ALPAGE: towards the setting-up of a collaborative work tool

Hélène Noizet, Alain Dallo, Georges-Xavier Blary, Laurent Costa, Frédéric Pouget

ARCHEOLOGIA
E CALCOLATORI

19

2008

ESTRATTO

All'Insegna del Giglio

ALPAGE: TOWARDS THE SETTING-UP
OF A COLLABORATIVE WORK TOOL

1. Introduction

In 2006, the French National Agency for Research (ANR) funded an interdisciplinary research program called ALPAGE (diachronic analysis of the urban space in Paris: a geomatic approach). It aims to develop a Geographic Information System (GIS) for the city of Paris within its boundaries before 1860 (The Wall of The Farmers General).

Led by the LAMOP[1], this program involves three other laboratories: ARSCAN[2], LIENSS[3] and L3I[4] and gathers some twenty people with three main types of skills, historical skills (with historians and archaeologists in Paris in the Middle Ages), geographical skills (especially GIS) and computer skills (in the field of automatic vectorization of images).

Our approach is, on the one hand, to reconstruct the pre-Haussmannian Paris plots map from the islets cadastre made by Philibert Vasserot between 1810 and 1836. On the other hand, we are setting up the first two historical information layers: the medieval and modern ones.

Since this tool is technologically complex and since it is intended to be a reference work for any further historical study about Paris, the working methods implemented to make the tool are part of the work itself. The way an interdisciplinary research is conducted determines the success of the research itself. The scope of reflection is wide: from a technical aspect – how to make a database – to a cognitive one – how to modelize a database in order to understand the history of an area – and through a human one – how to involve researchers from different horizons in a single program. As the reports of the meetings available on the LAMOP's website[5] show it, the project team has gradually formed what organization theorists call a "community of practice" which goes beyond a mere data infrastructure and a technical platform (Caron 1996; Pornon 2007). Our goal is not only to work with a GIS but to work together with a GIS as well.

[1] Laboratoire de Médiévistique Occidentale de Paris – UMR 8589, Université de Paris 1.
[2] Archéologie et Sciences de l'Antiquité – UMR 7041, Universités of Paris 1 and Paris 10.
[3] Littoral, environnement et sociétés – UMR 6250, Université de La Rochelle.
[4] Informatique image et interaction – EA 2118, Université de La Rochelle.
[5] Since September 2006, 11 meetings were held: http://LAMOP.univparis1.fr/LAMOP/LAMOP/LAMOPII.ALPAGE.htm.

The ALPAGE program, which was started in Autumn 2006, will not be completed until the Fall of 2009, but the team is already thinking about how to organize the further collaborative work, especially on how to access the GIS data, how to share them between researchers and other users and how to complement them. Often considered as trivial, the issue of the use of a tool by the researchers seems nevertheless to be decisive and intrinsically linked to the success of a GIS. That is why the setting-up of a platform for collaborative work can not be improvised. To efficiently fulfill the needs and expectations of the users, this reflection needs to be started at a very early stage.

2. The shaping of a historical GIS for Paris

This historically focused GIS for the city of Paris is principally meant to be a research tool that allows synchronic and diachronic spatial analysis.

From a synchronic point of view the GIS shall allow thematicians, mainly medievalists, to spatialize their historical data, for example, the tax households studied by Caroline Bourlet[6] from the tax rolls of the early 14th century. These thematicians will then be able to interrogate and manipulate their data to test spatial assumptions, which has been difficult up to now. The historical information layers can be topographic (buildings), territorial (seigniorial and parish constituencies) or social (tax households, paving expenditures, etc.). These layers will be built by different researchers on the basis of their own interests and without a defined program.

2.1 *The drawing of a reference plan for the late Middle Ages*

In order to locate the medieval historical data, we created a reference plan for the late Middle Ages (Fig. 1). Called ALPAGE 1380, this plan includes the contours of the street-blocks from the historical centre of Paris in 1380. It was formed using several documents from different sources and eras. The basic information (contours of the islets) comes from the CNRS map (Leuridan Mallet 1975), but their shape had to be adjusted to fit the plans by Vasserot which were georeferenced (by the laboratory LIENSS) in comparison to the current Parisian plots map provided by the APUR (Atelier Parisien d'Urbanisme).

2.2 *Creation of a plots map of the early 19th century*

The creation of the early 19th century plots map, from plans by Vasserot (Fig. 2), aims to allow diachronic spatial analysis.

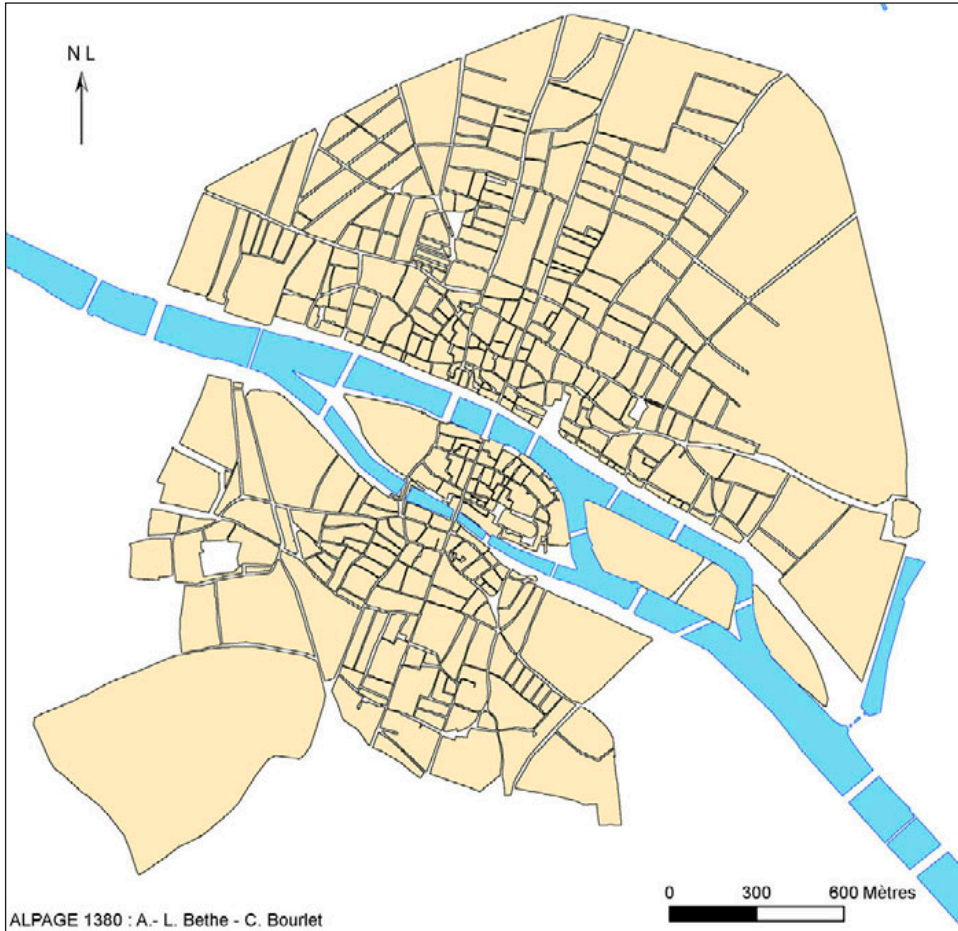[6] Institut de Recherche sur l'Histoire des Textes, CNRS.

Fig. 1 – Plan ALPAGE 1380.

The goal is to study the urban morphology in the long term, showing that there are, in addition to planning, other explanations, linked to social practices: the way people have lived in an area, their needs, what they did, partly explain the preservation of certain plots and parts of the street system.

How is the urban fabric formed and how does it evolve? What are, at different moments, its social uses? The aim is to observe the transmission of the urban forms through the constant mutation of the social use of the urban area. That is because societies are constantly reinvesting the inherited shapes, with different specific goals and contexts so that these urban shapes are updated and maintained in the long term. Indeed, once produced, urban
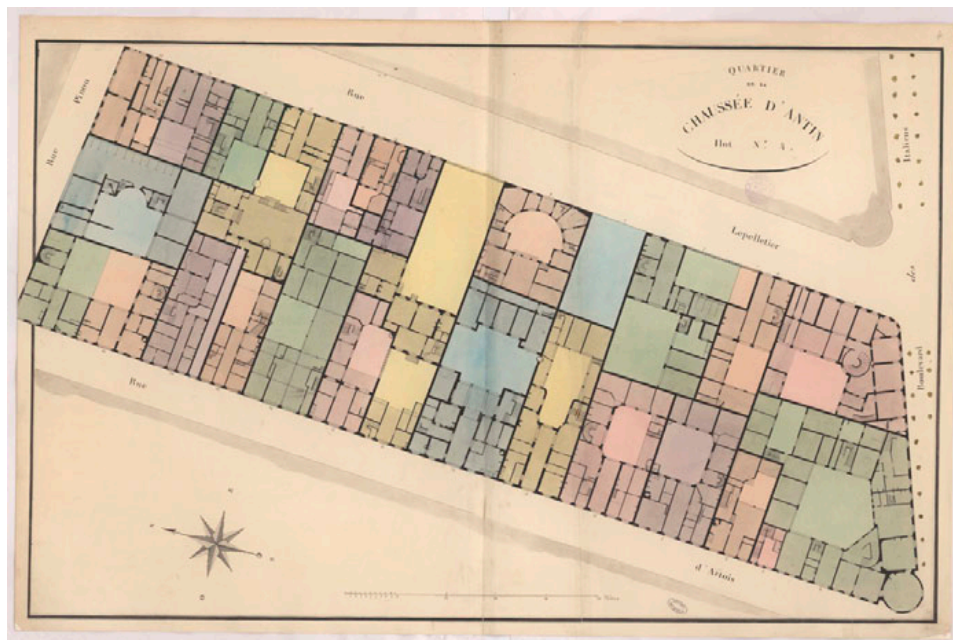
89

Fig. 2 – Plan by Vasserot: quarter of the Chaussée d'Antin, block No. 4.

forms (street-system, plots and buildings) have, by their materiality, a longer life than the social convention that made them emerge. The shape will be maintained beyond the social context that has brought it forth and will be reused later in a new social configuration which will have nothing to do with the previous ones.

This idea of multiple urban temporalities and the interest of confronting the short period of social conventions to the longer special forms, appears in the works of Marcel Roncayolo (1996, 2002), Bernard Lepetit (1993, 1995) or Gérard Chouquer (2000, 2003, 2007).

Studying the relationship between the structure and the uses of urban space requires to have, on the one hand, layers of historical information which spatialize social activity (broadly defined) and actors living and involved in this urban space at different times (see above) and, on the other hand, layers of vectorized plots, not just street-systems.

That is why the ALPAGE program is georeferencing and semi-automatically vectorizing the 912 islets from the Vasserot plans (Fig. 2). The vector layers of the plots thus made (polygons and segments) will allow to study precisely the morphological structure of the city inherited from the early 19th century: before Haussmann's upheavals. The spatial analysis tools will then be implemented: sizes

Fig. 3 – Advanced georeferencing of the Vasserot islets plans in June 2008.

and types of blocks, plots and streets, density calculations, orientation calculations, degree of openness of islets, degree of street-system connectivity, etc.

Between January 2007 and June 2008, 475 plans (i.e. 52%) were georeferenced by Anne-Laure Bethe[7] (Fig. 3). This georeferencing is conducted using as a reference the current plots map provided by the APUR with the help of documents of intermediate periods like the plans known as the Berty plans.

2.3 *The integration of historical information into a GIS and its use*

Once the reference plans of the historical centre of Paris from 1380 and the early 19th century were completed, historians were able to integrate historical data, spatializing them on the basis of information provided by these reference plans.

The chosen methodology is to use the skills of everyone to build a collective tool. The GIS also has to be built in order to enable common use by

[7] LIENSS, UMR 6250, CNRS – Université de La Rochelle.

91

people with different backgrounds, whatever their field (mainly historians and archaeologists, but also potentially architects, urban planners, etc.), the period studied (Middle Ages, modern or contemporary era, etc.), or the level of research. Indeed, the audience may be composed of experienced researchers or young researchers like master's and doctoral students, which will be able to question spatially their historical sources as soon as the reference layers exist. This search tool will then also be used as a teaching tool.

Beyond the research in history, this tool should also be used for development or for different knowledge purposes. Inasmuch as the Vasserot map is fitted on the current plots map conceded by the APUR, the former fabric can be concretely reflected in the studies carried out upstream of construction projects by professionals working in the Parisian region. We, historians, hope that actors in urban planning and development will actually refer to the historical information available.

Finally, when the issue of accessing the data is resolved, the GIS may also be used for general cultural purposes, through its possible use as a collection of georeferenced plans. For example, the exhibitions that are regularly carried out in any particular area of Paris and which always attract many people, need to fit the old plans into the new ones, which is currently done for each event unless it is integrated into a single tool. Moreover, our approach for the recovery of the old Parisian plots map had already been undertaken by architects of the School of Paris-la Villette in the 1980s (DARIN 1998, 66-67). In 1981, Bruno Fortier and Antoine Grumbach worked intensively with their students to assemble these same Vasserot islets plans, aiming to present the pre-Haussmann Paris at the Universal Exposition that was to be held in Paris in 1989. This project, like the Universal Exposition, failed.

As with any collective tool, the condition for the success of such projects (commonly called a "Rube Goldberg machine" or a "Heath Robinson contraption" when they are not efficient) lies in the ability to create a real dynamic of work. A collaborative GIS can not be a mere accumulation of data collected without consultation. We need to define the modalities for the administration of the GIS, in other words an organizational structure that allows different levels of production and use of the data and that validates who does what and how. We must think about the procedures of data access, use and edition by different users in order to set up a suitable technical platform (computers, software, networks).

3. Types of users and working conditions

Two main types of users of a GIS can be distinguished, consumers and producers of data, the needs of the latter being more complex to be met.

The consumers' needs can be summed up in a few classic features: inter-active display of the data with the selection of the desired layers, displacement and change of the scales, cross sorting by simple queries, export and display of the data in a common format to be used in different contexts or software. The web technologies of these features are known and widely practiced. Several solutions exist, whether free or proprietary. We have to accurately assess the needs and anticipate the consultation practices of these data. Thus some users, whether thematicians or general public, will immediately have specific requests, while others will first have an overview of the potential of the application before conducting a search. This means organizing the data in order to browse quickly the all the resources, among which a "tour" category should be included for those who visit the site without prior questions. Some visitors will only come once, while others will come several times, sometimes to view the same data.

For these users it may be interesting to predict that they have personal bookmarks in their web browser pointing to objects they regularly visit (for example, their street in Paris). There should be different possible scales of consultation to resolve the problem of generalization. It is necessary to devote time to those questions if we want the users to make the resources theirs.

As for the producers, in addition to the previous requirements, the major issue is the online edition, namely the creation via the Net of geometric and attributive data stored in a remote server. This kind of architecture requires to create unique objects, to capture the geometric and attributive data (hence the drawing), give the evidence and metadata. It also has to be able to georeference online plans that thematicians themselves will have collected through their disciplinary skills. This set of needs is less common and appears more difficult to meet than the previous ones, in particular to draw the georeferenced objects.

Moreover, these data producers (mostly historians and archaeologists), have often not been trained to GIS. They will therefore have to make a significant investment to learn how to use the tool. It is important that this investment should be minimal: users do not need to become geomaticians. The user interface should take this issue into account.

Although there are two types of data producers: the master's or doctoral student and the colleague researcher (archaeologist or historian), they face the same kind of constraint, which prevents most of them from working in a classic professional context (in the office) as most of the geomaticians do.

As for the students, more and more of them have to take jobs that take up the time that was previously devoted to their studies. Now many students work alongside their studies which often makes their schedules incompatible with those of an institutional structure like a laboratory (closed in the evening, on Saturdays, etc.). It is useless to ignore the socio-economic reality, especially since there are fewer and fewer students, at least in French universities.

The researchers are, by the nature of their work, independent and manage their schedules in an autonomous way. They have to go to the library, the archives, on excavation sites, etc. It is therefore difficult to compel them to office hours to feed a GIS, which will substantially increase the risk of failure of the program. In addition to these specific professional practices, there is the constraint of the time that is lost in transport, which is very important in the region of Paris: the time spent in transport reduces greatly the travel ability. All these practical realities are not trivial and determine the success of a program.

However, the objective of the ALPAGE program is to create a work dynamic in the long term, which is the only way to enrich and thus perpetuate the tool. Databases are useful only if they are perpetually transformed by users. Indeed, what is the use of an extremely complex GIS with a very elaborate CDM (Conceptual Data Model) if it is not filled by thematicians?

It is therefore necessary to establish a collaborative platform that takes into account the needs, skills and constraints of the users.

4. Ways of producing data

4.1 *From offline to online work*

The most classic and most technically simple way of feeding a GIS is to work offline on a dedicated computer. The thematician defines his contribution with the administrator, accumulates the necessary sources in libraries or archives for the production of his data and then creates his objects in the GIS, in a laboratory with such a computer, in this case ARSCAN or LAMOP. This type of local organization meets the need of centralizing the data production within a GIS. However, the reasons that will prevent Parisian thematicians from using such a platform were explained above. At the outset, this solution will obviously not help to develop the use of a GIS among thematicians. Nevertheless, it is still possible to use this kind of platform, particularly as a temporary one, provided laboratories have the adequate equipment (and thus the structural funding), which is far from being the case.

Nevertheless, the way forward is the webmapping, which provides a centralised management of data production, while allowing to work anywhere. After the objects definition phase with the administrator of the GIS and the collection of the sources, the phase of the definition of objects in the GIS could be done directly online in a server/clients relationship, which solves a large part of the problems mentioned above. However, this type of operation creates other needs than those of conventional micro-computers. This includes persons employed on a permanent basis to maintain the tool technologically, which appears to be more complicated to achieve than on a dedicated computer.
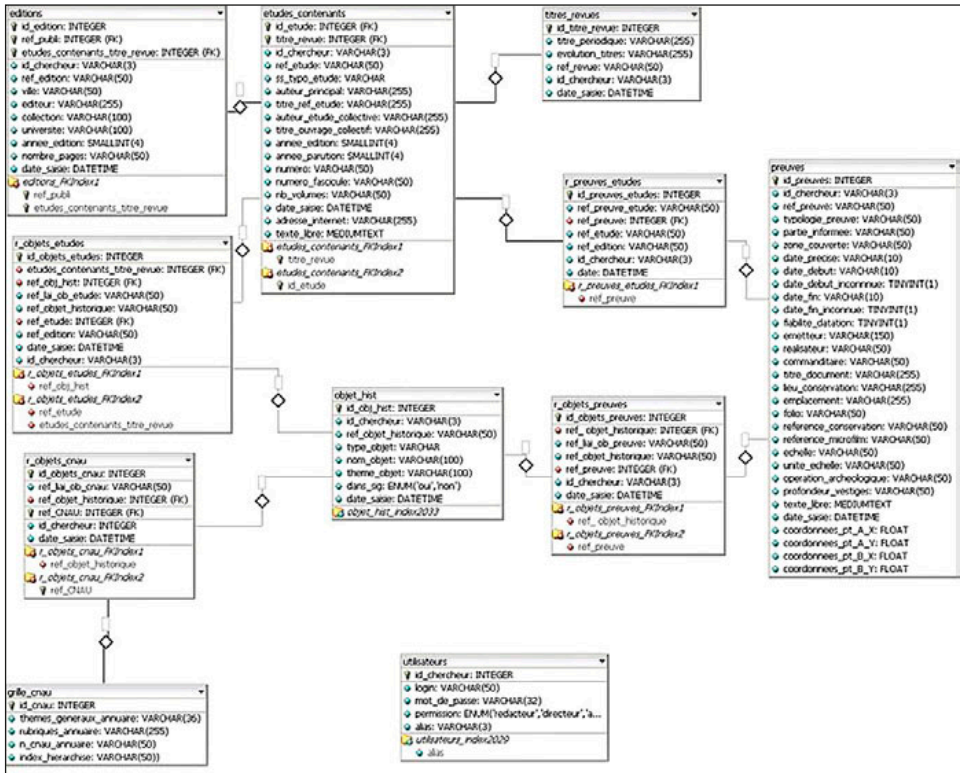
Fig. 4 – Conceptual model data ALPAGE-References.

Prior to the establishment of an online solution to meet the needs of users and contributors, ALPAGE project implemented a temporary solution that can be presented as being halfway between these two.

On the one hand, the working process is indeed based on a centralized management of the objects identifiers via the web (to ensure the uniqueness of IDs) and on the other hand on the independent definition of the objects by thematicians on their own computer using the same reference map for all of them. Once the objects are defined, it will be necessary to copy them on the server which will allow a consolidation of the information and therefore centralized management.

For the first phase of the work, the DBMS ALPAGE-References, mainly built by Alain Dallo (Fig. 4), has been adapted and posted on the website of the LAMOP by Georges-Xavier Blary (Fig. 5).
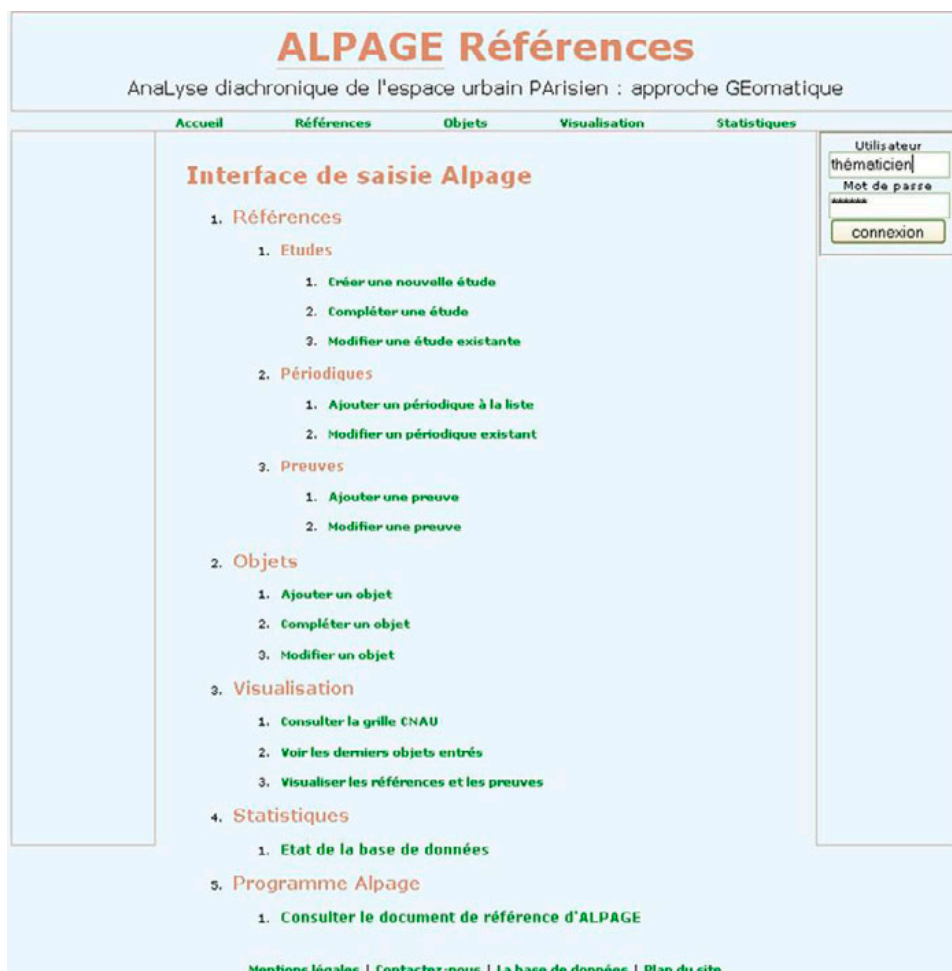
Fig. 5 – Web interface of DBMS ALPAGE-References.

4.2 *Structuring the DBMS References*

The thematicians build their historical layers, combining geometric objects and attributive data.

Each object must have a unique identifier, stored in the objects table. The existence of objects is established through one or more evidence pieces, to draw its contour (table of evidences).

These evidences can be a manuscript in archives, a plan of archaeological excavations, a monograph devoted to that object or only part of a publication

(e.g. a few pages). Thus, as a single study may contain much evidence for different objects (Berty *et al*. 1866-1897), a table of studies separate from that of the evidence has been created so as not to fill the database with multiple occurrences of the same reference. In this case, the proof is precisely located within this study in the "evidence" table (pages of a book for example). Finally, a study can have multiple editions, which induces an "editions" table.

Concerning the relations between these tables, there are often not one but several evidences for a single object. In addition, the same evidence may be used for several objects (such as a plan). Finally, a study may contain several types of evidence. So, links between the "objects", "evidence" and "studies" tables must be established.

To indicate the urban function of topographical historical objects, the table of classifications of the CNAU (Centre National pour l'Archéologie Urbaine) which is indirectly related to objects (n to n relationship) is used. Finally, a table of titles of magazines is made to avoid data entry errors.

This rather complex database, linking objects, evidence and studies, was initially developed under Access. Each table includes many fields in order to specify different types of evidence (handwritten or edited text, archaeological evidence, iconographic evidence, planimetric proof, etc.). Thus, according to each type of evidence, only certain fields have to be filled. The forms being very complex, the idea was then to move quickly to an online database. Each user, depending on the type of his source, is then proposed an appropriate form.

4.3 *DBMS References online*

The difficulties linked to the choice of working on the spot were already numerous despite the relatively low number of thematicians participating in the program. It was therefore decided to lift some of these constraints by migrating the "References" database from an off-line version under Access to a MySQL one allowing the data capture and search online.

The choice of MySQL was justified by its availability on the server of the LAMOP and by the fact that this is one of the most widely used database management systems on the Internet at the moment. Moreover, it can be installed relatively easily on computers without connection to the Internet which allows the off-line use of interfaces primarily designed for the web.

The context of the use of "References" requires to pay special attention to its ergonomics. The creation of objects requires many complex manipulations like finding or creating evidence and studies. But the multiplication of manoeuvers significantly increases the risk of wrong data entry and selection errors. Although they can be promptly corrected by the administrator of "References", it seemed unwise to resort systematically to this solution. But insofar as the number of contributors is probable to grow, corrections could monopolize much time. That is why the interface was designed to prevent

the most common errors that could be made by using the possibilities of each technology, which allow the control of the validity of the data at three levels: database server (MySQL), web server (PHP), client (Javascript).

First, particular attention has been given to the choice of type[8] and length and the possibility of redundancy or lack of data in MySQL (Fig. 4). Thus, when information is not exactly the type and length expected, the transmitted data are not recorded in the database. For example, it is not possible to enter string data in a numeric field. Moreover, if the set of data to be recorded is identical to a previous registration or if it contains an empty field, it will not be registered, thereby avoiding duplication or partial information.

PHP, the language used to link forms and the database, can check more precisely the data than with the DBMS before sending them to MySQL, including checking the validity of a set of data according to other data. For example, when entering a new study, if the thematician states that he wants to record a thesis, it is possible to verify that the "University" field has been filled.

Javascript allows to make similar checks on the client computer. Although this solution has the advantage of reducing the burden on servers and thus increase the speed of the interface, it has not been selected for security reasons. Indeed, it is possible to disable Javascript on client computers and thus circumvent the audit. However, this language has proved useful for the entry box.

Entering text data implies many potential errors, including typographical errors, which depreciate the database by increasing its rate of silence (because the recorded data do not correspond to those searched). These faults are very difficult to control with the methods outlined above. That is why it seemed appropriate to filter a maximum of the data during the entering in order to minimize the risk of errors. In order to do this, the AJAX method based on Javascript was used. This technique allows to query a database asynchronously, thereby leading to its end the necessity of reloading a page to obtain new information sent by the server. It is therefore possible to suggest that specific data be included in a field based on data from other fields while information is being entered. This method limits the errors by showing only the relevant fields and information and by standardizing the data entered in the database.

In addition to increased ergonomics, the advantage of the online version of "References" is to pool the work of the thematicians. The previous version, offline on Access, was necessarily difficult to update. In order for each thematician to have the latest updated data at his disposal, all the files locally

---

[8] MySQL offers a thirty types of data (MySQL 5.0 Reference, chap. 10: http://dev.mysql.com/doc/refman/5.0/en/data-types.html).

produced had to be merged. This operation, already time-wasting, requires paying great attention to the IDs of each record. Without the establishment of a technique ensuring the uniqueness of the recordings, some would be crushed. However, avoiding the loss of recordings can prevent from protecting oneself against duplication. So, doubles would then have to be treated retrospectively. Finally, the last drawback of working offline is having to enter information that may already have been entered, which represents a waste of time.

The web version of "References" seeks to remedy these defects. Indeed, the centralized management of data allows to share instantly all the records and guarantees the work on the most recent data because the data source is common to all thematicians and is updated in real time. The programming languages for Internet allow greater latitude in controlling and presenting the information.

This interface allows, in the context of feeding a GIS offline, to produce as quickly and easily as possible a truly unique identifier to ensure data integrity for the next phases of work. This documentary database is the only one available online. The attributive, historical and heuristic databases, specific to each researcher, are currently only developed locally on the computers of the researchers. Thus, each thematician, who has been conferred access codes, connects via the web, creates the identifier of the object and indicates the evidence that he used to create this object. Then the second phase of the work starts. Offline, each thematician creates the object in a GIS software installed on his machine. He draws the object and therefore provides the attributive data, to which he integrates the ID produced online. As for now, this solution works well, because there are not many thematicians.

Indeed, this procedure requires to implement on each computer tools (a software to connect to an ftp server, a GIS software, possibly a DBMS software) and data (current plots maps, Vasserot's raster images). The technical question is therefore deferred to the user: it is up to him to have a powerful computer, to install the data and software (GV-SIG mainly because it is free and demands few computer resources). But this assumption is virtually unobtainable: it is necessary to help the thematicians with the use of the tools, and this has a cost both in time (a need to re-explain the same procedures, each moving at his own pace) and financially (a participation in the purchase of a laptop and external hard drives). Another drawback concerns the issue of data dissemination: some of them are derived from trade with external partners (urban planning agency) and are copied on personal computers. How can we control the copy and the spreading of the data since it is on several computers?

It is therefore clear that this solution, though it works at the present time, is a temporary solution. It will be much less convenient if the number of researchers increases significantly, which is our goal. In the perspective of a work dynamic where many people contribute to enrich the GIS, it seems

that the solution of webmapping adapts best to our needs. We have to take these practices into account, which is why this symposium was organized as these methodological issues arise without doubt everywhere. If the problem of the data access is recurrent, there is however no single or ideal solution since each time we have to take into account the human, technical and geographic contexts specific to each research program.

Hélène Noizet, Alain Dallo, Georges-Xavier Blary
UMR 8589 – LAMOP LAboratoire de Médiévistique Occidentale de Paris
Université de Paris 1 Panthéon-Sorbonne

Laurent Costa
CNRS – UMR 7041 ArScAn Archéologies et Sciences de l'Antiquité

Frédéric Pouget
UMR 6250 – LIENSS
Université de La Rochelle

REFERENCES

Berty A., Legrand H., Tisserand L.-M., Vacquer T., Platon C. 1866-1897, *Topographie historique du Vieux Paris*, 6, Coll. Histoire Générale de Paris, Paris.

Caron C. 1996, *Cadre descriptif des projets d'implantation de technologies géomatiques dans les organisations*, Thèse de Doctorat, CRG – Université de Laval (http://www.bibl.ulaval.ca/).

Chouquer G. 2000, *L'étude des paysages. Essais sur leurs formes et leur histoire*, Paris, Errance.

Chouquer G. 2003, *Crise et recomposition des objets: les enjeux de l'archéogéographie. Introduction*, «Études rurales», 167-168, 13-31.

Chouquer G. 2007, *Quels scénarios pour l'histoire du paysage? Orientations de recherche pour l'archéogéographie*, Coimbra-Porto.

Darin M. 1998, *The study or urban form in France*, «Urban Morphology», 2, 2, 63-76.

Lepetit B., Pumain D. (eds.) 1993, *Temporalités urbaines*, Paris, Anthropos Economica.

Lepetit B. (ed.) 1995, *Les formes de l'expérience. Une autre histoire sociale*, Paris, Albin Michel.

Leuridan J., Mallet J.-A. 1975, *Plan restitué de Paris en 1380*, Paris, CNRS (2ᵉ éd. 1991, 3ᵉ éd. 1999).

Pornon H. 2007, *Bilan et perspectives de 20 années de géomatique*, «Géomatique expert», 57, 36-60.

Roncayolo M. 1996, *Les grammaires d'une ville. Essai sur la genèse des structures urbaines à Marseille*, Paris, Ehess.

Roncayolo M. 2002, *Lectures de villes. Formes et temps*, Marseille, Parenthèses.

ABSTRACT

Alpage program, based on a collaboration between historians, geographers and computer researchers, aims to build a historical GIS of Paris. First, we reconstructed the pre-Haussmannian plan of Paris by georeferencing and vectorizing the survey of the cadastre made by

P. Vasserot (1810-1836). Then, on these fundamental layers, historical information layers, like medieval and modern ones, will be built by different researchers according to their specific interests (churches, town walls, fiefs, parishes, etc.).

Since this tool is technologically complex and since it is intended to be a reference work for further historical studies about Paris, we must immediately take into consideration the organization of further collaborative work. Accessing the GIS data, both to share them with different types of users and to edit new data in it, is an essential question, although it is often considered a trivial one. The issue of the use of a tool by the researchers seems nevertheless to be decisive and intrinsically linked to the success of a GIS. Meeting the needs and expectations of the users, webmapping might be a good solution for editing the geometric and attribute data in a GIS. But today, setting up this kind of platform for collaborative work is still difficult and time consuming. That is why, for Alpage, a temporary solution was found, revealing what organization theorists call a "community of practice": it combines a centralized management of the references and object identifiers via the web and an independent edition of the objects by thematicians on their own computers. For this reason the DBMS ALPAGE-References has been adapted and posted on the website of the LAMOP.