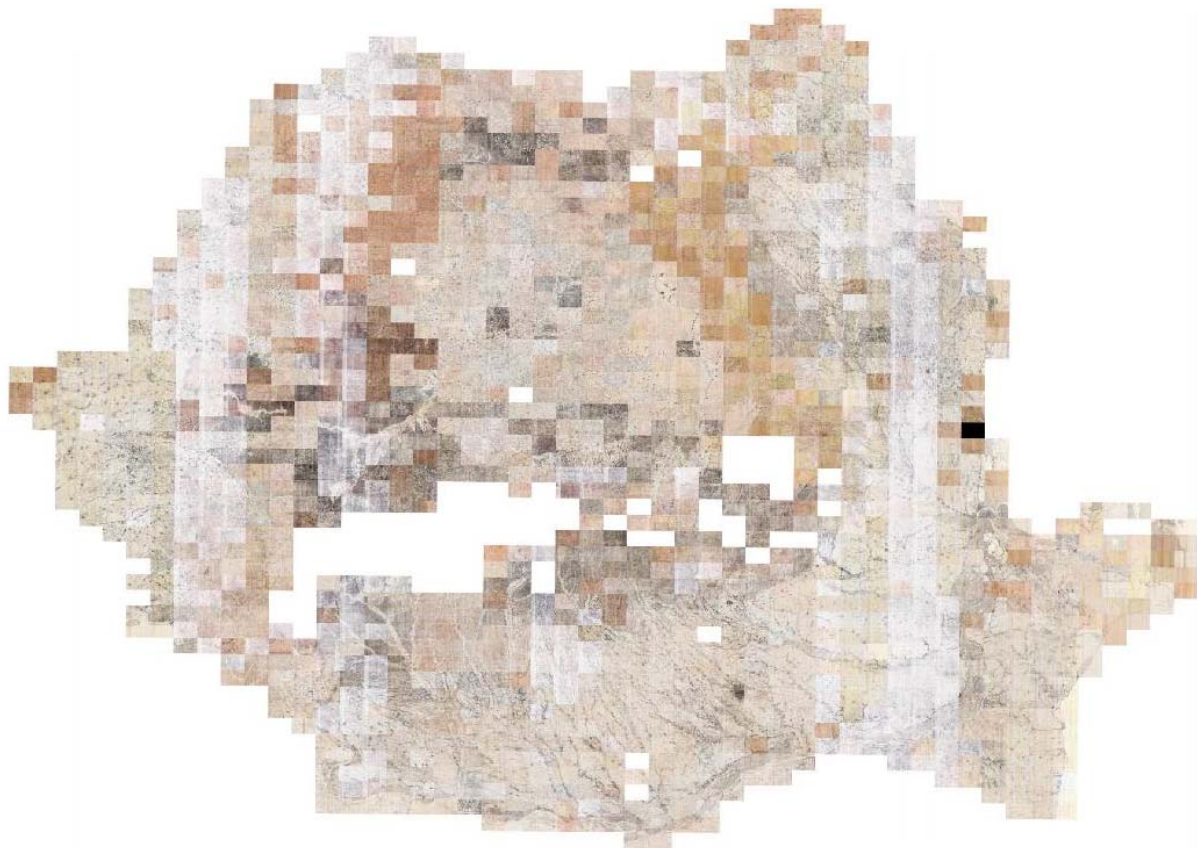1. Create a virtual mosaic of 441,671 by 311,892 pixels (1.6 m cellsize): gdalbuildvrt merge.vrt *.tif (Figure 5);

Figure 5: The resulted map mosaic.

2. Build tiled pyramid levels of the mosaic: gdal_retile.py -levels 8 -ps 2048 2048 -r bilinear -v -co "TILED=YES" -co "BLOCKXSIZE=256" -co COMPRESS=JPEG -s_srs EPSG:31700 -targetDir pt merge.vrt

3. Register each level in GeoServer using ImageMosaic plugin (ImageMosaic allows the creation of a mosaic from a number of georeferenced rasters).

4. Aggregate the ImageMosaic levels using ImagePyramid plugin. After this step, the WMS service is ready for production.

5. Create cached tiles of WMS service, for three coordinate systems (EPSG:4326 for Google/Yahoo/Bing/OSM Maps integration; EPSG:900913 for Google Earth integration; EPSG:31700 official Romanian coordinate system), using GeoWebCache. The cache is served through a number of standards: WMS, WMS-C, WMTS, TMS, KML, Bing Maps, Mobile GMaps.

The processing time for these final steps was around six-seven days. Finally, the users are able to freely access the georeferenced 1: 20 000 scale Lambert-Cholesky maps as plain files or through standard web service within web clients or thick desktop clients. Figure 6 presents an example of map access through the popular Google Earth application.
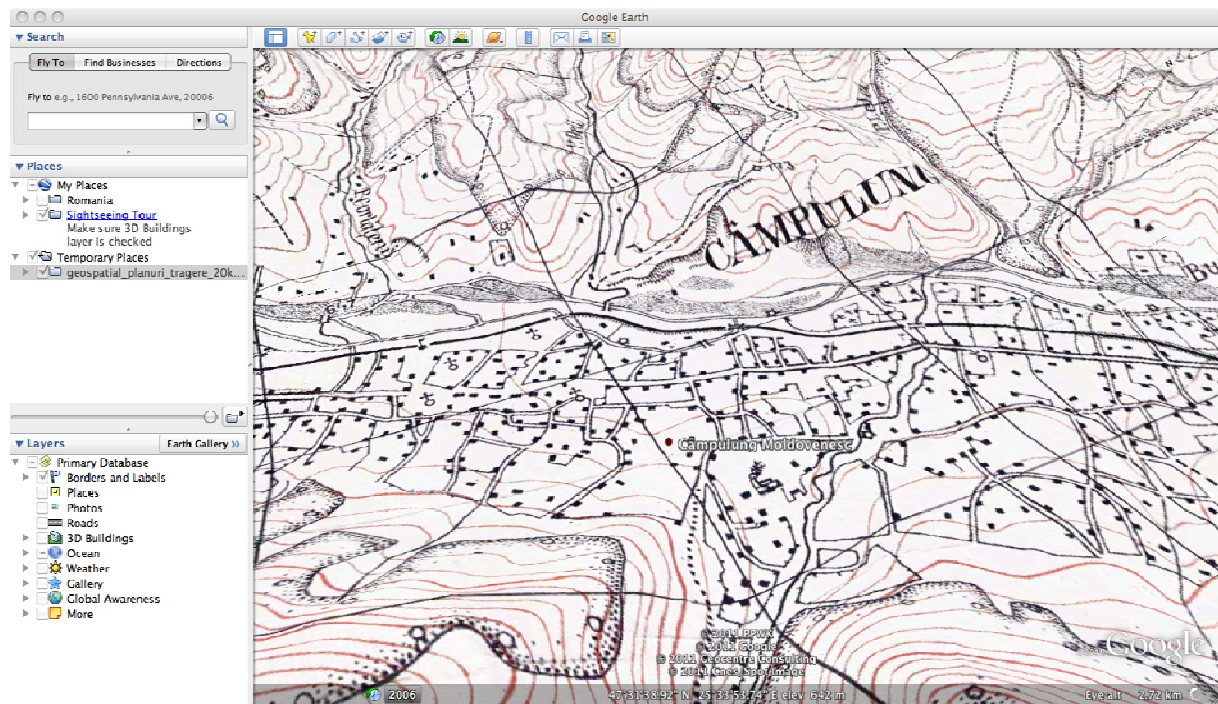
Figure 6: Example: map access through Google Earth.

Additionally, a webmapping application, which integrates all the old maps published within eHarta project, was implemented on geo-spatial.org. The users can use the left menu to adjust the map content by selecting the visible layers or by adjusting the layers opacity. The top toolbar contains useful instruments for map navigation (Figure 7).
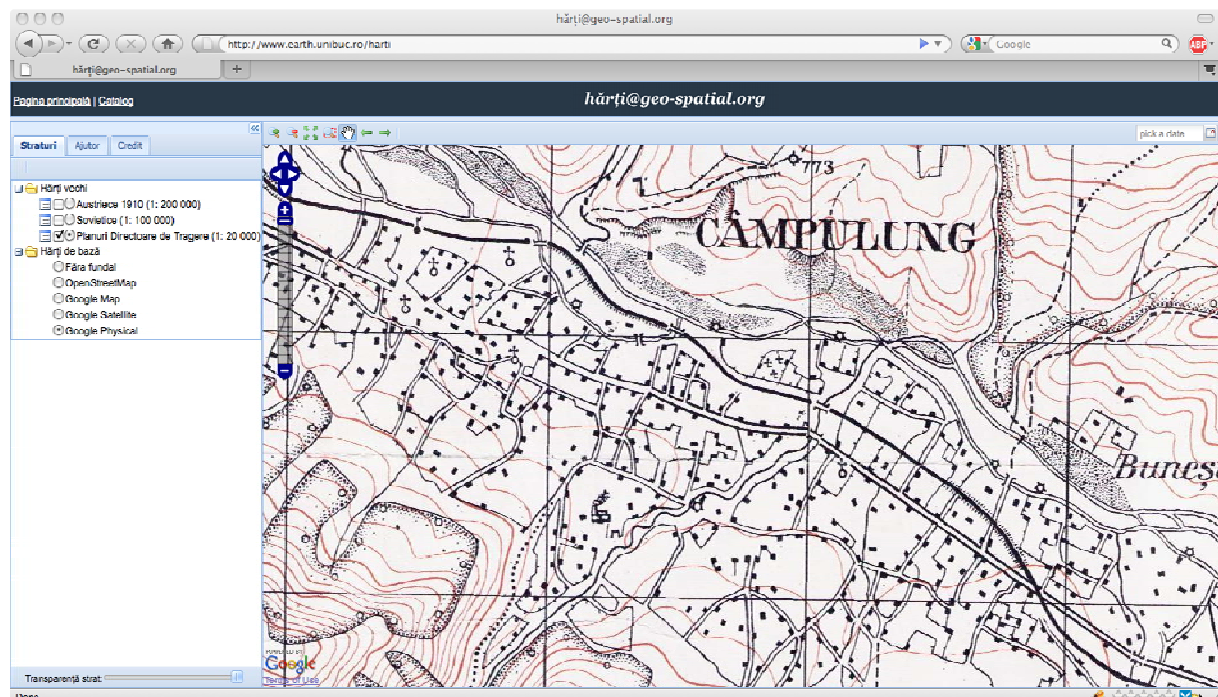

Figure 7: Example: map access through geo-spatial.org webmapping application.

Conclusions

The use of georeferenced historical cartographic documents opened new perspectives in understanding how actual processes are influenced by decisions taken in the past. Maps series like the Romanian 1:20,000 'Lambert-Cholesky' can be successfully georeferenced and included in such applications. The role of community, empowered with easy to use web services, is essential when the volume of data is high and the available resources are low. The domain of free and open source geospatial applications is growing fast and provides reliable solutions for all the stages of georeferenced maps management.

Acknowledgements

The authors express their full gratitude to the other 206 people that contributed so far to the success of the eHarta Project.

References

Přidal P. and Žabička P. (2008). Tiles as an approach to on-line publishing of scanned old maps, vedute and other historical documents, *e-Perimetron*, Vol. 3, No. 1. ISSN 1790-3769, p. 10-21; www.e-perimetron.org/Vol_3_1/Pridal_Zabicka.pdf

Timár G., Biszak S., Székely B., Molnár G. (2010). Digitized Maps of the Habsburg Military Surveys – Overview of the Project of ARCANUM Ltd. (Hungary), *Preservation in Digital Cartography - Archiving Aspects*, ISBN 978-3-642-12732-8, p. 273 – 283;

Rus I., Balint C., Crăciunescu V., Constantinescu S., Ovejanu I., Bartos-Elekes Zs. (2010), Automated georeference of the 1: 20,000 Romanian maps under Lambert-Cholesky (1916-1959) projection system, *Acta Geodaetica et Geophysica Hungarica*, p. 105-111

GeoServer Documentation: http://docs.geoserver.org/

GeoWebCache Documentation: http://geowebcache.org/docs/current/