ArSol: An Archaeological Data Processing System

Philippe HUSI – Xavier RODIER

CNRS, Université de Tours, Laboratoire Archéologie et Territoires CITERES (UMR 6173), Tours, France
{philippe.husi; xavier.rodier}@univ-tours.fr

Abstract

The ArSol (*Archives du Sol*: Soil Archives) system has been used by the "Archéologie et Territoires" Laboratory (CNRS – Tours University) since 1990 for processing archaeological data. It can be used for all stratigraphic excavations and has the dual purpose of data management and research. It was constructed as an open system that is flexible and above all not conditioned by the integration of predefined thesauri. The ArSol client-server system is designed to integrate data from different sites. It comprises two main components:

1) The *Archives de fouilles* module, based on the stratigraphic recording system used since 1970, concerns stratigraphic data and artefacts for site analysis. It offers tools for data management, location and spatial analysis;

2) The BaDoC module, a pottery database, which can process the data according to different quantitative techniques regardless of the diversity of the typological assemblages and the scale of archaeological contexts. This database can be used to manage large quantities of pottery artefacts, but it has been developed primarily for research in this field. Results of database queries are automatically exported onto a spreadsheet in the form of a data matrix, including all the information (positive or null values), a prerequisite for all statistical data modelling. This database allows the integration of external information, such as the results of statistical dating models, and exports other information such as dating of archaeological contexts from pottery.

The location of data stored in both the *Archives de fouilles* and *BaDoC* modules is based on a chain of spatial information acquisition and processing, linking stratigraphic data and digital surveys in a Geographical Information System. The GIS is based on a link between 4th Dimension ™ and ArcGIS ™ software through an ODBC (Open Database Connectivity) interface.

For excavation analysis, for example, ArSol enables all the information from cartographic data to be consulted, plans to be established automatically according to the chronology included in the stratigraphic database, or the distribution of artefacts to be mapped.

ArSol is designed both as a recording and data management research tool for use during excavation, and as an exploratory analysis data system for post-excavation work. Finally, the system makes it possible to switch rapidly from intra- to inter-site scales.

Keywords

site records, stratigraphy, pottery, processing data, spatialisation

1. Introduction

The *ArSol* (Soil Archives) system has been developed by the *Archéologie et Territoires* Laboratory to process the data of the excavations carried out by its members. It is also used for other excavations and by other operators, such as INRAP or regional authority departments. It consists of two modules: *Archives de fouilles* for stratigraphy, and *BaDoc* for pottery.

The *Archives de fouilles* module is used for managing and processing stratigraphic data and artefacts with a view to analysing the excavated sites. This module was developed in 1990 for the excavation of the medieval parish centre of Rigny (Indre-et-Loire) (Zadora-Rio 1994) and then adapted for other sites, in particular for excavations in the city of Tours. The *Archives de fouilles* module is based on the manual stratigraphic recording system used since 1969 for excavations in Tours (Galinié 1980; Galinié and Randoin 1987), and which was itself based on work carried out in Winchester at the end of the sixties by Martin Biddle (Biddle 1969; Barker 1969, 1977; Harris 1975, 1979). It was constructed as an open system, i.e. it is flexible and above all is not constrained by an integrated thesaurus, and can be used for all stratigraphic excavations. Since 1998 it has gradually been incorporated into a graphic information processing chain developed by Xavier Rodier which links digital site surveys with stratigraphic data in a geographic information system. This system allows

both the management and post-excavation processing of site data (stratigraphic and spatial analysis), and also the incorporation of data which are useful on a small scale into other information systems. For example, data from the Tours excavations have been incorporated into ToToPi, a Geographic Information System for studying the historical topography of pre-industrial Tours (Galinié *et al.* 2004; Rodier and Saligny 2007; Rodier and Saligny 2008). At the excavation scale for example, all the information can be consulted from cartographic data, plans can be drawn up automatically from the chronological periods included in the stratigraphic base, and the distribution of artefacts can be mapped.

BaDoC, (*Base de Données Céramique,* Pottery Data Base), first developed by Philippe Husi in 1991 for pottery in the city of Tours, is a system for processing the data from all stratigraphic excavations, whatever the typological facies and chronology of the site. The aim of this system is pottery management, but above all research, since the main developments concern typology, quantification and dating on the scale of the site or the town, or more broadly on the scale of the chosen stratigraphic analysis. The interface of this base has been developed in such a way that archaeologist who do not have very good computer skills can learn to use it quickly. This tool has been used for all the ceramic studies of the *Archéologie et Territoire* Laboratory, including those forming part of university degrees (master's, doctoral theses) (Husi 2000; Husi (dir.) 2003; Bellanger *et al.* 2006a, 2006b; Bellanger *et al.* 2008).

The choice of software was based on commercial programmes used since the system was first developed in 1990. Using free software was not then an option, particularly in archaeology. The software used is 4th Dimension for DBMS and ArcGis for GIS, linked by an ODBC (Open DataBase Connectivity) interface. While ArcGis is widely known, 4th Dimension is less so. It is a robust DBMS which can manage complex structures and has an application development language. It provides several access levels for different user skills: personalized interface, advanced use, programming/development. It enables a client/server and multi-platform application to be used and can be accessed via the internet.

2. *Archives de fouilles*, recording site data

The procedure used involves a dual approach to the archaeological interpretation of an excavation. The first is inductive and is based on processing stratigraphic data following a process of successive groupings (functional, spatio-functional, spatio-temporal) of stratigraphic units, atoms of information, while the second approach is deductive, looking at a ranked periodization specific to the excavated site, based on hypotheses of how phases of site occupation fit into a historical model (*Fig. 1*). The archaeological interpretation is always at the interface of these two simultaneous approaches. This procedure on the one hand ensures the robustness of the analysis data by using information systems, and on the other hand it avoids the illusion that the input of a broad range of very precise data will produce the precise answer to the question asked. That arises from confusion between a purely descriptive, supposedly objective and exhaustive recording system, and an information system designed to deal with a particular issue.

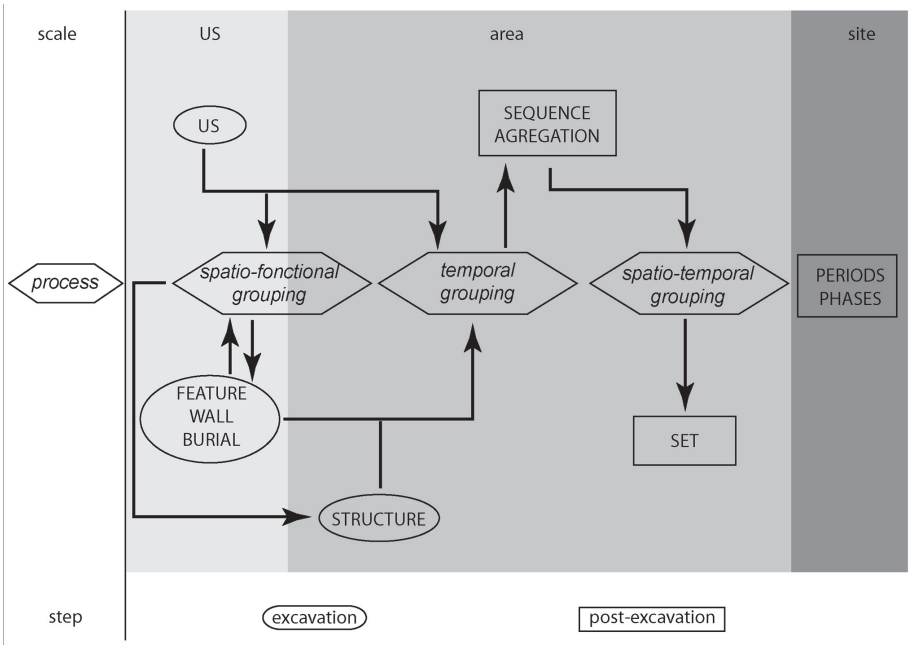

Fig. 1. Flowchart showing processing of archaeological data.

The method of recording and processing stratigraphic data is based on a two-step process of successively

grouping entities, first on site and then during the post-excavation study. The smallest recording unit on the site is the *stratigraphic unit* (US). During the excavation, the US can be grouped by *feature* (pit, post hole, cooking pit, floor, trench, etc.), and/or by *structure* (construction, room, building, etc.). Two specific series can be distinguished for walls and burial places which are unequivocally identifiable features. This initial grouping procedure is spatio-functional. The post-excavation study starts by drawing up a stratigraphic diagram to show the relative chronology between layers, i.e. the order in which they were deposited. After that, two grouping levels are possible, *sequence* or *aggregation*, depending on the complexity of the stratigraphy. This involves a temporal grouping of US in relation to a single event (construction of a wall, use of a cooking pit, occupation of a room, etc.). A third spatio-temporal grouping into sets occurs with the incorporation of dating elements.

The periods and their subdivision into phases correspond to a chronological division which cannot be inferred from the successive groupings.

To match this process, a computer system must meet the following objectives:

1. manage the overall data and documentation
2. authorize the reiteration of recordings in the data base during the post-excavation study
3. provide a level of comparison of the inter-site data, while allowing the specifics of a site to be characterized.

The *Archives de fouilles* module does of course deal with information management issues: recording stratigraphic data and associated artefacts, indexing the documentation (photos, surveys, files, etc.), report writing. It includes a module to help produce stratigraphic diagrams, process stratigraphic relationships, check the coherence of the recorded reports and errors, as well as produce a diagram of the relations. This involves integrating the "Stratifiant" algorithm designed by Bruno Desachy (2005).

The *Archives de fouilles* design looks at the process of analyzing data which involves reiterating the data base during the post-excavation study. At each grouping stage (sequence, aggregation, set), this new information must in fact be assigned down to the smallest recording level, the US. For example, with the proposed input method, when inputting an *aggregation* it is possible to select the *sequences* which it contains from a list in which only those which are not already assigned to an *aggregation* appear. The *aggregation* number is automatically assigned to each of the selected *sequences*, as well as to the US of which it is composed.

Another original aspect of the Excavation archive is its aim to preserve an open and adaptable system. Cataloguing archaeological entities (artefacts, features, structures, etc.) is usually based on one or more thesauri developed at length with a view to exhaustiveness at the risk of pre-interpreting the data. Without wanting to put forward completely open fields offering no possibility of queries or comparison, we have opted for an intermediary solution consisting in distributing information under several headings. Taking artefacts, for example, four fields are useful for interpreting objects:

1. free **description**
2. **identification** selected from a list which can be expanded at will
3. a cataloguing **function** in a closed list of 13 general categories
4. **usage,** selected from a list which can be expanded at will in order to develop a classification specific to the nature of the site.

In this way, interpreting the usage of the object in its context is distinguished from its theoretical cataloguing function. For example, an object identified as a ring will be catalogued as a "personal household object", but its usage will differ according to whether it was discovered in a funerary or production context. This way of structuring information, valid for the whole system, means that the software tool will not replace reasoning in the chain of inference when interpreting excavation data.

3. BaDoC, a tool for studying large quantities of pottery

Two levels for studying pottery can be envisaged. Firstly, a rapid and systematic examination of the material which is useful for the archaeologist during the excavation, from which the potential of the study can be ascertained through selection of the corpus to be processed. This phase is carried out in the Archives de fouilles module, via a specific table, containing only a few general descriptors per US as follows: the most relevant typological pottery elements, the approximate dating of the US, the vessels to be drawn, the overall weight of the pottery for logistical

purposes and the research potential of the collected pottery. The BaDoc module corresponds to the second level, with a detailed study of the pottery, involving all the typological and functional descriptors as well as a thorough quantitative processing of the artefact. The use of such a tool is only of interest when applied to a reasoned corpus, and thus selected after the examination of the previous phase.

This data base (BaDoc) is used to study large quantities of pottery artefacts. It has been developed for research in this field, rather than just for the management of data. The proposed system is open for two reasons: 1) because it can be adapted for all types of artefacts from stratified excavations (it is not linked to a predefined thesaurus); 2) because external information can readily be incorporated (in Tours, the results of a statistical pottery dating model) and other information exported (dating of the archaeological contexts).

Although no thesaurus has been pre-defined, for the BaDoc to function properly it must be linked to a typological reference system. For pottery from the 6th to 17th centuries, and the Centre-West region of France, catalogues of vessel shapes and pottery fabrics have been drawn up actively and collectively on an internet site called ICERAMM (an acronym in French for information about medieval and modern pottery) (http://iceramm.univ-tours.fr/). One objective for the future is for this internet site, involving most researchers in the field at a national level, to become the tool holding all the regional typological reference systems for the period.

Rather than give a detailed presentation of the BaDoc (*Fig. 2*) it seems more useful here to focus on the innovative advantages of the system.

3.1. Quantification

The system allows the pottery to be quantified by *Sherd count* or *Weight,* or in more complex ways such as *Estimated vessel equivalents (Eves), Minimum vessel count (MINVC),* or *typology vessel count*, depending on the archaeological analysis selected. The outcome is a matrix of data incorporating the missing data, resized according to the query and automatically created in a spreadsheet (Plug-in 4D View). Since each new typological element is integrated into the database, even if it is not in the selection in question, the creation of automatically resized matrices ensures a dynamic structuring of the quantified data. These data, created at the time of the query and therefore easy to reiterate, are essential for statistically processing a large volume of data.

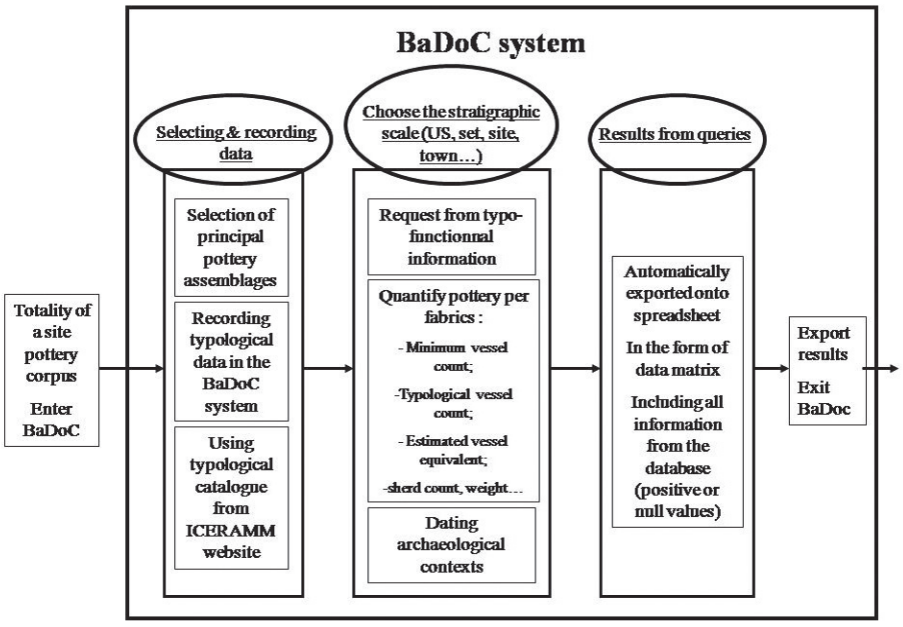

Fig. 2. The BaDoC system.

1) Vessel count in 2 separate levels of a midden			
US of midden	**Rim**	**Base**	**MINVC**
US 1	**3**	1	3 MINVC
US 2	1	**2**	2 MINVC
Total			**5 MINVC**
2) Vessel count in the whole of the midden			
The whole of the midden	**Rim**	**Base**	MINVC
Total	**4**	3	**4 MINVC**

Fig. 3. Simple example showing two ways of quantifying the pottery (MINVC) in a single archaeological context.

Unlike many other systems, the *Minimum vessel count (MINVC)* is not fixed, calculated in a field, when the archaeologist studies the material. The only information which is input into the BaDoc is that which will allow it to be calculated. The difference is enormous, because in the

contrary situation, quantification does not really take account of stratigraphic scale changes. In the BaDoc, the MINVC is reconstructed at every query. Thus, taking the example of pottery in a midden, the overall calculation of MINVC in the pit as a whole will be different from, and not the sum of, the count for each stratigraphic unit of which it is composed: it is calculated from primary typological information contained in the database. This difference will affect the results, as it gives real amplitude to the data, depending on the analysis scale chosen (*Fig. 3*).

3.2. Dating

It was decided not to include automatic dating in this system as it is risky, the archaeologist having no real control. Instead, the approach involved is working towards helping decision-making through dating indicators. In this way, the user automatically obtains the following information from queries made at a particular stratigraphic level (US, Set, etc.) (*Fig. 4*):

1. The total Minimum vessel count (MINVC); the quantitatively best represented pottery fabric.
2. The so-called "reference" chronological bracket during which the majority of fabrics are likely to be contemporary (between 1275–1450).
3. The Minimum vessel count of each fabric, whose chronological bracket is earlier than and separate from the so-called "reference" bracket, and which corresponds to residual material (750-925).
4. Finally, the Minimum Vessel count of the few fabrics whose chronological bracket is later than and separate from the so-called "reference" bracket, with two possible interpretations depending on the nature of the context and the quantity of the pottery: (i) these fabrics date the archaeological context, (ii) they represent the proportion of intrusive material (1600-1650).

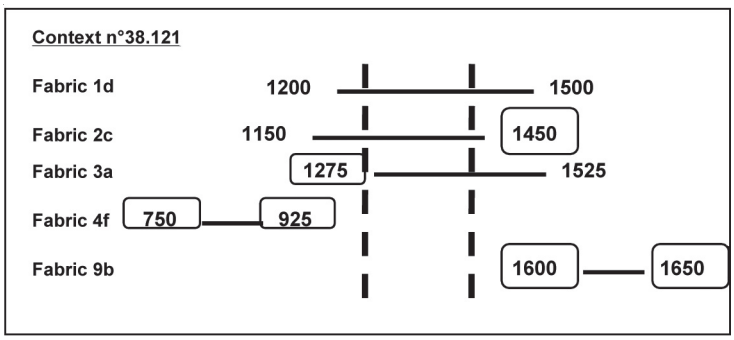

Fig. 4. Chronological brackets of the 5 fabrics making up the pottery assemblage of a fictional archaeological context.

It is not a question here of explaining how the chronological brackets of the pottery fabrics are established (simple typological comparison or, as in Tours, drawing up statistical dating models), but rather of emphasizing the form taken by the information in the system (Bellanger L. *et al.* 2008). For each fabric, one field is given with the lower chronological limit, and the other with the upper chronological limit, representing the probable life-span of the pottery fabric (emergence, predominance, disappearance). It is by bringing together the various chronological details and archaeological interpretations that the context can be dated.

3.3. The socio-functional interpretation

As for dating, the system only provides quantified information liable to assist the user make archaeological interpretations. Apart from the *Minimum vessel count* and the *typology vessel count*, the system automatically generates functional tables summarizing quantified information at a specific stratigraphic level. In this way, for each type of recipient it is possible to see the number of items and their theoretical function (tableware, cooking, etc.), and the number which show traces of fire indicating their culinary use. Summarizing information in this way avoids long database manipulations which too often lead to errors.

3.4. An open and accessible system in which the user is in control

Navigating the BaDoc follows the protocol of a pottery study: primary data input which is as limited as possible, with maximum automatic processing of the most repetitive tasks; data preparation with a view to their quantitative analysis; functional chronology, with a large proportion of pre-programmed actions; easy consultation of the data base in the form of interfaced queries.

Finally, we need to go back to two aspects which I consider to be important. First, the fact that this system can be used on all sites involving stratigraphy, whatever the period or form of pottery. The system can easily be adapted by modifying the content of the fields, which affects neither the structure of the database, nor the calculation procedures. Secondly, no programming is involved in archaeological decision-making: only the

calculations are automated. In other words, the user is always in control of the decisions regarding the archaeological interpretation with no risk of being subject to the system. Apart from learning to use the tool, the user must take decisions based on experience, not only of the pottery analysis protocol but also of processing large quantities of material from stratified excavations.

4. Data spatialisation

In *ArSol*, the spatialisation of descriptive data recorded in the *Archives de fouilles* and *BaDoc* modules is based on a chain of spatial information acquisition and processing operations in a Geographic Information System.

The 1/20 site survey on paper or tracing paper has been replaced by geo-referenced digital photography. This involves taking a digital photo of the element to be surveyed (US, feature, wall, burial, structure) as vertical as possible, having first placed four control point on the same horizontal plane. The x, y, z and id (identifier of the surveyed element) coordinates of the four landmarks are read by tacheometer. A fifth coordinate, or more if necessary, in the centre of the structure will enable an altitude to be recorded in or on the element.

All the following stages are carried out in ArcView™. By transferring the landmark coordinate files, points can be created to geo-reference and then rectify the photos. The contour of the surveyed elements is vectorized and an identifier provides a link with the descriptive data of the *Archives de fouilles* module.

This change of practice gains time, but also entails profound modifications to ways of working whose implications must be analyzed and acknowledged. On site, these can be seen on two levels: work organization and the methods used, and the way data is interpreted. For example, printing photos is essential on the site in order to be able to note on them all the information required for interpretation. Likewise, their definition and geo-referencing frees the archaeologist from stone-by-stone surveys.

This data organization can be used in three ways:

1. consultation of the soil archives as a whole from spatial objects and the possibility of spatial queries (location, proximity, distance, etc.)
2. production of thematic plans based on all the data recorded in ArSol (by structure, phase, artefact distribution, etc.) using automatic page layout models.
3. access to calculation tools based on the spatial properties of the objects themselves (location, boundaries, surfaces, topology, distance, etc.), aiding the exploratory analysis of the data.

5. Outlook

In conclusion, we would like to look at four ways that *ArSol* might be developed:

1. Structuring, allowing non-stratified data to be recorded.
2. User interface, for which we hope to improve the input-analysis-consultation ergonomy.
3. Normalization, dealing with the question of storing, archiving and inter-operability.
4. Development to allow *ArSol* to be used via internet.

The objective is to offer a system meeting the demands of the process of interpreting archaeological data, taking account of what computerisation can bring to the process itself, in terms of data manipulation for example.

References

Barker, Philip (1969). Some aspects of the excavation of public buildings. *World Archaeology* 1, 220–230.

Barker, Philip (1977). *Techniques of archaeological excavations*. Londres.

Bellanger, Lise, Philippe Husi and Richard Tomassone (2006a). Statistical aspects of pottery quantification for dating some archaeological contexts. *Archaeometry* 48, 169–183. Oxford.

Bellanger, Lise, Philippe Husi and Richard Tomassone (2006b). Une approche statistique pour la datation de contextes archéologiques. *Revue de Statistique Appliquée*. LIV (2), 65–81.

Bellanger, Lise, Richard Tomassone and Philippe Husi (2008). A statistical approach for dating archaeological contexts. *Journal of Data Science* (JDS) 6/2. Revue en ligne, http://www.sinica.edu.tw/~jds/Content-v-6-2.html

Biddle, Martin and Birthe Kjolbye-Biddle (1969). Metres, areas and robbing. *World Archaeology* 1, 208–219.

Desachy, Bruno (2005). Du temps ordonné au temps qualifié : application d'outils mathématiques au modèle d'analyse stratigraphique d'Edward Harris. *Bulletin de la société préhistorique française* 102/4, 279–740.

Galinié, Henri (1980). De la stratigraphie à la chronologie. In: A. Schnapp (dir.) *L'Archéologie aujourd'hui*. Paris. 63–86.

Galinié, Henri and Bernard Randoin (1987). Enregistrement et exploitation des données des fouilles de Tours. Actes de la table ronde. *Enregistrement des données de fouilles urbaines*. Tours: Centre National d'Archéologie Urbaine, 61–73.

Galinié, Henri, Philippe Husi, Xavier Rodier, Christian Theureau and Elisabeth Zadora-Rio (2005). ARSOL. La chaîne de gestion des données de fouilles du Laboratoire Archéologie et Territoires. *Les petits cahiers d'Anatole* 17. 27/05/2005. 36.772 signes. http://www.univ-tours.fr/lat/pdf/F2_17.pdf

Galinié, Henri, Xavier Rodier and Laure Saligny (2004). Entité fonctionnelles, entités spatiale et dynamique urbaine dans la longue durée. *Histoire & Mesure* XIX-3/4, 223–242. Paris: CNRS. http://histoiremesure.revues.org/document761.html

Harris, Edward (1975). The stratigraphic sequence: a question of time. *World Arcaheology* 7, 110–121.

Harris, Edward (1979). Principles of archaeological stratigraphy. London: Academic Press.

Husi, Philippe (2000). Quantification et datation en céramologie (le nombre minimum d'individus : la technique de quantification la mieux adaptée à la datation des contextes archéologiques à partir de l'exemple de Tours. *Les petits cahiers d'Anatole* 6. 2 ill. 65.800 signes. http://www.univ-tours.fr/lat/Pages/F2_6.html

Husi, Philippe (dir.) (2003). La céramique médiévale et moderne du Centre-Ouest de la France (11e–17e siècle), chrono-typologie de la céramique et approvisionnement de la vallée de la Loire moyenne, supplément à la *Revue Archéologique du Centre de la France* 20, 110. et un CdRom. Tours: FERAC.

Rodier, Xavier and Laure Saligny (2007). Modélisation des objets urbains pour l'étude des dynamiques urbaines dans la longue durée. In: M. Batton-Hubert, T. Joliveau and S. Lardon (dir.) *SAGEO 2007. Rencontres internationales Géomatique et territoire*. CdRom. Série : Aménagement du territoire. AgroParisTech-ENGREF Clermont-Ferrand. http://www.emse.fr/site/SAGEO2007/CDROM/p34.pdf

Rodier, Xavier and Laure Saligny (2008). Social features, Spatial features, Time features: An urban archaeological data model. In: A. Posluschny, K. Lambers, and I. Hertog (eds.) *Layers of Perception, Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA) Berlin, Germany, April 2–6, 2007*. Bonn: Kolloquien zur Vor- und Frühgeschichte, vol. 10, 2008.

Zadora-Rio, Elisabeth (1994). Le système de gestion des données de fouilles de Rigny. *Le médiéviste et l'ordinateur* 29, 31–35. http://lemo.irht.cnrs.fr/29/mo2913.htm