

Eric Grosso*

SEREDONA: A web service platform to integrate historical vector data into current reference frame

Keywords: web platform; OGC services; data integration; historical geographic data; old maps; georeferencing

Summary

This article has the main objective to introduce the web platform SEREDONA. This platform aims to suit the needs of users who want to georeference historical data, in providing tools to integrate easily historical geographic vector data into current georeferenced frameworks. SEREDONA has been built around the main idea that the georeferencing process of historical data is maybe not necessarily easy for people dealing with historical data such as historians, archaeologists or ecologists. Therefore, the choice is done to provide tools in a pedagogic way through a web platform based on a architecture of OGC (Open Geospatial Consortium) standard services such as Web Map Service, Web Feature Service and Web Processing Service.

Introduction

Over the last few decades, historical geographic data has been digitised for archiving and preservation purposes. This phenomenon has consequently contributed to a wider diffusion of this data. Although it is possible to use data in its actual state or digitalised, users realize that this data becomes more useful if its content is directly integrated into a recent frame of reference, thus enabling that data to be manipulated using a Geographical Information System (GIS) (see e.g. Sanderson and Brown 2007). Indeed, the process of integration of historical data into a recent frame of reference increases the possibility of analysis with current data. An integration process is thus essential. Currently, the main method used to integrate this data consists of georeferencing the digitalized image onto the ancient document.

Nevertheless, some users such as archaeologists, historians, ecologists or more generally research scientists, begin to vectorise this data in order to obtain better analysis and to improve the study of geographical phenomenon. Currently, several projects use this solution, e.g. the ALPAGE project (Noizet and Grosso 2011) or a study of the French forest evolution (Dupouey *et al.* 2007). Moreover, some new solutions about the georeferencing of vector data begin to be proposed in the International Cartographic Association commission on "Digital Technologies in Cartographic Heritage" (Cajthaml 2010 ; Grosso 2010). These solutions of historical data integration should allow data enrichment (Hampe and Sester 2002), enable the detection and correction of inconsistencies (Egenhofer *et al.* 1994) and also the evaluation of data quality. Unfortunately, the tools to integrate historical geographic vector data are not yet easily available for users.

In this context, this paper aims to introduce a solution to solve the problem of historical geographic vector data integration by explaining the building of SEREDONA, a data integration web platform, for which the French acronym is for "SErvice de REcalage de DONnées Anciennes"

* PhD Student, Institut Géographique National, Laboratoire COGIT, 73, avenue de Paris, 94360 Saint-Mandé – Université de La Rochelle, Laboratoire L3i [eric.grosso@ign.fr]

and which can be translated as follows: "Web service platform for the adjustment of historical data".

This platform aims to suit the needs of users who want to georeference historical data, in providing tools to integrate easily historical geographic vector data into current georeferenced frameworks. SEREDONA is built around the main idea that the georeferencing process of historical data is perhaps not always easy for people dealing with historical data such as historians, archaeologists or ecologists. Therefore, the choice is made to provide tools in a pedagogic way and through a web platform to allow an easy access for users. Finally, to ensure the interoperability and the compliance with the technical solution chosen in the European INSPIRE directive, the platform is based on a architecture of OGC (Open Geospatial Consortium) standard services such as Web Map Service, Web Feature Service and Web Processing Service.

After explaining the motivations for this work, the functionalities, the architecture and the pedagogic content used in the platform will be described.

Motivations

Thanks to the digitalisation of historical data, many people, including ecologists, glaciologists, archaeologists, historians and more generally research scientists, began to access data easily and thus were able to study these data. To improve their analysis, they have to georeference the data into a current reference framework in order for example to compare data at two different times. To help them in these tasks of georeferencing and analysis, software solutions have been developed. Solutions for georeferencing were first provided in desktop GIS software (both proprietary and open source GIS software). Then the Old Maps Online project helped users with online publishing and georeferencing of digitised historical maps (Pridal and Zabicka 2008) based on the observation that the ability to use old maps is critically under-utilised in the current Internet environment. Finally, the project MapAnalyst (Jenny *et al.* 2007) provided a tool to help users visualise easily planimetric accuracy of historical maps.

More recently, another approach based on the vectorisation of historical data has been increasingly used. This approach allows users to go one step further in their analysis. Indeed, vector data enables users to apply some spatial analysis processes, e.g. to compute the main orientations of the building walls in a town or to detect some patterns in a urban structure. The detailed reasons of this approach are explained in Grosso (2010). To be able to compare these historical vector data with current data, users have to georeference their data, as it is done for raster data. Indeed, the vectorisation is generally processed after the digitalisation to keep the best raster quality and also to allow all the possibilities of georeferencing (using different geometric transformations). Unfortunately, except for the uncompleted tool of spatial adjustment provided in ArcGIS, a proprietary GIS software published by ESRI, the georeferencing of vector data is not currently available for users.

Based on this last assessment, we decided first to develop a georeferencing tool to deal with vector data as a standalone application. This tool has been successfully tested in three different projects involving historical vector data. Then, to allow users to access these integration processes and thus to solve the problem of the quasi non-diffusion of these processes, we decided to develop this application as a web solution. The OGC standard web services has been chosen to assure the interoperability of the platform in terms of web solution but also in terms of the compliance with the technical solution selected in the European INSPIRE directive. Finally, thanks to our work in collaboration with historians and archaeologists during three different projects, it appeared that it

was not necessarily easy for them to deal with the georeferencing process (both for raster and vector data), e.g. to choose an adapted geometric transformation in a specific context or to evaluate the results. Therefore, we thought that it was important to design our platform from a pedagogic point of view, in order to fit the needs of users such as historians or ecologists and to allow them an easy access.

Here we describe the functionalities and the architecture of the web platform SEREDONA that was developed, to understand its design and its pedagogic content.

Functionalities and architecture of the web platform SEREDONA

From a global point of view, SEREDONA is built around seven main modules as illustrated in Figure 1: a users management module, a data management module, a metadata management module, a data validation module, a visualisation portal, a covisualisation portal and a georeferencing portal. Data (users information, metadata) and geographic data are stored in database management systems. Nevertheless, geographic data can only be accessed through OGC web services of diffusion, i.e. through Web Map Service (WMS) and Web Feature Service (WFS). Every process is computed based on OGC Web Processing Service. Finally, in order to make the platform accessible to foreigners, it has been internationalised. French and English languages are provided and the language displayed is the one which fits the “locale” parameter or the country code of the browser.

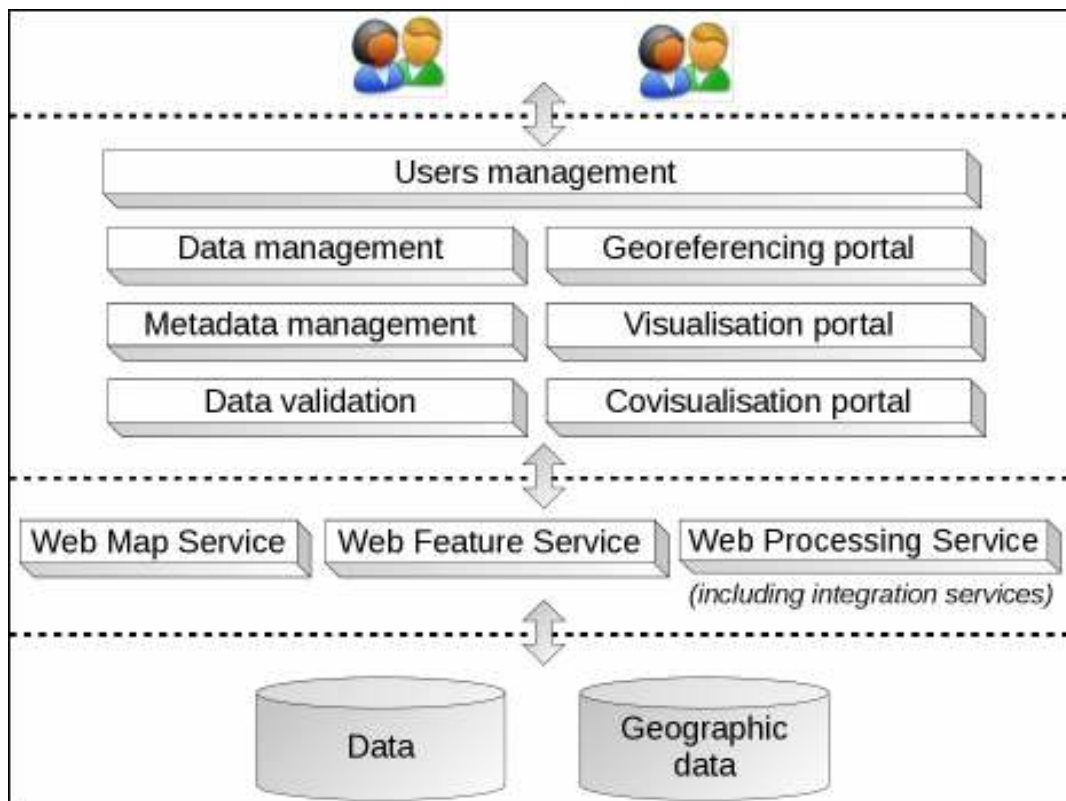


Figure 1: Simplified architecture of the platform SEREDONA.

We now describe the different modules of SEREDONA. This description is predominately based on a pedagogic point of view rather than on a technical one. Indeed, the technical aspects are already detailed in Grosso (2011).

Users management

To access SEREDONA, users have to create an account (registration). They can access their account information, view and modify their personal details, but also save their preferences, or stop a task in progress and get on with it later. But mainly the users rights serve as a basis for defining the rights of users spatial data. Currently, a collaborative way of working is still under development. Indeed, a classical example would be that several users need to work at the same time on the same project and therefore need to access the same data. Rather than working separately, the aim is to share data, information and results between a group of users.

Online data management

In order to display or to process spatial data online, users have to manipulate their data through the OGC web services of diffusion, i.e. WMS and WFS, or data in KML (Keyhole Markup Language) format, which is an OGC data standard. Users such as historians, archaeologists or ecologists manipulate mainly geographic data in a ESRI Shapefile format. Thereby, being rarely non specialists in GIS, let-alone in spatial data conversion (for KML) or spatial data server installation (for WMS and WFS), they encounter difficulties in displaying their data online.

To solve this problem, SEREDONA is designed to be able to upload a compressed archive (“zip” format) which contains one or several ESRI Shapefiles provided by a user, and to load dynamically these data into the SEREDONA spatial data server which is GeoServer (<http://geoserver.org>) – more precisely, the users data are loaded into the database management system (DBMS) PostgreSQL/PostGIS (<http://postgis.refrations.net/>) and a dynamic link between the DBMS and the SEREDONA spatial data server is created. To be loaded in the spatial data server, a projection must be associated to the data. Unfortunately, this is usually not the case with the historical vector data. To enable the data loading, a fake projection file is generated using a WGS84 projection.

At the end of this step, users can access their data through WMS and WFS. Nevertheless, this access is restricted to the user who owns these data, and the data displaying or processing can be done only from the SEREDONA servers. This restriction enables users to keep a full control of their data, control which is often requested by users in a context of web applications. Thanks to a collaborative management, users could share their data by giving access to their data (or a part of it) on a read-only basis to another user.

Metadata management

The metadata management module enables to add metadata about users data. The metadata framework used in SEREDONA is based on the ISO-19115 metadata standard (ISO 2003). This standard consists of more than 400 metadata elements. To fit with geographic historical data, we have defined a metadata framework which is both a subset of the ISO-19115 metadata standard – keeping only the metadata elements which are in relation with historical data – and an extension

of this standard (Grosso 2011a). This extension can be a specialisation of existing metadata elements – e.g. the description of the digitisation process, the spatial transformation used to georeference data and its parameters – or a new metadata element – e.g. the description of the historical data quality and the sources which allow determining it. This metadata framework is applied on three different data levels: on a “Data series” level – e.g. Cassini maps –, on a “Dataset” level – e.g. one sheet of Cassini maps –, and on a “Theme” level – e.g. the windmills in Cassini maps. The use of such a metadata framework has three objectives: 1. to help users discover, explore and exploit historical data, 2. to capitalise and share experts knowledge, and 3. to share knowledge about the georeferencing process.

Online data visualisation and covisualisation

The objective of the data visualisation module is to allow users to easily display their data on the web by simply uploading them (in a ESRI shapefile format) into the SEREDONA platform without having to install a server or to convert data. Users can access their data through webpages which are automatically generated. The data displayed is based on the OpenLayers library.

To improve the simple online data visualisation, an online data covisualisation module has been created. Thus users are able to display their data with data produced by institutions or national mapping agencies as illustrated in Figure 2, or when the collaborative process will be available on the platform, with data provided by other users. The historical data needs to be georeferenced to enable their covisualisation with data coming from a current reference framework.

The covisualisation module means that users can compare the results of a georeferencing process using two (or several) different spatial transformations. This way, users can choose the most adapted one and analyse their data in a current context.



Figure 2: Covisualisation of georeferenced Cassini vector data and the data.

Data validation

Some errors, both geometric and topological, could result in the process of data vectorisation such as illustrated in Figure 3. This is particularly true about the vectorisation of historical data which

is realised mainly by non geographical information experts. Due to these errors users have difficulties applying analysis processes on historical data.

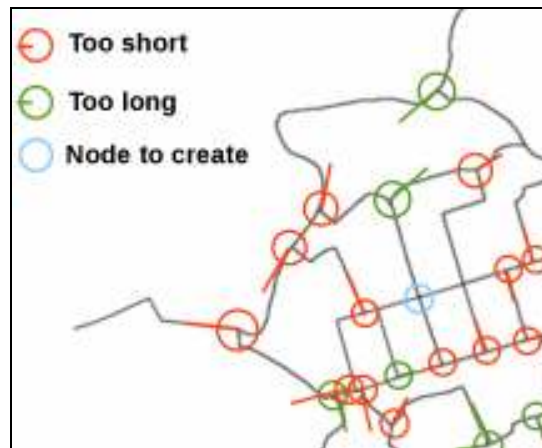


Figure 3: An example of map of topological errors.

Therefore, a validation process appears crucial, especially in the context of a services based platform. Indeed, if a geoprocessing service encounters a problem in a context of service chaining, it usually implies a global error for the entire process (Grosso et al., 2009). The data validation module allows users to check if their geographic data is valid or not, considering the correctness of geometries, the homogeneity of geometric type (point, linestring, etc.), the topological errors, etc.

This data validation module is implemented through a Web Processing Service and based on the JTS – JTS Topology Suite – library (<http://www.vividsolutions.com/jts>). JTS enables us to check if a set of geographic data is valid or not at geometric and topological levels, based on the fact that JTS conforms to the Simple Features Specification for SQL (OGC standard). The data validation module creates dynamically web pages with the results of this validation. These results are presented with the use of statistics and maps of validation errors.

Online georeferencing process

The georeferencing process is the key element of SEREDONA. After having uploaded data into the platform, users have a choice between four different solutions of georeferencing which appear helpful for users:

- georeferencing historical data into a current reference frame directly, which is the most common solution – this can be called an absolute direct georeferencing process –,
- georeferencing historical data into a current reference frame indirectly. This solution allows users to use intermediate data to georeference data into a current reference frame, due to the fact that it is impossible to find enough ground control points to process the georeferencing – this can be called an absolute indirect georeferencing process –,
- georeferencing historical data into other historical data, which can be useful to study the difference between two maps at two different periods, – this can be called a relative georeferencing process –,
- georeferencing current reference data into historical data, which can be useful for example to improve the vectorisation of historical data. Indeed, if computer scientists want to vec-

torise the symbols of windmills in a map, they can use the fact that windmills are very close to the hydrographic network. The georeferencing of the current hydrographic network into the old map to vectorise provides a context where the vectorisation process can be applied – this can be called an inverse georeferencing process –.

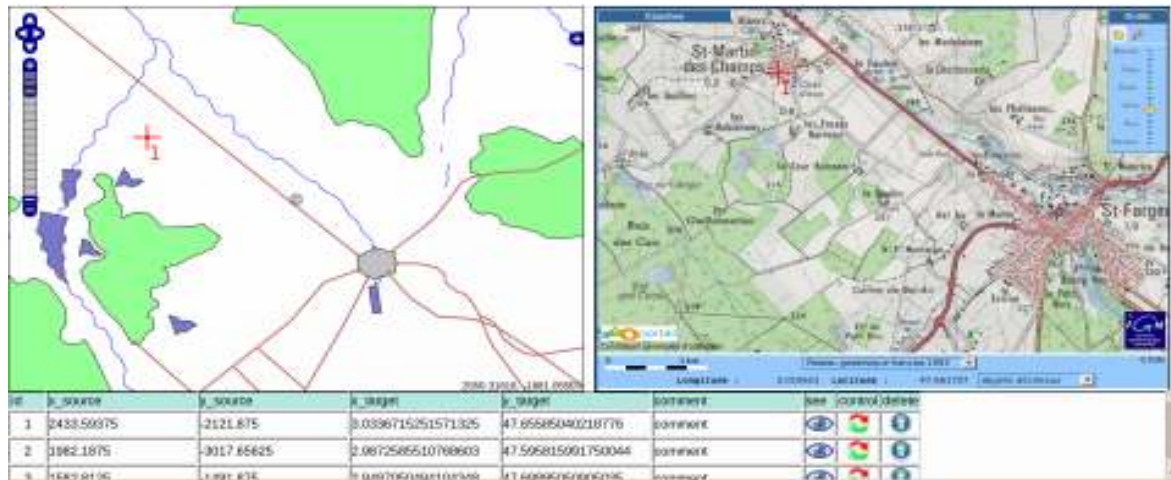


Figure 4: Web interface to add the ground control points.
Left: Cassini's maps, Right: current geographic reference framework.

Once users have selected one of these options, two map displayers are shown, e.g. in the absolute direct georeferencing process, the first one contains user data and the second one contains institutional data as illustrated in Figure 4. The institutional data here are the data provided by the national mapping agency through its national geoportal which contains both images of vector database and maps: SEREDONA uses the French geoportal to georeference French historical data, otherwise it uses data provided through Google maps.

Then, users have to select GCP (the web interface allows to select GCP only, or to select GCP and other points which will be used to control independently the quality of the transformation) and launch the computation of the georeferencing process. Different levels of users have to be considered for this task, from beginners to experts. If users are beginners, the platform computes several possible spatial transformations (experts can decide to do the computation with only the transformation(s) they choose): affine transformations, Helmert transformations (four or seven parameters), first or higher order polynomial transformations and thin-plate spline method. A table with complete information about transformations is displayed to help the beginners to choose the transformations adapted to their needs: map of residual errors and transformation grid (see Figure 5.A) ; residual errors on each GCP, Root Mean Square (RMS) and RMS associated to cross-validation (see Figure 5.B) ; map of interpolated errors (see Figure 5.C) ; histogram of the residual error frequencies (see Figure 5.D). As described, a cross-validation technique is used to evaluate the results. Indeed, to test the geometric accuracy of a result, one of the best ways is to test this accuracy with a set of independent GCP. The problem is that it is not always possible to have this independent set of points, mainly in the georeferencing of historical data where it is sometimes difficult to find enough GCP to be able to compute a transformation. Therefore, the geometric accuracy is evaluated thanks to a cross-validation technique (McGwire 1996).

Based on this evaluation, users can decide to remove some control points and to launch once again the georeferencing process. At the end, the new georeferenced data can be saved through the online data management module.

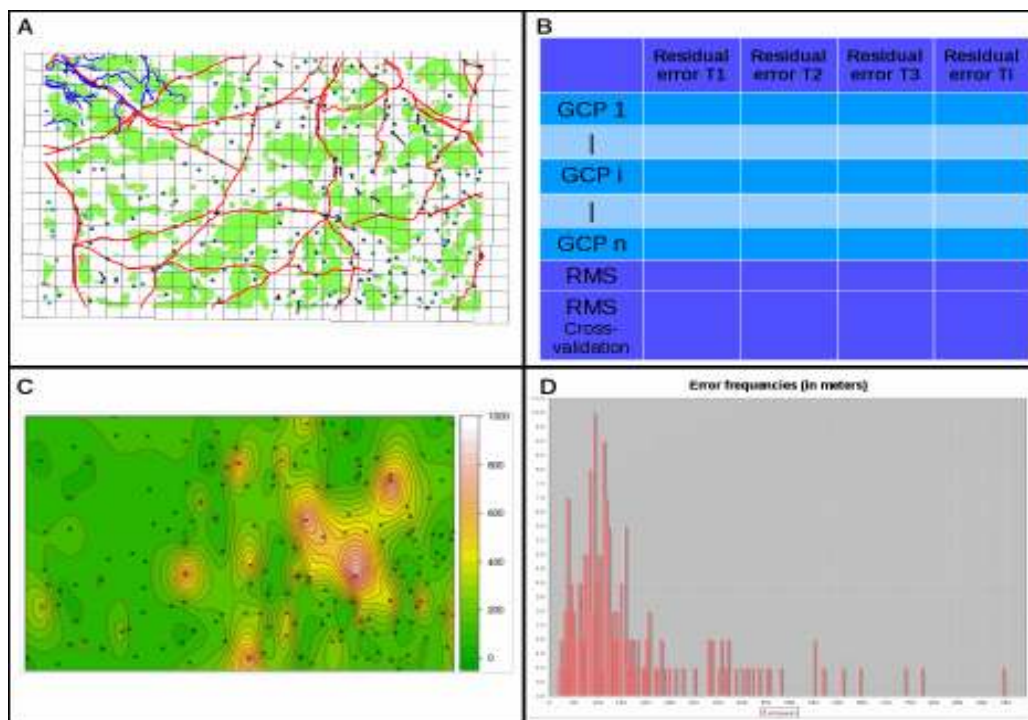


Figure 5: A. Map of residual errors ; B. Table of comparison between the residual errors of different transformations ; C. Map of interpolated errors ; D. Histogram of the residual error frequencies.

If classical georeferencing tools are used the georeferencing process will have to be tested with different mathematical transformations separately and then the results between the different strategies compared manually. The advantage of this module is that it allows users to compare with ease the results of different georeferencing transformations simultaneously, both numerically using the table of comparison between the residual errors of the different transformations, and visually using the covisualisation data module.

Platform based on OGC web services

The choice of using the OGC services has been made as it provides a platform compliant with the technical solution chosen in the European INSPIRE directive. Therefore, all the data visualisation is provided through Web Map Service, the data diffusion through Web Feature Service and all the processes through Web Processing Service (developed on top of the 52°North WPS platform (<http://52north.org/communities/geoprocessing/>)). The aim is to have a platform which fits a full distributed environment and thereby allows users to add their own geographic data diffusion servers – e.g. GeoServer or MapServer – and/or their own WPS servers to apply processes which are not necessarily available in SEREDONA. Moreover, the use of Open Source components enables the reuse of developed components in a more global spatial data infrastructure or separately in other contexts.

Conclusion

This paper introduced the web platform SEREDONA which aims to provide a solution to the problem of historical geographic vector data integration into current reference frameworks.

The functionalities of this platform have been presented through the description of the different modules: a users management module, a data management module, a metadata management module, a data validation module, a visualisation portal, a covisualisation portal and a georeferencing portal. Each module has been developed as much as possible in respect of the needs of users and to be pedagogic. In term of architecture, this platform is based on OGC standard Web services to allow a compliance with other Spatial Data Infrastructures based on spatial standards, with the idea that each module can be reused in another infrastructure.

The SEREDONA web platform is still under development and should be complete at the end of April 2011, and we intend to add more functions in the near future.

References

- Cajthaml J. (2010). Digital technologies in analysis of the Müller's map of Bohemia, 5th ICA International Workshop on Digital Approaches in Cartographic Heritage, 22–24 February, Vienna (Austria).
- Dupouey, J.-L., J. Bachacou, R. Cosserat, S. Aberdam, D. Vallauri, G. Chappart and M.-A. Corvisier De Villèle (2007). Vers la réalisation d'une carte géoréférencée des forêts anciennes de France. *Revue du Comité Français de Cartographie (CFC)*, Vol. 191, 85-98.
- Egenhofer M., E. Clementini and P. Di Felice (1994). Evaluating inconsistencies among multiple representations. In: *The Sixth International Symposium on Spatial Data Handling (SDH)*, Edinburgh, Scotland, 901–920.
- Grosso E. (2011). SEREDONA: a web platform to integrate historical geographic data into current georeferenced frameworks. *Proceedings of the 25th Conference of the International Cartographic Association*, 3-8 July, Paris (France), *Advances in Cartography and GIS Science*, Selection from ICC Paris, Vol.2, pp. 117-128, Springer.
- Grosso E. (2011a). Proposition pour une gestion unifiée des données anciennes. *Revue Internationale de Géomatique / European Journal of GIS and Spatial Analysis*, Hermes Lavoisier (in press), 20 pages.
- Grosso E. (2010). Integration of historical geographic data into current georeferenced frameworks: A user-centred approach. In: *e-Perimtron*, Vol. 5 (3), ISSN: 1790-3769, 107-117.
- Grosso E., A. Bouju and S. Mustière (2009). Data Integration GeoService: A First Proposed Approach Using Historical Geographic Data. In J. D. Carswell, A. S. Fotheringham, and G. Mcardle (eds), *9th International Symposium on Web and Wireless Geographical Information Systems (fully refereed proceedings)*, *Lecture Notes in Computer Science (LNCS)*, Springer-Verlag, Berlin, Heidelberg, 103-119.
- Hampe M. and M. Sester (2002). Real-time integration and generalization of spatial data for mobile applications. In: *Geowissenschaftliche Mitteilungen, Maps and the Internet*, Wien, pp.167–175.
- ISO (2003). ISO-TC211 19115 Geographic Information, Metadata. International Standard.
- Jenny, B., W. Adrian and H. Lorenz (2007). Visualising the Planimetric Accuracy of Historical Maps with MapAnalyst. *Cartographica*, 42-1, 89-94.
- McGwire, K.C. (1996). Cross-validated assessment of geometric accuracy. In: *Photogrammetric engineering and remote sensing*, vol. 62, 10, 1179-1187.

Noizet H. and E. Grosso (2011). The ALPAGE project: Paris and its urban area at the intersection of history and geography (9th-19th century). Proceedings of the 25th Conference of the International Cartographic Association, 3-8 July, Paris (France).

Pridal, P. and P. Zabicka (2008). Tiles as an approach to on-line publishing of scanned old maps, vedute and other historical documents. In: e-Perimetron, Vol. 3 (1), ISSN: 1790-3769, 10-21.

Sanderson, E. W. and M. Brown (2007). Mannahatta: An ecological first look at the Manhattan landscape prior to Henry Hudson. *Northeastern Naturalist*, vol. 14, 4, 545-570.