RESEARCH ARTICLE

Dissemination and geovisualization of territorial entities' history

Christine Plumejeaud-Perreau[1,2],
Eric Grosso[1,3], and Benjamin Parent[1]

[1]COGIT Laboratory, French mapping agency (IGN), Saint-Mandé, France
[2]LIENSs Laboratory, U.M.R. CNRS, La Rochelle, France
[3]School of History, Classics and Archaeology, University of Edinburgh, Edinburgh, United Kingdom

Abstract: This paper describes an innovative solution for geovisualization of the demographic and administrative history of French municipalities, named "communes" in French. This solution allows for the open dissemination of such data. The challenge is to provide a web interface for unskilled users in order to help them understand complex information about the demographic evolution of French territories. Our approach combines interactive thematic, spatial, and temporal views. We describe our architecture, based on open-source technologies, and the organization of this imperfect geo-historical information in our spatiotemporal database. Our second contribution concerns the concept of an *acquaintance graph* that has been used to obtain an efficient design with good performance in our geovisualization website.

Keywords: geovisualization, web mapping, spatiotemporal databases, imperfect data, census data, historical GIS, spatial events

1 Introduction

The large-scale adoption of OGC standards and associated technologies for web mapping [20] has led to a dramatic increase in websites showing maps and data connected with the evolution of various phenomena. In the context of census data, the OECD Explorer[1],

---

[1]http://stats.oecd.org/OECDregionalstatistics

Google Public Data Explorer[2], GapMinder[3] project, the Web Atlas of World Population[4] [26] provide some examples of web-based geovisualization tools that are addressed to a large audience of non-specialist users (such as the general public). Those web applications make extensive use of interactive and dynamic maps and charts to portray demographic phenomena.

However, today such sites do not take into account the full complexity of demographic evolution implied by the life of socio-economic units [8]. This complexity is due to constantly changing boundaries and names of the territorial units upon which census collection is based. Despite some new proposals for viewing such information [28], demographic data is not usually displayed along with the evolving limits of these territories. Previous proposals are also typically limited to recent census data (since 1950 only), and typically do not allow for free access or the download of the data. Historical demography is essential for basic population research and policy analysis because models and descriptions of the past underlie theories of change and projections into the future. Demographic variation may have a variety of causes, such as the changing spatial extents of the territorial units. Demographers must refer to multiple historical sources in order to understand demographic variations, unpicking which changes are the results of spatial variation in territorial boundaries, and which are not. Allowing online access to all the collected information (not only demographic data, but also names and spatial extents of historical territories) can therefore assist in efficient analysis of this data.

In the US, the Minnesota Population Center has led a project of open dissemination and visualization of historical census data and boundaries dating back to 1850 [19]. However, the resulting interface[5] is dedicated primarily to the download of various data sets, selected by a dynamic query, rather than to online geovisualization of those data sets. The site lets users choose from amongst many options for data extraction in order to construct a complex query that satisfies exactly their needs (see Figure 1). However, while this type of interface may be valuable for specialized scientific researchers, it is not well-adapted to the needs of non-specialist users who, for example, may simply wish to investigate the history of a town based purely on its name.

In France, the history and demography (e.g., number of inhabitants) of French municipalities (communes) is known since 1791 [22]. This information is disseminated through a website [23]. At the time this site was released (2006) the functionalities of this site were innovative. For instance, in addition to demographic data, the site allows users to search for a commune using toponyms, and displays the history of relevant communes in among the results. Similar results have now been achieved in other projects (that are not focusing on demographic data) such as PELAGIOS[6] and GAP[7] (Google Ancient Places [15]). GAP is a tool for exploring and reading texts that reference historical places, based on RDF triples that link to services such as *GeoNames*[8]. In the historical domain, many new projects for mapping historical facts and events have recently emerged [10]. Rumsey [30] has argued that interactive maps could be helpful for pedagogical purposes, provided their use was intuitive enough [24].

---

[2]http://www.google.com/publicdata/directory
[3]http://www.gapminder.org
[4]http://www.ined.fr/en/everything_about_population/interactive_maps
[5]http://sda.usa.ipums.org/cgi-bin/sdaweb/hsda?harcsda+all_usa_samples
[6]http://pelagios-project.blogspot.fr
[7]http://gap.alexandriaarchive.org/gapvis/index.html
[8]http://www.geonames.org/about.html

Figure 1: Browsing the demographic database of the Minnesota Population Center (IPUMS, http://sda.usa.ipums.org/cgi-bin/sdaweb/hsda?harcsda+all_usa_samples).

In France, GeoPeuple is a project that aims to take a fresh look at historical French censuses. GeoPeuple combines historian-demographers (LaDéHis Laboratory of the School for Advanced Studies in the Social Sciences, EHESS) and computer scientists (team MALIRE from LIP6 laboratory) with researchers that specialize in geographic information science (COGIT laboratory of the French Mapping Agency, IGN). One of the aims of the GeoPeuple project was to digitize historical topographic maps (such as Cassini and Etat-Major maps) in order to build a vector-based spatiotemporal database and provide spatiotemporal analysis of the collected topographic and demographic information at the same time [29]. A second aim was further to assist in pedagogical dissemination of such information. In this way, the evolution of historical boundaries of communes could be presented in a web interface, as well as providing open access to this data.

The challenge facing this system is to enable the export of data using open standards, but also to offer an intuitive view on such data. Historical map data is rather complex, in terms of changing boundaries, and also the imperfect nature of such historical data. Many spatiotemporal models and implementations have been proposed in order to solve the so-called "split tract problem" [11]. These include space-time composite models [17, 25]; ontological knowledge representation [9, 16]; object-oriented models [35]; and event-oriented models [36]. However, few of these models address in a seamless interface the combined problems of imperfect knowledge and visualization of boundary changes.

This paper is a detailed extension of a previous communication presented at the OGRS symposium, [12], which demonstrated our prototype. The first section describes in detail the nature of the information within our model, as well as the architecture of our open-source infrastructure that allows dissemination of open data about historical boundaries.

The second section focuses on the design of the web interface for geovisualization and analysis of such complex data. The interface is available online at http://geopeuple.coriolys. fr. Finally, in the conclusions we provide some insights into current and future work.

2   Historical boundary data

This section describes the challenges in modeling the demographic and administrative history of French municipalities. Specifically, in addition to the challenge of modeling evolving boundaries, the inherent imperfection in historical information is important to consider and model.

2.1   Handling hierarchies of evolving territorial units

Census data is collected on territorial statistical units having a name, a code, and a boundary. Frequently, these units may be grouped together into a hierarchy in order to form wider territorial statistical units, up to the state level. This hierarchy of territorial units is termed a *nomenclature*. In the European Union, the nomenclature for territorial statistical units (NUTS) is widely in use. NUTS has been updated at least seven times since its creation in 1980. In France, the COG [14] (Code Officiel Géographique) is based on the French municipality as the smallest territorial unit. The COG has been used since 1801, but has only been fully documented by INSEE (French National Institute for Statistical and Economical Studies) since 1943. The complete demographic and administrative history of communes from 1791 to today has been collected by historian-demographers like Claude Motte [22]. This substantial undertaking required the collection of census data for every commune, as well as information about the history of their various designations (toponyms), and relationships to other territorial entities. This database captures many different types of spatial events (absorption, reincarnation, merge, split, redistribution) in the evolution of communes, along with names, census figures, and other relationships to the COG and other nomenclatures.

A commune can belong to multiple nomenclatures that may be current or no longer existing. For instance, communes, which generally correspond to ancient parishes, have belonged to a specific hierarchy called "généralités" in French. Généralités initially emerged in the 15th century to help in organizing the raising of taxes [21]. Today, there exist some political attempts to group communes into new administrative structures named "intercommunalités" in order to reduce the fragmentation of the French territory. There are currently 36,700 communes in France, not many fewer than the 44,000 communes created in 1791, but many more than the 15,865 German units in 2012 [33]. In summary, multiple evolving hierarchies of territorial units need to be modeled to capture these complex interrelationships.

2.2   Uncertain boundaries

Generally, historical boundaries of territorial units are uncertain because there frequently exist no systematic maps of the historical subdivisions, at least up to 1950 and in some cases even subsequently. In certain cases, it is possible to compute historical boundaries based on knowledge of past events that have affected territorial units, as is the case for

French communes. Thus, when one commune has split into multiple pieces, it is sufficient construct the geometric union of today's commune boundaries in order to get the historical boundaries of past communes. For example, Val d'Orvin was formed from the merger of Trancault, Bercenay-le-Hayer, and Bourdenay in 1973 and existed until 1999. Equally, the reincarnation (restoration) of Trancault, Bercenay-le-Hayer, and Bourdenay allows known boundaries to be reinstated for historical time periods (from 1832 to 1973 for Trancault, Bercenay-le-Hayer, and Bourdenay), as shown in Figure 2.
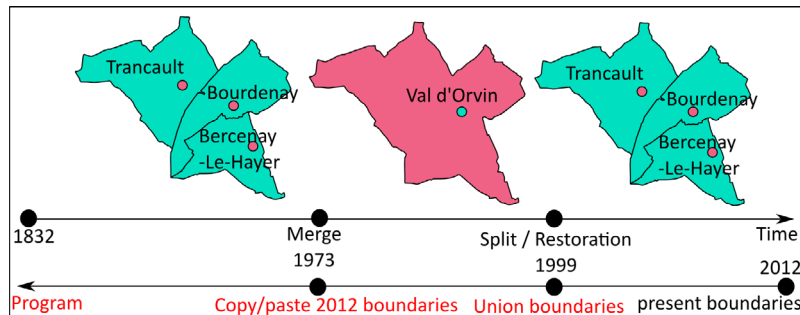

Figure 2: Simple cases of boundaries to reconstitute: split, reincarnation.

However, in other cases where two or more communes have merged in order to generate a new commune (like Charmesseaux, Trancault, Villeneuve-aux-Riches-Hommes which gave rise to Trancault in 1832), their historical boundaries may remain uncertain. For example, are the historical boundaries best represented by the black dashed lines or the white dashed lines in Figure 3? The same difficulty arises when two or more communes have had a piece of territory annexed in order to build a new commune, such as in the case of Echirolles, in 1833, constituted from parts of Grenoble, Eybens, and Jarrie (Figure 3).

2.3   Handling imperfect historical information

Going further into the past also leads to further uncertainty. Incompleteness and inaccuracy are common in older nomenclatures, even if not a significant problem in today's (such as NUTS which exists only since 1980). For instance, historical census data may be derived from old, hand-written books that may not be easy to decipher and are sometimes conflicting. In such historical sources, communes are designated by their toponym; but a toponym may change or disappear, perhaps because of a merge event. In addition to spelling mistakes, such changes can lead to substantial uncertainty for historians. The patient work of Claude Motte has helped to unravel such problems in the case of the French communes, using multiple sources and the knowledge that deductions might need to be retracted in the light of new historical sources that become available in the future. In the presence of contradictory sources, some historical knowledge is inevitably imprecise (e.g., "this merge seems to have happened between 1800 and 1805" or "this change of toponym occurred before 1793"). This contrasts strongly with more modern sources, where accurate information about the dates of events is typically available (e.g., "those three communes were merged on the 11 January 1973, after the official source referenced 'A08-12-1972'.")

This project does not aim to *solve* conflicting historical information as Claude Motte did. Instead, our aim is to provide a framework to handle the results of such historical
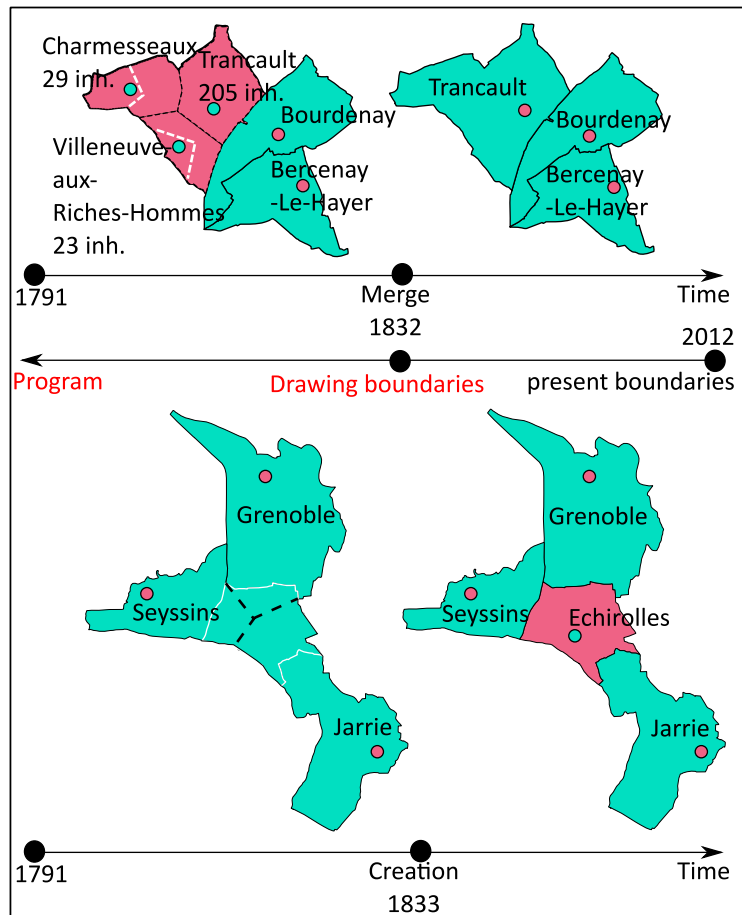
Figure 3: Complex cases of boundaries to reconstitute: merge, redistribution.

investigations of uncertain boundary evolution, and support to the dissemination of this information. Thus, it is essential to maintain fine-grained lineage metadata about historical information, being as accurate as possible concerning the dates of events.

3 Architecture

In order to handle hierarchies of evolving territorial units, we propose a data model which handles territorial units as objects with versioned attributes (toponyms, boundaries, codes, etc.). Our approach is based on an identity paradigm, particularly suitable for queries on evolving objects [34, 35]. Furthermore, evolving territorial units are associated with the set of events that have occurred to those entities (merges, splits, restorations, changes of name, of superior unit, of central place, etc.) [28]. Thus an event-oriented model is best suited for queries about such events [36]. This model, tested on the NUTS [27], is not limited

to queries about French communes. It is designed more generally for evolving territorial units, i.e., for any nomenclature and hierarchy of levels.

One option considered was to use ontological mechanisms to capture boundary changes, such as proposed by [16] or [9]. For example, [9] has addressed similar challenges in the context of Swiss municipalities since 1960. Even though the Federal Statistical Office (FSO) has recorded changes of names, superior entities, and events since 1960, this work does not aim to reconstruct historical boundaries. Instead, all boundaries are those existing since 1990, because geometry data sets were provided only since 1990 by the FSO and since 2000 by the FOT (Federal Office Of Topography). Such ontological mechanisms are helpful to make inferences about changes, but do not allow the reconstruction of historical boundaries as we aim to achieve. Nor can these existing approaches explicitly handle imperfection in the dating of events, as we aim to in this work.

3.1 Dating model

Metadata, compliant with ISO 19115, is used to store the fine-grained lineage of data sources. This lineage is required in order to document the imperfection of the data: uncertainty, incompleteness, and inaccuracy. Associated with each class and relationship in the model (for example, *Boundary*, *Toponym*, *Census*, *CentralPlace*, *Belonging*, *ChangingEvent* for instance), this metadata indicates the name of the contributor, the sources of the information, and the date of the update of the information (transaction time).

The imperfect nature of historical data has a critical impact on the temporal aspects of the model, since many events are not known precisely. Furthermore, many dates may indicate that the entity or one of its attributes (e.g., code, name, boundary) is still valid today. In the past, some models have used a date in the far future, like "2500:01:01," in order to indicate that data is "still valid now." However, such an approach is both inelegant and potentially invalid if the database survives more than four centuries (we hope so...). In PostgreSQL, it is possible to use the keyword "infinity" in order to express an unspecified end-date. But using a keyword specific to any DBMS will prevent the database from being fully compliant and interoperable. This issue is reminiscent of the past discussions on the semantics of "now" in databases [6]. Similarly, in our solution, we associate the notion of "still valid now" with a Boolean attribute (*isForever*) set to true. Thus, to avoid potential implementation incompatibilities we explicitly store information about unbounded end dates as an element within our data model.

Our dating model is shown in Figure 4. The model allows timestamping of any valid item of the model through the use of the *HistoricalDate* object. *HistoricalDate*s have a unique system identifier, an integer which may be referred to by any timestamped attribute, which can reduce redundancy in date storage. A *HistoricalDate* is always associated: a. to a temporal resolution *DateResolution* (which can vary from a second to a century); and b. to a temporal operator *DateOperator*. The *DateOperator* specifies whether this date is given as an instant (e.g., the census happened on a precise day in 1793), or as an interval (e.g., the merging event occurred between January 1800 and January 1805). Dates can also be specified using a textual form, sometimes used by historians to describe a historical event (such as the "Death of the Emperor Napoléon"). The advantage of this form is that it leaves open the debate concerning the uncertainty of an event date, and allows the implicit representation of imperfect knowledge for integration with future historical findings.
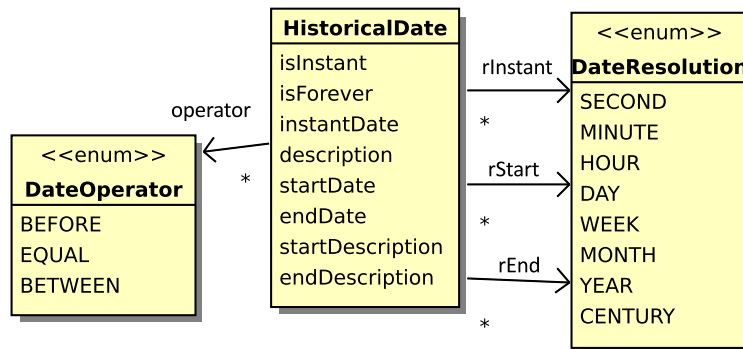
Figure 4: Dating model for imperfect data used in GeoPeuple.

This basic dating model could be further extended, for example, using a fuzzy approach (e.g., [31]), whereby a date matching's confidence could be calculated for each item. We might imagine linking such confidence intervals to maps and charts for displaying uncertainty, such as using a boundary's transparency level to depict level of confidence. However, for now, this basic model supports the full semantics required by historians, thus answering the main need of historians to be able to retrieve as much information as they have entered in the system.

3.2   Principles of boundary reconstruction

Concerning the reconstruction of historical boundaries, our algorithm queries events sorted in reverse chronological order, and uses known boundaries as far as possible to explain unknown historical boundaries, as illustrated in Figure 2. In the case of the French communes, more than half of the boundaries can be inferred in this way, because there were many boundary merges followed by reincarnations, and simple splits in French territorial history. Next, we make the assumption that the location of the central place of each historical commune endures: it must always be spatially inside the boundary of the commune. In fact, the coordinates of such locations are frequently known. In most cases, such central places have either not disappeared, or have been documented by archaeologists and historians due to their historical significance as main human settlements with solid buildings.

In the absence of any clues as to the location of a boundary, our heuristic is to then apportion area in proportion to population in each commune at the time of the event. For instance, after the creation of Echirolles in 1833, Grenoble lost an eighth of its inhabitants (modulo the natural increase/decrease rate of population in this region at that time). Thus, in the absence of any other information about the boundary location, that part of the boundary crossing the Grenoble area would be drawn in order to take only the eighth of the Grenoble land area. This crude approximation assumes a uniform population density, which is of course far from the geographic reality.

In at least some cases a human operator may be able to set a territory boundary manually, for example to match known topographic patterns. In these cases any available cartographic information (e.g., photography, textual source, historical or recent topographic maps) could be used in support of estimated historical boundaries, which frequently follow roads, the borders of woodlands, or rivers.

4 Implementation

The spatiotemporal model was implemented using an open-source DBMS, PostgreSQL, with its spatial extension PostGIS. The focus here is on two major difficulties: the temporal aspects of the model, and the rebuilding of historical boundaries of communes.

This temporal model addresses historians' needs in term of content, but also in terms of computing performance. A set of SQL procedures have been stored inside the database in order to accelerate the processing of temporal queries. For example:

- Get the system identifier of a date equal to 1793 with year-level precision (e.g., returns "13" as an integer identifier).
- Get the system identifier of a date interval between to 1800 and 1805, with year-level precision (e.g., returns "189" as an integer identifier).
- Get the oldest date of a historical date having the identifier 189 (e.g., returns "01-01-1800" as an SQL Date-type instance).
- Is the historical date with identifier 13 (strictly) before the date with 189 as its system identifier? (Returns "true" as a Boolean).

Instead of using a full temporal algebra (e.g., Allen [1]) this set of procedures allows for the computing of precedence and equality between dates (specified as intervals or instants). Precedence and equality are necessary and sufficient for the application of qualitative reasoning tools, such as the one developed by [7] based on theory of S-languages [32].

Our method for computing historical boundaries is based on that developed by Hervé Le Bras, from the LaDéHis laboratory, who has programmed FORTRAN code in order to automate this task. As outlined in Section 2.2, the procedure consists of backtracking from present boundaries back into the past, using knowledge about territorial events that changed commune boundaries, and using demographic data and central place locations. To achieve this, the spatiotemporal model of territorial entities and events is required. First, the program navigates the historical graph in order to reconstruct logical boundaries, as explained in Figure 2. This stage is able to handle events of type "split," "merge," and "reincarnation." Then, more challenging cases are handled based on the principles discussed in Section 2.2, summarized in Figure 3. The procedure initially computes unknown boundaries using known recent boundaries as a first constraint; known locations of central places as a second constraint; and areas in proportion to inhabitants as a third constraint (see Section 2.2).

The process might stop there. However, in addition our system offers the option of manually processing difficult cases in order to further enhance the results. Thus, an operator can use an interface to superimpose historical maps or photographs. This interface also provides the operator with constraints such as the locations of historical places and the limits of recent boundaries. The interface allows the operator to interactively compute the relative proportions of populations expected to be distributed as a result of redrawing of historical boundaries, indicating if putative boundaries conform to known demographic distributions. In this way, 18,152 boundaries have been inferred over the whole France, many cases of which are shown in Figure 5. The FORTRAN code used is currently being ported to Python, in order to build an open-source plugin for QGIS that could be reused by anyone.
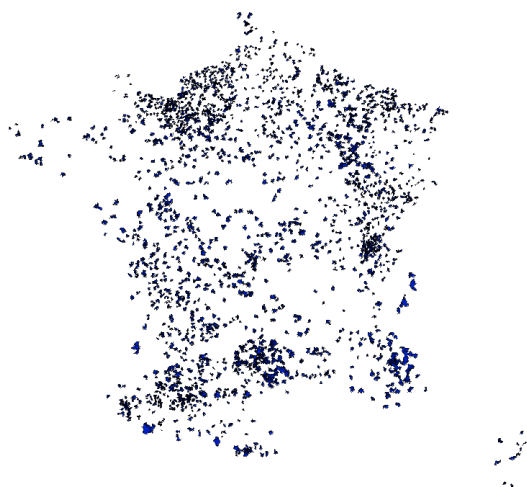
Figure 5: Set of historical boundaries that have been reconstructed (only difficult cases).

4.1   Architecture for open dissemination of historical boundary data

The architecture design is driven by the desire to provide free access to downloadable historical data, within the current open-data context. A conventional download page, or even a web interface to query the database, would be one mechanism to satisfy this desire, as used by the Minnesota Population Center (see the Figure 1). However, using desktop software the data would require special handling, for example in order to recreate manually the dynamic links originally defined in the DBMS.

Thus, an alternative solution is to develop a website dedicated to the geovisualization of this data. This alternative could provide user-friendly access to historical data, and thus appears more flexible and usable than a simple data download site. However, it also presents two major drawbacks. First, such a website would not necessarily allow full access to the database, being limited to the original use for which the website was designed. Second, the results of queries displayed using HTML cannot be easily handled by users. Creating a mash-up of HTML results and external data is expected to be a complex undertaking for novice users.

Consequently, it was decided to provide a modular solution, by developing a web API that allows users to easily query the database; serves several formats handled directly on the Web; and supports the construction of the GeoPeuple website on top of these API capabilities. The overall architecture is summarized in Figure 6. The web API is based on the RESTful API initially, but future adaptations could use other technologies. The point is to provide a documented interface based on web standards that is independent of our specific implementation.

At the API level, our choice is identical to that of the *AddressingHistory*[9] project [18]. However, our implementation adds temporal criteria. The RESTful web API serves W3C and OGC standard formats, such as HTML (by default), XML, and KML, as well as JSON and GeoJSON to ease data handling.

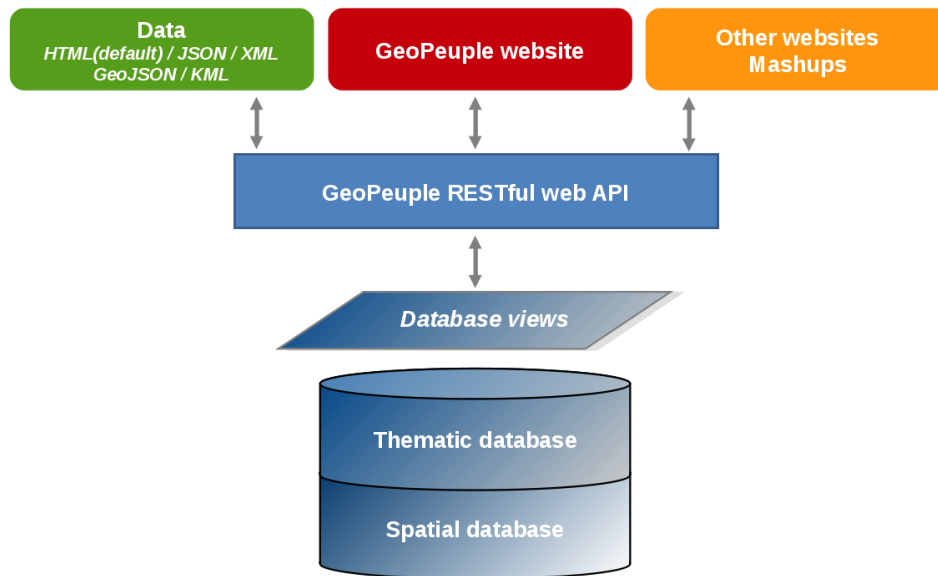[9]http://addressinghistory.edina.ac.uk

Figure 6: Global architecture of data diffusion.

In order to release the code under an open-source license, 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 geohistorical database. Java and the Spring framework were used to develop the RESTful web API. Hibernate was used to map the object-oriented domain model to the database. Finally, Tomcat was used as a Web application server, and GeoServer to share spatial data through the Web Map Service (WMS).

5   Web analysis interface

This penultimate section focuses on the geovisualization aspects of this research. Our goal is not only to disseminate open data, but also to help users in understanding complex information about the evolving boundaries of communes since 1791 and assist in interpreting demographic dynamics at local levels. As in [28] and many previous works in visual analytics [2–5], our aim is to offer various synchronized points-of-view of the studied phenomena in an interactive and dynamic way.

The resulting website is composed of a first page allowing users to search based on various criteria (thematic, spatial, temporal, or spatiotemporal) using the capabilities of the web API. The main work concerns the geovisualization of the results. These are displayed on a second web page (see Figure 7) using four synchronized tabs: "Cartes" (maps), "Historique" (historical graph), "Démographie" (demography), and "Densité" (density). The tabs allow users to access commune information corresponding to these different research criteria.
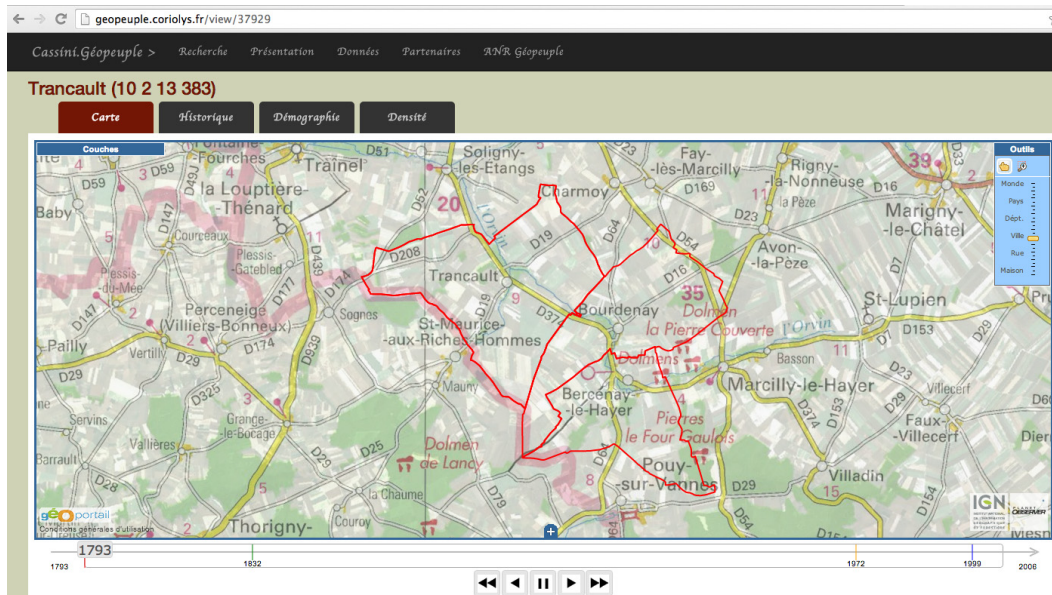
Figure 7: GeoPeuple website currently hosted by Coriolys (http://geopeuple.coriolys.fr) showing map view page, and access to other tabs.

The first tab, "Cartes," shows the spatial view, which contains a map with a temporal slider in order to visualize evolving boundaries (Figure 7). The third tab ("Démographie") displays the demographic curves of each commune. The last tab ("Densité") shows the evolution of population density through time. Thus, our approach is similar to the leading websites for viewing census data (including as already introduced OECD Explorer, Google Public Data Explorer, GapMinder, and the Web Atlas of World Population [26]), except that in our site we display historical data.

Our key innovation compared to these sites, and to other sites dedicated to historical demography data (e.g., the Cassini website [23]) is in the second tab, "Historique," which presents the *systemic diagram*. This diagram, exemplified in Figure 8, represents in a schematic way the relationships between communes by way of territorial events (merges, splits, restoration, shown using grey links) and changes of names (shown using red links).

For example, a search for a commune named "Charmesseaux" reveals via the systemic diagram that Charmesseaux was merged into Trancault in 1833; and then Trancault itself was merged into Val d'Orvin in 1973; before the restoration of Charmesseaux in 1999. The form of the diagram is not unique, because for a given date the display order for entities is not fixed. However, in drawing this graph, the algorithm takes care to group together entities having being involved in the same event for the subsequent dates. Groups are ordered in chronological order of events. As far as possible, this approach avoids crossing edges in the graph.

Drawing this kind of graph necessitated the design of an efficient exchange format between server- and client-side components. In fact, web mapping tools using WMS or WFS for data exchange generate a request from the client each time a spatiotemporal extent is modified by the user, leading to a high computational overhead. This may be accept-
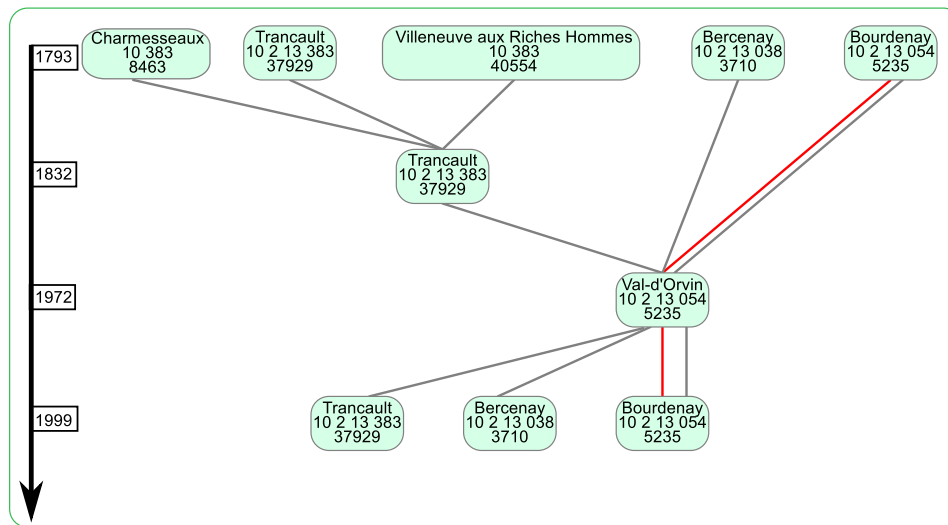
Figure 8: Systemic diagram of Charmesseaux and its neighboring administrative entities over time.

able when using a cloud-based distribution of data and servers, such as used by Google. However, our open architecture should not require hundreds of servers in order to fulfill geovisualization needs. To be free and easy to deploy requires a lightweight, standalone architecture. Furthermore, the client is also built using open-source technologies, with as few specialized frameworks as possible in order to simplify the architecture and minimize maintenance. Our solution uses HTML 5, Javascript, JQuery, and three specialized libraries for charts and map-making: HighCharts[10] to create the demographic evolution charts (Highcharts is free for non-commercial use); Raphaël JS[11] to create the systemic diagram; and OpenLayers[12] to display maps. Building on the fact that the RESTful API serves JSON, which is fully compliant and easy to use with JavaScript technologies, we embed all the commune's history and relationships (with related history) into one unique exchange format, expressed in GeoJSON. We call this the *acquaintance graph*.

5.1 Constructing acquaintance graphs

An *acquaintance graph* contains all communes having had in common some parcels with the commune searched for by the user in the database. Each commune is described by its various toponyms, codes, and boundaries, as well as by all the territorial events in which it was involved, and by its census values since 1791.

There are two reasons to serve data in such a way:

(1) For performance, since it avoids many exchanges between server and client. There is no need to make multiple requests of the server when one wishes to present data concerning neighboring communes to that first searched.

---

[10]http://www.highcharts.com
[11]http://raphaeljs.com
[12]http://openlayers.org

(2) For pedagogic purposes, because it helps the user to understand that rapid changes of population in one commune may be due to changes in boundaries, such as merging between communes. Further, when census data is missing (such as for Trancault between 1973 and 1999) the user can verify that this is due to a merging event and not to missing data.

Acquaintance graphs are pre-computed on the server by an iterative call to the database, searching for territorial units having events in common. An index connecting a commune's identifier to its corresponding acquaintance graph is built, in order to respond rapidly to queries from the client, since many communes share the same graph. The answer is both human-readable, and can be easily processed by a JavaScript interpreter (see the example in Annex 1).

5.2   Using GeoJSON to visualize boundary changes of territorial units

The tab named "Cartes" displays an interactive and dynamic map of evolving boundaries through time, using the GeoPortal API of the French Mapping Agency (IGN). For example, Figure 7 shows the boundaries of Trancault, Bourdenay, and Bercenay-le-Hayer in 1793 (in red) overlaid on the IGN road map. This extension of OpenLayers allows for the superposition of IGN topographic maps (current or historical ones, such as Etat-Major) and georeferenced aerial photography with the timestamped layers served using GeoJSON. A time-slider made with JQueryUI can be used either to select the date to display, or to animate automatically the evolution of the commune, via standard play/stop/rewind buttons. A special mechanism for handling layer transparency has been implemented in order to smooth the appearance and disappearance of each layer according to the valid time selected in the time-line.

5.3   Metadata for interactive demographic or density graphs

Finally the tabs named "Démographie" and "Densité" allow the user to examine more deeply how population of communes has evolved through the display of interactive population curves and density bar charts. These two views are complementary, since a commune like Bercenay in 1836 can concurrently experience increasing population (Figure 9) and decreasing density (Figure 10). Thus users can more easily understand that observed population increases are not due to a sudden demographic boom, but merely to an extension of the commune's area following a territorial event. Displaying the data in this way, therefore, suggests a different interpretation compared to increases of population within unchanging land areas. The options for interactivity (zoom, pan, and selection of one or more communes) are provided through the functionality of the HighCharts library, as well as the possibility to export those graphics to PDF or PNG formats. Finally, one can access metadata on the graphics, used for example to display the reason for missing values (as shown on Figure 9 for Charmesseaux, which disappeared after 1832).

6   Conclusion

This paper has presented the architecture developed in order to share the geo-historical database produced in the GeoPeuple project through an online geovisualization interface.
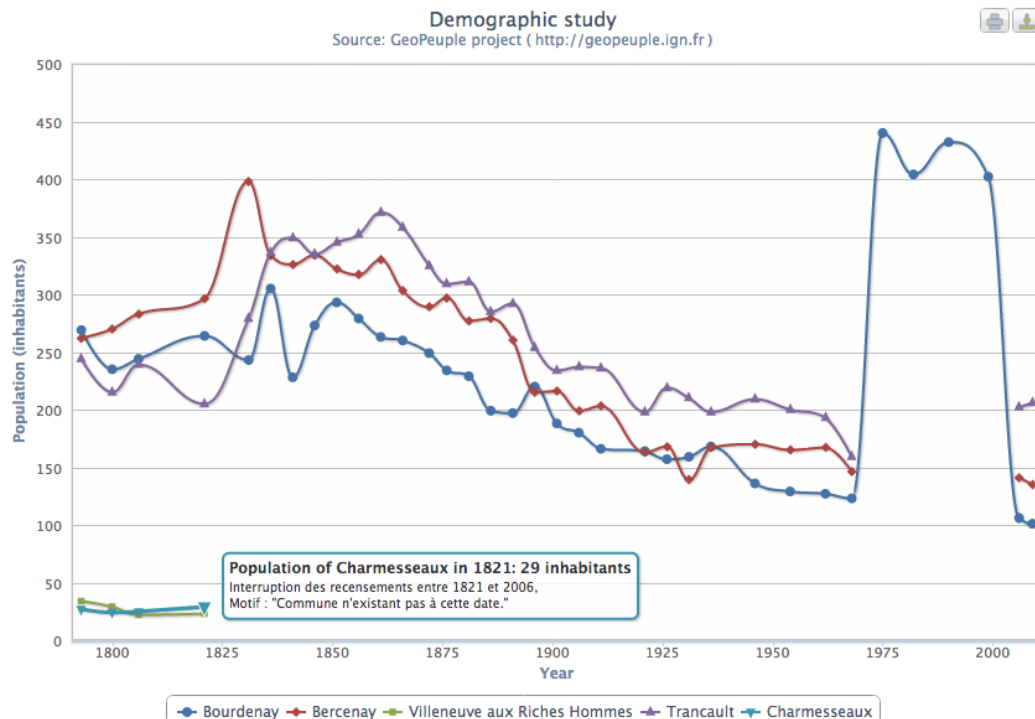
Figure 9: Demographic evolution of Trancault and its neighboring administrative entities over time.

Such historical content generally requires consideration of uncertainty, inaccuracies, and incompleteness. Our approach does not aim to unravel all the conflicts arising from this imperfect knowledge. Rather, the contribution of historians has been to conduct this investigation, and provide us with a corpus of expert knowledge about this uncertain past. Instead, this paper demonstrates that such historical content requires special handling: support for imperfect dating of events, and the substantial task of boundary reconstruction, based partly on demographic hypotheses.

We have addressed these challenges by proposing a model of imperfect data that satisfies historians' needs in terms of metadata capture and maintenance. Further, based on an event-oriented spatiotemporal model, the database has been constructed in order to capture and store historical boundaries, using procedures for semi-automatic acquisition of historical boundaries. The resulting architecture supports free sharing of this data through a web API.

Users can easily view and analyze the data (demographic figures, historical and contemporary boundaries) through a web browser. On the geovisualization side, the concept of the acquaintance graph, represented in a GeoJSON format, has been introduced in order to improve client/server exchanges and to offer a more complete and pedagogical view on demographic phenomena. This format has been used as the basis of four synchronized views of the evolving boundaries, the census data and associated metadata, and a systemic graph showing the evolution of communes in relation to each other. Finally, the design
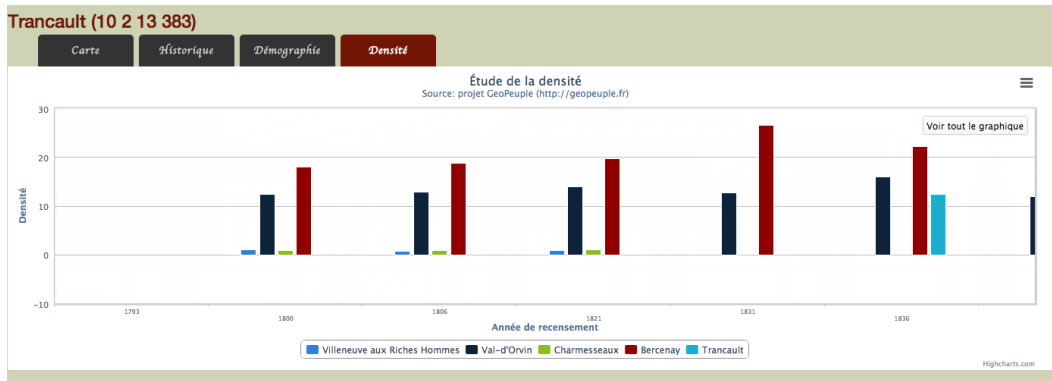
Figure 10: Density evolution of Charmesseaux and its neighboring administrative entities over time, focusing on the period 1793–1836.

also makes use of the GeoPortal API in order to help people to situate any event inside the present (or past) topographic context. All this has been achieved using standard open-source web technologies, such as HTML5, JavaScript, and SVG, in order to ease maintenance and deployment.

There are still improvements to be carried out before the deployment of this architecture on the "TGIR Huma-num"[13] site, a French national platform for humanities data. In connection with the other activities in the GeoPeuple project, one such improvement is to visualize in dedicated thematic layers historical topographic data (mills, farms, churches, castles, etc.) that have been digitized on Cassini and Etat-Major maps [29], although only in a small number of few areas (around Grenoble, Saint-Malo and Reims).

In the future, we also envisage citizens digitizing historical topographic maps through the use of an online crowd-sourced tool, much as can already be done with OpenStreetMap for contemporary topographical maps. Such historical data would be most useful for our understanding of the past of the humanity [13,18]. Our works brings this crowd-sourcing tool a step closer, since our database has been designed to support fine-grained lineage information including transaction metadata. Finally, we hope that the interest in historical data can be enhanced by our proposal for the interactive exploration of territorial entities' history, knowing that our model can handle any kind of territorial nomenclature, not only French ones.

Acknowledgments

---

[13]http://www.huma-num.fr

that is planned for release before the end of 2014 on the TGIR Huma-num site, http://sig.geopeuple.huma-num.fr.

References

[1] ALLEN, J. F. Maintaining knowledge about temporal intervals. *Communications of the ACM 26*, 11 (1983), 832–843. doi:10.1145/182.358434.

[2] ANDRIENKO, N., AND ANDRIENKO, G. *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer, Secaucus, NJ, 2005.

[3] ANDRIENKO, N., ANDRIENKO, G., AND GATALSKY, P. Exploring change in census time series with interactive dynamic maps and graphics. *Computational Statistics 16*, 3 (2001), 417–433. doi:10.1007/s001800100076.

[4] ANTONI, J.-P., KLEIN, O., AND MOISY, S. Cartographie interactive et multimédia: vers une aide à la réflexion géographique. *Cybergeo: European Journal of Geography*, 288 (2004), 12–24. doi:10.4000/cybergeo.2621.

[5] BRETAGNOLLE, A., LIZZI, L., MATHIAN, H., AND VAN HAMME, A. Animating the cities: Dynamic exploration of harmonized urban databases (United States, France 1800–2000). In *Proc. 25th International Cartographic Conference* (Paris, Palais des Congrès, France, Jan. 2011), Springer, pp. 130–147.

[6] CLIFFORD, J., DYRESON, C., , ISAKOWITZ, T., JENSEN, C. S., AND SNODGRASS, R. T. On the semantics of "now" in databases. *ACM Transactions on Database Systems 22*, 2 (1997), 171–214. doi:10.1145/249978.249980.

[7] DURAND, I. A., AND SCHWER, S. R. A tool for reasoning about qualitative temporal information: The theory of S-languages with a Lisp implementation. *Journal of Universal Computer Science 14*, 20 (November 2008), 3282–3306. doi:10.3217/jucs-014-20-3282.

[8] FRANK, A. U., CHEYLAN, J.-P., AND RAPER, J. *Life and Motion of Socio-Economic Units*. European Science Foundation GISDATA. Taylor and Francis, London, May 2001.

[9] GANTNER, F. *A Spatiotemporal Ontology for the Administrative Units of Switzerland*. PhD thesis, University of Zürich, Switzerland, April 2011.

[10] GIGAND, J.-F. Space and time mapping of historical facts. In *Proc. International Cartography Conference* (Paris, France, 2011), International Cartographic Association, Springer, pp. 223–234.

[11] GREGORY, I. Time-variant GIS databases of changing historical administrative boundaries: A European comparison. *Transactions in GIS 6*, 2 (2002), 161–178. doi:10.1111/1467-9671.00103.

[12] GROSSO, E., PLUMEJEAUD, C., AND PARENT, B. GeoPeuple project: Using RESTful web API to disseminate geohistorical database as open data. In *Proc. Open Geospatial Research & Education Symposium (OGRS)* (Yverdon-les-Bains, Swiss, 24-26th October 2012), O. Ertz, S. Joost, and M. Tonini, Eds., OGRS, pp. 222–228.

[13] GUILLAUD, H. Qu'apportent les digital humanities? In *Read/Write Book 2: Une intro-duction aux humanités numériques*, P. Mounier, Ed. OpenEdition Press, Marseille, 2012, pp. 69–73.

[14] INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES. History of French administrative entities. http://fr/methodes/nomenclatures/cog/historique.asp, 2012.

[15] ISAKSEN, L., BARKER, E., KANSA, E. C., AND BYRNE, K. GAP: A NeoGeo approach to classical resources. *Leonardo 45*, 1 (2012), 82–83. doi:10.1162/LEON_a_00343.

[16] KAUPPINEN, T., VÄÄTÄINEN, J., AND HYVÖNEN, E. Creating and using geospatial ontology time series in a semantic cultural heritage portal. In *Proc. 5th European Se-mantic Web Conference (ESWC)* (2008), Springer, pp. 110–123.

[17] LANGRAN, G. E., AND CHRISMAN, N. R. A framework for temporal geographic information. *Cartographica 25*, 3 (1998), 1–14. doi:10.3138/K877-7273-2238-5Q6V.

[18] MACDONALD, S. AddressingHistory: A Web 2.0 community engagement tool and API. *IASSIST Quarterly 35*, 3 (2011).

[19] MACMASTER, R. B., LINDBERG, M., AND VAN RIPER, D. The National Historical Geographic Information System (NHGIS). In *Proc. International Cartographic Conference (ICC)* (2003), pp. 88–102.

[20] MITCHELL, T. *Web Mapping Illustrated: Using Open Source GIS Toolkits*. O'Reilly Media, Sebastopol, CA, 2005.

[21] MOTTE, C. *Evolution of Administrative Divisions in Europe*. Palgrave Macmillan, Bas-ingstoke, UK, 2013, ch. France, pp. 12–35.

[22] MOTTE, C., SÉGUY, I., AND THÉRIER, C. *Communes d'hier, communes d'aujourd'hui. Les communes de la France métropolitaine, 1801–2001. Dictionnaire d'histoire administrative*. Institut National d'Études Démographiques, Paris, France, 2003.

[23] MOTTE, C., AND VOULOIR, M.-C. EHESS Cassini website. http://cassini.ehess.fr/cassini/fr/html/1_navigation.php, 2006.

[24] NIVALA, A.-M., BREWSTER, S., AND SARJAKOSKI, T. Usability evalua-tion of web mapping sites. *Cartographic Journal, The 45*, 2 (2008), 129–138. doi:10.1179/174327708X305120.

[25] NORMAN, P., REES, P., AND BOYLE, P. Achieving data compatibility over space and time: Creating consistent geographical zones. *International Journal of Population Geog-raphy 9* (2003), 365–386. doi:10.1002/ijpg.294.

[26] PISON, G., MATHIAN, H., PLUMEJEAUD, C., AND GENSEL, J. Exploring world demography on line. In *Proc. 25th International Cartographic Conference (ICC)* (Paris, France, 2011), Springer, pp. 45–59.

[27] PLUMEJEAUD, C. *Models and methods for evolutive spatio-temporal information*. PhD thesis, Université de Grenoble, France, 2011.

[28] PLUMEJEAUD, C., MATHIAN, H., GENSEL, J., AND GRASLAND, C. Spatio-temporal analysis of territorial changes from a multi-scale perspective. *International Journal of Geographical Information Systems 25*, 11 (November 2011), 1597–1612. doi:10.1080/13658816.2010.534658.

[29] RUAS, A., PLUMEJEAUD-PERREAU, C., NAHASSIA, L., GROSSO, E., VOULOIR, M. C., AND MOTTE, C. GéoPeuple: The creation and the analysis of topographic and demographic data over 200 years. In *Cartography from Pole to Pole* (Dresden, Germany, 2013), M. Buchroithner, N. Prechtel, and D. Burghardt, Eds., Lecture Notes in Geoinformation and Cartography, Springer, pp. 003–018.

[30] RUMSEY, D., AND WILLIAMS, M. Historical maps in GIS. In *Past Time, Past Place: GIS for history*, A. Knowles, Ed. ESRI Press, Michigan, US, 2002, pp. 1–18.

[31] SCHOCKAERT, S., AND DE COCK, M. Temporal reasoning about fuzzy intervals. *Artificial Intelligence 172*, 8–9 (2008), 1158–1193. doi:10.1016/j.artint.2008.01.001.

[32] SCHWER, S. S-arrangements avec répétitions. *Comptes Rendus de l'Académie des Sciences, Mathématiques 334* (2002), 261–266.

[33] SOHNLE, J. L'intercommunalité, une alternative aux fusions des communes? *Pouvoirs Locaux II*, 81 (2009), 74–80.

[34] STOCK, K. Spatio-temporal data management using object lifecycles: A case study of the australian capital territory spatial data management system. *Journal of Spatial Sciences 51*, 1 (2006), 43–56.

[35] WORBOYS, M. F. A generic model for spatio-bitemporal geographic information. In *Spatial and Temporal Reasoning in Geographic Information Systems.*, M. Egenhofer and R. Golledge, Eds. Oxford University Press, Oxford, UK, 1998, pp. 25–39.

[36] WORBOYS, M. F. Event-oriented approaches to geographic phenomena. *International Journal of Geographical Information Science 19* (2005), 1–28. doi:10.1080/13658810412331280167.

7 Annex

```
1  //acquaintance graph in GeoJSON format
   {"communeList":[//list of all communes having existed or still existing involved
      in the acquaintance graph.
3    {"id":37929, //commune having the system id "37929"
      "evolution":[ //chronologically ordered list of changes that have affected
          commune "37929"
5      //'year' gives the lowest bound of the change date, in a java date format,
            while 'displayDate' gives the exact historical date as string format
        {"date":1793,"displayDate":"1793","InseeCode":"10 2 13 383","motif":0,"
            officialName":"Trancault","otherToponyms":[]},
7      {"date":1832,"displayDate":"1832","InseeCode":"10 2 13 383","motif":32,"
            officialName":"Trancault","otherToponyms":[]},
       {"date":1972,"displayDate":"1972","InseeCode":"10 2 13 383","motif":31,"
            officialName":"Trancault","otherToponyms":[]},
```

```
9          {"date":1999,"displayDate":"1999","InseeCode":"10 2 13 383","motif":21,"
               officialName":"Trancault","otherToponyms":[]}
         ],
11       "eventList":[ //chronologically ordered list of territorial events that
             involved many communes with commune "37929"
           //'id' is the system identifiant of the event
13         //'motif' is a code indicating the nature of the event and what happened to
               the commune during this event
           //'init' is true if the commune was existing before the event
15         //'end' is true if the commune was existing after the event
           {"id":7392,"motif":32,"generalMotif":3,"motifName":"Absorbing commune of the
               merge","date":"1832","displayDate":"1832","comment":"null",
17           "entitiesList":[{"id":8463,"motif":31,"init":true,"end":false},{"id"
               :37929,"motif":32,"init":true,"end":true},{"id":40554,"motif":31,"init
               ":true,"end":false}]},
           {"id":1044,"motif":31,"generalMotif":3,"motifName":"Merge","date":"1972","
               displayDate":"1972","comment":"null",
19           "entitiesList":[{"id":3710,"motif":31,"init":true,"end":false},{"id":5235,
               "motif":32,"init":true,"end":true},{"id":37929,"motif":31,"init":true,
               "end":false}]},
           {"id":7393,"motif":21,"generalMotif":2,"motifName":"Restoration","date":"
               1999","displayDate":"1999","comment":"null",
21           "entitiesList":[{"id":3710,"motif":21,"init":false,"end":true},{"id":5235,
               "motif":22,"init":true,"end":true},{"id":37929,"motif":21,"init":false
               ,"end":true}]}
         ],
23       "census":[//chronologically ordered list of census for commune "37929"
           //'pop' is the number of inhabitants
25         //'order' is the census order in time series
           //'isMissing' indicates if the value is missing, and the 'missinReasonCode'
               indicates the reason
27         //'area' gives the area of the commune at that time, it values −1 if the
               commune was not existing or if the boundary is unknown.
           {"pop":244,"date":1793,"order":1,"isMissing":false,"missinReasonCode":"","
               area":"−1"},
29         {"pop":215,"date":1800,"order":2,"isMissing":false,"missinReasonCode":"","
               area":"−1"},
           {"pop":239,"date":1806,"order":3,"isMissing":false,"missinReasonCode":"","
               area":"−1"},
31         "..."
           {"pop":159,"date":1968,"order":29,"isMissing":false,"missinReasonCode":"",
               "area":"27"},
33         {"pop":0,"date":1975,"order":30,"isMissing":true,"missinReasonCode":"...",
               "area":"27"},
           {"pop":0,"date":1982,"order":31,"isMissing":true,"missinReasonCode":"...",
               "area":"27"},
35         {"pop":0,"date":1990,"order":32,"isMissing":true,"missinReasonCode":"...",
               "area":"27"},
           {"pop":0,"date":1999,"order":33,"isMissing":true,"missinReasonCode":"...",
               "area":"27"},
37         {"pop":202,"date":2006,"order":34,"isMissing":false,"missinReasonCode":"",
               "area":"27"},
           {"pop":206,"date":2009,"order":35,"isMissing":false,"missinReasonCode":"",
               "area":"27"}
39       ]
       },
41     {"id":40554,//commune having the system id "40554"
         "evolution":[//chronologically ordered list of changes that have affected
             commune "40554"
```

```
43        {"..."
          }
45      ],
      "eventList":[//chronologically ordered list of territorial events that
           involved many communes with commune "40554"
47        "..."
      ],
49      "census":[//chronologically ordered list of census for commune "40554"
          "..."
51      ],
    },
53    "..."
    ],
55 "footprints": //list of all boundaries of each commune, structured by layers which
       are chronologically ordered
    {"BBox":{"type":"Polygon","coordinates":[[[734864,6801065],
57   [734864,6811485],[747185,6811485],[747185,6801065],[734864,6801065]]]},
    //'BBox' is the bounding box of the maximum spatial extent of all communes
        included in acquaintance graph
59   "layers":[//chronologically ordered list of layers
      {"id":0,"startDate":"1793", //fist layer, valid from 'startDate' (1793) given
          as java date format untill next one
61      "geometries":[//list of boundaries, with 'id_entity' giving the id of the
           corresponding commune
         {"id_entity":37929,"boundary":{"type":"MultiPolygon","coordinates"
            :[[[[744763,6806027],[744777,6805872],
            [744757,6806059],[744763,6806027]]]]}},
63        "..."
        ]
65    },
      {"id":1,"startDate":"1832","geometries":["..."]}, //second layer, valid from '
          startDate' (1832) given as java date format untill next one
67    "..."
    ]
69   }
}
```

Listing 1: Extract of a GeoJSON format describing an acquaintance graph.