GeoPeuple project: using RESTful Web API to disseminate geohistorical database as open data

Authors

- **Eric Grosso**, IGN - French Mapping Agency, *France*
- **Christine Plumejeaud**, IGN - French Mapping Agency, *France*
- **Benjamin Parent**, IGN - French Mapping Agency, *France*

KEYWORDS : GeoPeuple, demographic evolution, Web API, geohistorical database, old maps, France

Introduction

For 250 years the French territory has considerably densified throughout population growth and technological advances, thus changing the topography of the French landscape. Moreover, industrialisation has generated a significant rural exodus. The study of the relationship between the evolution of the French landscape and the French population distribution evolution is relevant, both from a historical point of view and from a land settlement point of view: if the past is better understood, it is possible to better anticipate future needs. But this topographic past has been transmitted to us in the form of old maps (for example Cassini maps from the eighteenth century and Etat-Major maps from the nineteenth century), which offer a codified trace following the map specifications relative to their eras.

The GeoPeuple research project aims to analyse and extract by vectorisation the contents of old maps – Cassini maps (1:86,400 scale), Etat-Major maps (1:80,000 scale) and 1960's topographic maps (1:25,000 scale) – and to use current topographic databases in order to build a geo-historical (or spatio-temporal) database, and to study the densification of French territories. This project, funded by the ANR (French National Research Agency), brings together different partners: the COGIT laboratory of the French

Mapping Agency (IGN), the LaDéHis (historical demography laboratory) of the EHESS (School for Advanced Studies in the Social Sciences) and the MALIRE team of the LIP6 (Paris 6 laboratory in computer science).

Building such a geo-historical database requires different processes such as old maps scanning, identification of concepts and their organisation represented in an appropriate ontology, vectorisation (manual and automatic), georeferencing or data matching. We focus here on the diffusion process of the gathered data via an open Web server which allows for both an interactive exploratory spatio-temporal data analysis process of the change (administrative, demographic as well topographic) following precepts detailed in [1] or [2] and a free access to geo-historical data.

A geo-historical database gathering 200 years of evolutions

The geo-historical database produced during the GeoPeuple project contains demographic data (34 census since 1793 up to 2006), historical data describing the story of administrative entities and associated topographic data covering a 250 year period (since 1750 to present). The administrative entities in question are French municipalities ("communes" in French) and their upper entities. For all of them, the database has been designed to handle their change of name, border, membership to upper unit, or administrative center, all sorts of changes that happened frequently and have led to the disappearance of 7,300 communes from a total account of 44,000 since their date of creation in 1793. Changes are modeled as events that can occur either on a single entity (change of name for instance) or simultaneously on many entities, leading to the disappearance of entities by fusion inside others or the creation of new ones by land plot transfers of pre-existing entities or the division of pre-existing entities. Furthermore the design is not limited to administrative entities: it can handle various nomenclatures (religious, academic, fiduciary, judicial, etc.) and also particular cases such as the multiple co-belonging to various entities at the same time for one entity. In order to handle topographic data (roads, rivers, buildings, fabrics, mills, churches locations) collected throughout the digitalisation of Cassini maps (1759-1789) and Etat-Major maps (around 1890) as well as the border definition changes, this new database is now also a spatio-temporal one.

This is a new development with regards, for example, to the INSEE database [3], or to previous databases, whose content was explained in [4]. The current one is focused only on historical and demographic data and is connected to a Web interface [5]. Its success has proved the great interest that French citizens have for their villages' history. However this previous Web interface does not give a free access to these data nor does it allow for a real interactive exploratory data analysis process of this history in conjunction with topographic and demographic data.

The amount of work represented by the digitalisation of maps (even though limited to 4 maps covering 80 km by 50 km, for Saint-Malo, Reims, Grenoble and Agen locations) as well as building the map of French administrative entities' history is considerable, time consuming and very valuable for scientists such as historians, demographists or ecologists. Overall it has been funded with public funds and this is why the different partners of the project considered that these data are due to be accessible to all French citizens and decided to share this information, raising thus the question of the means of this diffusion.

Towards an open data diffusion allowing easy data handling and sharing

To answer this question, a first step is to provide everyone with free downloadable data, falling within the current open data context. However, even if data provided in this way is necessary, it is not sufficient. Indeed, this solution does not generally allow users to handle this data dynamically without reprocessing it using desktop software (mainly in order to manually recreate the dynamic links originally defined in the database management system), nor is it able to handle this data directly on the Web.

Therefore, the decision was made to create a website from which it is possible for example to search for the history of a place. Because the understanding of entity evolution is complex, based on different factors (topographic, politic and demographic ones), the results page is made of several synchronised views in four interactive tabs, with a time line similar to [6], [7] or [8]. The first tab displays topographic data at different periods, the second tab contains a systemic diagram showing the administrative entity relationships over time as shown in Figure 1, the third contains a chart showing the demographic

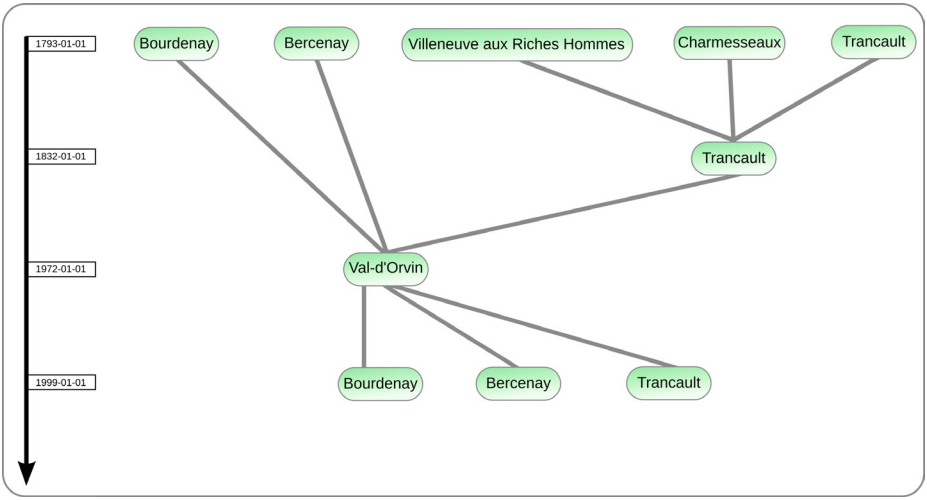

FIGURE 1

One website element : systemic diagram of Charmesseaux and its neighbouring administrative entities over time

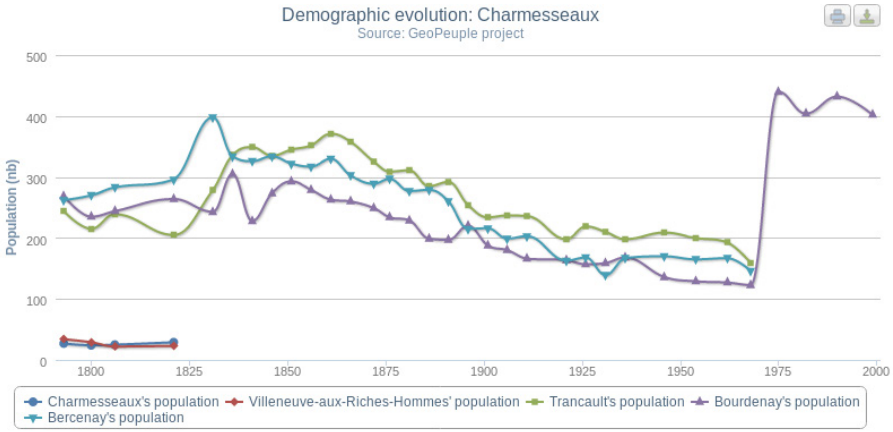

FIGURE 2

Another website element : demographic evolution of Charmesseaux and its neighbouring administrative entities over time

evolution of the studied administrative entity and its neighbouring entities as illustrated in Figure 2, and the last one combines the geometrical view of limits at various time with two other views: a demographic and systemic one presented simultaneously as proposed in [9].

This solution provides a user-friendly access to information and thus appears more flexible than just a simple download, however, it also presents two major drawbacks. The first is that a website does not allow full access to the information contained in the database, because it is limited to the original use for which the website was designed. The second is that the results displaying in HTML can not easily be handled by users. Therefore creating a mashup by crossing HTML results and external data appears quite complex for inexperienced users.

Consequently, we decided to provide a modular solution by developing a RESTful Web API which allows users to easily query the database, to serve several formats handle directly on the Web, and to build the GeoPeuple

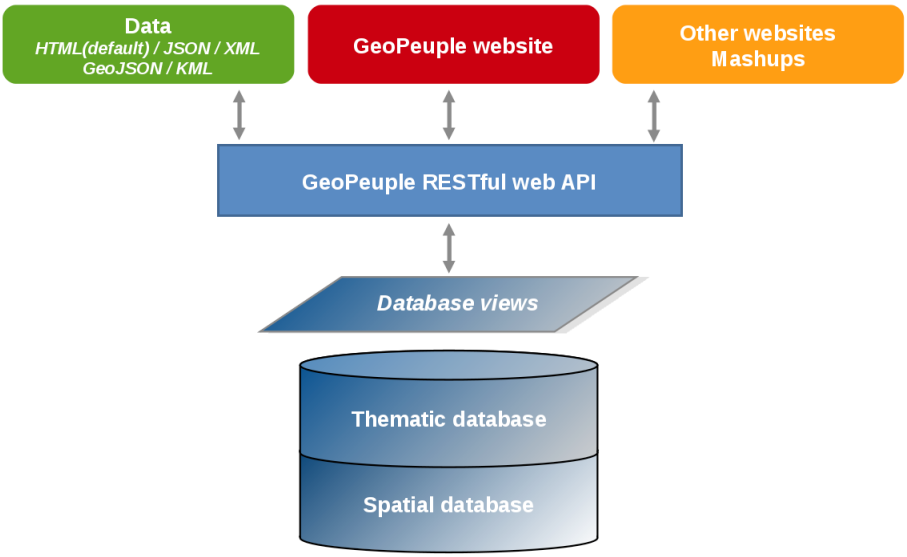

FIGURE 3

Global architecture of data diffusion

website on top. This global architecture is described in Figure 3. In order to help users query the Web API or to integrate the API easily into Web pages, HTML and JavaScript (AJAX) code snippets are provided. Furthermore, the code from this Web API will be released on an open source licence at the end of the project enabling the possibility for other projects to reuse similar code architecture.

Technical API description

In order to release the code under open source licence, the full architecture including the RESTful Web API and the website is based only on open source components. On the server side, PostgreSQL and PostGIS have been chosen to store and manage the geo-historical database, Java and the Spring framework to develop the RESTful Web API, Hibernate to map the object-oriented domain model to the database, Tomcat as a Web application server and GeoServer to share spatial data through Web Map Service. On the client side, jQuery is used to develop the code snippets, Raphaël JS to create the diagram showing the administrative entity relationships over time, Highcharts to create the demographic evolution charts (Highcharts is free for non-commercial use) and OpenLayers to display maps. Finally, the RESTful Web API serves W3C and OGC standard formats such as HTML (by default), XML and KML, but also JSON and GeoJSON to ease data handling.

Conclusion

This paper presented the architecture we chose to develop in order to share the geo-historical database produced in the GeoPeuple project. This architecture allows to share freely data that users can easily handle directly on the Web both by querying the RESTful Web API and by integrating HTML/jQuery code snippets directly on their own website.

A first advanced prototype has already been implemented (both server and client side). All the components used are open source allowing us also to release the RESTful Web API and GeoPeuple website codes under an open source licence (GNU Affero GPL v3 license). This will be done before mid-2013. Finally and as a perspective, the RESTful Web API could easily be extended to allow users to improve the database by crowdsourcing.

[1] ANDRIENKO, N., AND ANDRIENKO, G. *Exploratory analysis of spatial and temporal data*. Springer-Verlag, Berlin, 2005.

[2] ANTONI, J.-P., AND KLEIN, O., AND, MOISY, S. *Cartographie interactive et multimédia : vers une aide à la réflexion géographique*. Cybergeo : European Journal of Geography (Paris, France, 2004).

[3] Institut National de la statistique et des études économiques, 2012. History of French administrative entities. Retrieved September 7, 2012 from http://www.insee.fr/fr/methodes/nomenclatures/cog/historique.asp

[4] MOTTE, C., AND SÉGUY, I., AND THÉRÉ, C. *Communes d'hier, communes d'aujourd'hui*. Les communes de la France métropolitaine. Dictionnaire d'histoire administrative. Institut National d'Études Démographiques, Paris, 2003.

[5] MOTTE, C., AND VOULOIR, M.-C. EHESS Cassini website. Retrieved September 7, 2012 from http://cassini.ehess.fr/cassini/fr/html/1_navigation.php

[6] PISON, G., AND, MATHIAN, H., AND PLUMEJEAUD, C., AND GENSEL, J. *Exploring world demography on line*. In Proceedings of the 25th International Cartographic Conference (Paris, France, 2011).

[7] PISON, G., AND, MATHIAN, H., AND PLUMEJEAUD, C., AND GENSEL, J., 2011. Interactive maps of the world population. Retrieved September 7, 2012 from http://www.ined.fr/en/everything_about_population/interactive_maps/

[8] BRETAGNOLLE, A., AND, MATHIAN, H., AND VAN HAMME, A. *Animating the cities: dynamic exploration of harmonized urban databases (United-States – France, 1800-2000)*. In Proceedings of the 25th International Cartographic Conference (Paris, France, 2011).

[9] PLUMEJEAUD, C., AND MATHIAN, H., AND GENSEL, J., AND GRASLAND, C. *Spatio-temporal analysis of territorial changes from a multi-scale perspective*. International Journal of Geographical Information Systems, 2011.