A Computer-Aided System for Dynamic Pottery Classification Using XML

Maria Bonghi Jovino, Giovanna Bagnasco Gianni, Lucio G. Perego,
Elisa Bertino, Pietro Mazzoleni and Stefano Valtolina

Dipartimento Scienze dell'Antichità, Università di Milano

Abstract. The present paper focuses on the aspects concerning the improvement of the classification and the recording techniques of archaeological artifacts.
We propose a solution in which XML is used to build a computer-based system useful to group findings into a classification which can dynamically change over the time. The paper finally discusses the results achieved applying this methodology to the results of the excavations of the University of Milan at the archaeological site of Tarquinia (Italy).

1. Introduction

Traditionally, in every archeological site different dictionaries and structures of finds classification are used. Ad hoc solutions in which findings are assigned t fixed categories are usually applied according to traditional database techniques. Nevertheless, recent studies (Bonghi Jovino and Chiaramopnte Treré 1997; Chiaramopnte Treré 1999; Bonghi Jovino 2001) demonstrate how these structures are not static, instead they can be changed along with the contribution of the results of different kind of analysis. Recent chemical analysis, continuous extraction of new finds, and information coming from different sites, are pointing out that a more accurate and flexible classification system is required.
In the paper we introduce the problem and we describe how XML can be used to deploy dynamic classification schemas. Afterwards, we describe the architecture of the developed tool bound to help archeologists in going through the continuous improving of their classification criteria and in recovering similarities among findings previously classified. The paper is concluded with an outline of the current work and with the presentation of the results achieved using our system to organize finds extracted from the archeological site of Tarquinia (Italy).

2. The Example of Tarquinia and the Excavations of the University of Milan

The debate upon the problem of classification hereby presented broadens and develops furthermore the multiform investigations carried out since 1982, by the University of Milan, under the direction of Maria Bonghi Jovino, in the most outstanding areas of the ancient city of Tarquinia (Viterbo – Italy): the so called "Monumental Compound" and the Ara della Regina sanctuary.
As far as the "Monumental Compound" is concerned, the investigations have mainly explored the different phases of the development of the community of Tarquinia around a natural phenomenon constituted by a creak in the rock around which testimonies of the "sacred" and the "political" aspects have been brought into light for eight centuries since the 10th century B.C. The analysis, archaeologically speaking, has encouraged to extensively confront the different phases not only inside, in their inner values in the "compound", but also outside, with other testimonies of Etruscan area.
The results of the excavations and the investigations of the various disciplines included in the research have been fully published in the volumes of the serie Tarchna [1,2,3], direstedby Maria Bonghi Jovino. After the first volume devoted to the digging stratigraphy, the other two have been dedicated to the study of wares, steadily under the control of the scientific and naturalistic analysis. Important topics such as the relationship between local productions and imported wares have been investigated.
As far as the Ara della Regina sanctuary is concerned, during the last two decades intensive searches and philological investigations have taken place both inside and outside the "Cavalli Alati" Temple, around the south western corner of the basement.

2.1 The Problem of Classification

Since the beginning, in order to deal with such a complicated heritage and the connections existing among data, the archaeologists have been concentrated on the registration, management and analysis of structures and mobile finds, that have been collected more and more numerous during the excavations.
As it usually happens in an inhabited site, the most critical aspect is the study of pottery finds, since they are always (or almost always) very fragmentary. As a matter of fact the life of an inhabited site, either long or brief, hinders the preservation of the pots integrity. In the reality most fragments can generally be classified as "open shapes" or "closed shapes", that are not definable: in this case the external comparison, with necropolis sites where pots in their integrity are found, can create virtual reconstructions, which haven't

support in the concrete reality of the excavation. This compels the archaeologists to a complex approach, from the picking of the objects to the possible reconstruction of the pots shape.
This is the reason why the attention of the archeologists of the University of Milan has always been directed to the philological analysis of the finds and to their reconstruction within the context to whom the finds themselves belong.
The assistance of personal computers has been fundamental from the beginning, in order to create computer-aided archives suitable for registering and fast tracking all collected data and for recovering the links existing between them, through the interrogation of the structured files.
At the beginning the solution was found using a traditional relational database based on aprioristic categories. After many excavation campaigns and classification stages, this system has yet turned out to be too rigid, because it led to a slavish application of typological criteria of classification. The risk assessed was to overwhelm the reality of the documentation in favour of the logics of classification.
So a critical position came to light: to be equidistant on the one hand from the pursuit of the reconstruction of complete shapes, referring to necropolis sites, and on the other hand from a rigid application of the logics of classification, since both leading at leveling the ancient reality.
In order to maintain a continuous dialectic between the traditional studies and the innovating tools the purposes sorted out after many years of experience are:

- to interpret the documentation as it is, preserving its fragmentary character
- to reconstruct the reality within the context in order to preserve data
- to distinguish the conceptual levels between the classification and the internal analysis of the testimonial series.

In conclusion, our challenge was to find out a classification system allowing, through subsequent passages, to examine several aspects and to frame each fragment in morphological and typological groups (that will be the object of new analysis).
Some criteria are outstanding, e.g. the vase function, the surface (presence or absence of the decoration), the shape references, the cultural influences (recognized by the shapes and the decorations of the vases), the production (referred also to the aspect of the ceramic body), the chronological valences.
Therefore, a pottery find is recognized at first on the basis of the following headings:

1 “production” (e.g. “tarquinian” vs. other productions);
2 “category” (e.g. “ceramics” or “metallic objects");
3 “class” (a homogeneous whole of finds as far as morphological, functional and chemical and phisical characteristics are concerned);
4 “group” (a homogeneous whole of finds as far as typical characters can be immediately perceived);
5 “sub-class” (a homogeneous whole of finds as far as chronological, morphological, stylistical and functional characters are concerned).

Nevertheless these headings can dynamically change according to the new results of the reasearch. For example a “group” or a “sub-class” can be differently organised as soon as new peculiarities of the documentation come to light, as indicative archaeological features of a more articulated reality than what expected at the beginning of the process of classification.
As far as the reconstruction of the vases shapes is concerned, as already stated, the single finds can be insufficient in order to create a complete typology. Therefore, in order to avoid attribution mistakes (which are inevitable if we try to recognize an exact vase typology failing all diagnostic elements), we have developed a parameter which allows to refer all fragments to concrete “models”, easily identifiable during the excavation procedures from which the fragments themselves come: so the concept of “Pivot” ("Capofila") has been created.

2.2 The Concept of “Pivot”

Each pottery find that we collect in the excavated site has particular morphological characters: these characters are peculiar to the fragment itself, and just the fragmentary nature of the piece allows to reconstruct only the morphology of a part of a shape to whom the fragment has been referred.
This situation is worded through the concept of “Pivot”, commonly used in editing the excavations (Chiaramopnte Treré 1999; Bonghi Jovino 2001): it individualizes an element which, within the reality of the excavation, allows to classify the evidences without preconceived selections, based on complete finds: this parameter maintains the descriptive phase, even if it gives this phase a code that allows a structured rearrangement and, therefore, the creation of a database easily consultable.
So the Pivot represents, within the documentation, an unicum, that in the excavated site and at the time of its discovery has no comparison. This can concern, for example, the shape to whom the Pivot is referred, the surface, and the ceramic body. As far as these characters are concerned, to the Pivot can therefore be assigned n fragments endowed with the same characters, so that it's possible to define some “families”, in order to base on them the analysis of the archaeological reality of the site. These families can change with the progress of the works and the more careful investigation of the philological analysis.
The structure of classification which needs to be created has to take into account the variability of the available documentation, because each find can become a Pivot and so it can increase the number of morphological “families”: this structure always assures an extreme flexibility, able to adapt itself to new discoveries.
As a matter of fact, the continuation of the researches can bring us to considerable changes within the outline of the documentation that we have reconstructed: for example, the typological families, which are previously defined, can be subdivide into several groups, as regards the discovery of new Pivots or the re-interpretation of some finds formerly classified as Pivots; on the other hand, whole categories can be modified, thanks to the support of chemical or physical analysis: for example, they could report, on the basis of some characters that cannot be recognized with the naked eye, a non-local production; in this case, the archaeologists can create new “families”, each referred to a set of Pivots, which allow to broaden the knowledge of the ancient site.

2.3 Example

A very good example is given by the so called “depurated pottery”, whose surface is treated in different ways so that several sub-classes can be defined. For instance, we can enumerate the “acroma” (achromatic) and the “a bande” (decorated by stripes) (Bagnasco Gianni 1999).
In the first case, fragments which have no cover, or a cover that has the same colour of the ceramic body, are inserted in the sub-class of “ceramica depurata”.
In the second case, fragments which have a cover that is distributed by stripes on their surface, are inserted in the sub-class of “ceramica a bande”.
However, such a classification alone doesn’t allow to distinguish these sub-classes in a convincing way: as a matter of fact, the information we can get are determined by the dimensions of the fragments, which allow to reconstruct the decorative effect of the vase to whom the fragments belonged. In fact, a little fragment, even if it has an uniform cover, can belong either to a “a bande” vase or to a pot with total cover or to a vase with a figurative decoration.
Therefore, if the only element we can get is the ceramic body, it's difficult to assign the fragment to a particular category, because there could be no difference between the different decorations of a same shape.
It’s obvious that this situation affects the valuation of the incidence of the finds we are studying, either as regards their distribution into both sub-classes, or (if the finds are very fragmentary) as regards their links with other group of materials with a ceramic body of “depurated” kind.
Nevertheless, the continuation of the researches and a more careful knowledge of the links between the shapes and their classes can help us to exactly recognize the “individual”, even if we begin from a little fragment; now, on the contrary, the situation of the research doesn’t allow to obtain definitions more accurate than a generic and arbitrary belonging to sub-categories of “acroma” or “a bande” pottery.

3. Dynamic Classification Schema

In order to support the classification methodology proposed and to provide an efficient authoring tool to the archeologists who intend to organize their findings on such criteria, we therefore had to built a flexible schema that respected the fragmentary reality bound to change and deal with dynamic changes in organization.
In this section, we present several alternative solutions we have studied and we describe their impact over the systems today used into the excavation sites to classify and to study the finds.
In fact, one of the main goals is to smoothly adapt the systems normally used into the excavation sites to a more dynamic classification, avoiding to build a new system. This in order to maintain previous investments in technologies and to guarantee an easier adaptation to the new methodology.

3.1 Classification using Relational Database

As introduced into Section 1, in the excavation site of Tarquinia a system based on a well known relational database (i.e., Microsoft Access®) was in use to store and classify findings. Without entering too much into the technical details, the findings were organized into a single table composed by information identifying the finding (e.g., identification number, dimensions, color), information (or attributes) about the discovery (e.g., date of discovery, name of the archeologists) and information concerning the class the finding belongs to.
Although this solution seems at first reflecting the classification described in Section 1, it presents one main problem: the solution does not guarantee data consistency, since the table doesn’t preserve the logical dependencies among data. In other words, an archeologist could specify a finding and add it to the database without noticing it logically incorrect according to the classification schema given in Section 1.1.

Id-number	...	Class	Group	Sub-Class

Fig. 1. Shows an extract of the table with some of the attributes used to classify the findings.

In fact, in order to support those consistency checks, the database should be enhanced by an expert programmer with ad hoc applications which should be periodically rewritten or modified as often as the classification schema changes. Moreover, the organization proposed in Figure 1 cannot be easily extended to support Pivots changing over the time, without the work of a person with advanced knowledge of databases.
To solve the problem of data consistency, a possible solution would have been re-organizing data into multiple tables like the ones shown in Figure 2.
Using such a multi-table organization, the system can automatically check for inconsistencies without building ad hoc external programs. In fact, the logical dependencies among data are specified into the tables (Figure 2 shows all the combination between Class and Group which are logically correct). Using multiple tables to represent the classification will therefore solve the problem of checking for data inconsistency. However, the solution keeps presenting several problems. First, it might lead to an explosion of the number of tables

Class	Value
Cl1	Impasto
Cl2	Bucchero
Cl3	Depurata

Group	Value
Gr1	Acroma
Gr2	Dipinta

ClandGr	Class	Group
ClandGr1	Cl1	Gp1
ClandGr2	Cl1	Gp2
ClandGr3	Cl2	Gp3

Fig. 2. Multiple-Tables classification.

when the classification schema grows. Second and more important, a multi-table classification is rigid and it cannot be easily modified to support dynamic changes. Finally, the solution could not be exported into other excavation sites which are adopting the single-table schema (like the site of Tarquinia) without heavily modifying their entire structure.
Single-Table and Multiple-Tables solutions pointed out the need of a system which does not modify the existing database schemas and which however is flexible enough to support dynamic changes. Another very important factor is that the above solutions do not support Pivots which are the basic components of the classification methodology proposed.

3.2 Classification using XML

XML – Extensible Markup Language – is a well known formalism today used to storing and interchanging any type of data, including archeological information (Crescioli, D'Andrea and Niccolucci 2002; Niccolucci 2002; Schloen 2001; Vezzoso, Bocchini, Locati and Romagnoli 2001).
There are several recognized advantages in the use of XML technology: among the others there is the flexibility of the data organization which guarantees data usability from different systems, the possibility to look and to modify data by any text editor, and the compatibility with all new database technologies.
We adopt XML to build a formal classification schema which specifies not only the hierarchy among attributes, but also the possible values that at a certain point of the hierarchy, an attribute can assume. Figure 3 shows an example of the classification schema created.

```
<Find>
<Production> name=tarquinian, descr="…"
   <Category> name=ceramica descr="…"
      <Class> name=impasto descr="…"
      <Group> name=dipinta descr="…"
         <Sub-cl> name= villanoviana descr="…"</Sub-cl>
      </Group>
      <Group> name="acroma" descr="…"
      <Sub-class> name= villanoviana descr="…" <Sub-cl>
      </Group>
   </Category>
</Find>
```

Fig. 3. XML schema for findings.

With this formal schema, there isn't any possibility of adding into the database a finding with an inconsistent classification because all the possible combinations are coded into the organizational schema. A simple and unique application has been created to parse the XML file and automatically check all the possible inconsistencies. Therefore, if a user inserts a finding having a wrong (or incomplete) classification, the system will automatically identify the inconsistency and will require the user to revise the classification. Moreover, in case the classification changes, the schema can be easily changed directly by archeologists by using a normal text editor without modifying the existing database schema.
XML is also suitable to include the notion of Pivot. In order to add Pivots into the classification, we had to take into consideration that Pivots are not just another abstract "heading" like "Production", "Category", "Class", and the other elements defined into Section 1.1. Instead, a Pivot is a fragment itself, and it has some features which are shared among all its' "family members", whereas some other attributes are proper of the Pivot and not shared with the others.
To give an example, a find coded aa1-3, which has been identified as a Pivot, has some proper features (such as color brown and dimension equal to 40x10mm) which are not exported to its family member whereas it is correct to say that its image (which is taken manually by the archeologist when a Pivot is extracted from the ground), can be also associated to all the fragments associated to it.
Therefore, we decided to build a schema where Pivots are maintained into the classification at the same level as other findings. Nevertheless only attributes identifying the Pivot, such as the identification number and the reference to its image, are maintained separately and are referred to by all findings associated to that particular Pivot.
Figure 4 shows an example of this classification where fragment aa1-3 has been classified as Pivot for the findings aa3-21, aa21-12, and obviously aa1-3.

```
<Pivot inv_num =aa1-3 image=aa1-3.jpg>
   <fragment inv_num = aa3-21
      surface = 12 Ceramic_body = 23 … >
   <fragment code=aa1-3
      surface = 3 Ceramic_body = 13… >
   <fragment code = aa21-12
      surface = 9 Ceramic_body = 20 … >
</Pivot>
```

Fig. 4. XML schema to support Pivots.

Using such an organization, not only fragments can be easily associated to one of the existing Pivots, but the classification can also evolve over the time whenever the discovery of new finds or the results of chemical analysis require to revise the previous assumptions. This can happen thanks to the possibility of adding/removing the categories and re-organizing the existing ones, by only typing the modification into the schema using any text editor.

4. Implementation

Figure 5 shows a screenshot of the implementation of our classification system built for the archeologists working in the Etruscan site of Tarquinia (Italy) in order to properly select a Pivot. As you can notice from the figure, the application supports multiple comparison criteria including visually comparison among stored images and drawings, through which the users are helped to identify the proper category and to achieve the various categories employed in a specific archaeological site.
Moreover, the graphical user interface is not generated in advance but it is dynamically created starting from the information loaded into the XML schema. This feature guarantees applicability of our system on top of existing classification schema.

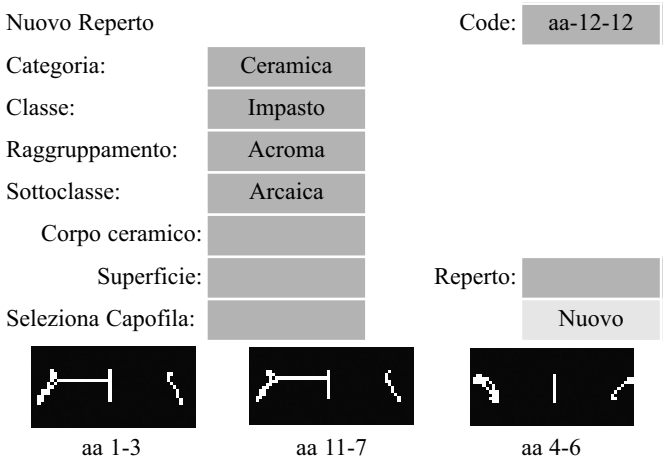

Fig. 5. Tool to classify new findings.

5. Conclusion

The paper presents the work done in the excavation site of Tarquina to support a dynamic classification methodology for findings and to improve recording techniques daily used during the excavation process.

Our system extends the current state of the art in two directions. First, the system makes use of XML to create a tool which visually supports dynamic changes on the classification schema. The second advantage is the wizard application which assists archaeologists in their training cataloguing artefacts, as well as external users.

Our system helps master students and foreign researchers working on the site in easily acquiring the archaeological classification. Moreover, scholars in charge of the excavations can easily control the process as well as update the categories organization, according to the conditions previously described. At present, we are extending our tool in order to semi-automatically discover similarities among dictionaries used in different archeological sites thorugh the use of ontologies.

Acknowledgments

The work of P. Mazzoleni and S. Valtolina is partially funded by DELOS Research Network.

References

Bonghi Jovino, M. and Chiaramonte Treré, C. (eds), 1997. *Tarchna I. Tarquinia. Testimonianze archeologiche e ricostruzione storica. Scavi sistematici nell'abitato (campagne 1982-1988)*. Roma.

Chiaramonte Treré, C. (ed.), 1999. *Tarchna II. Tarquinia. Scavi sistematici nell'abitato. Campagne 1982-88. I materiali 1*. Roma.

Bonghi Jovino, M. (ed.), 2001. *Tarchna III, Tarquinia. Scavi sistematici nell'abitato. Campagne 1982-88. I materiali 2*. Roma.

Bagnasco Gianni, G. 1999. La ceramica etrusca depurata acroma e a bande. In Chiaramonte Treré 1999, 99-176.

Crescioli, M., D'Andrea, A., and Niccolucci, F., 2002. XML Encoding of Archaeological Unstructured Data. *Proceedings of CAA 2001*. Oxford.

Niccolucci, F., 2002. XML and the Future of Humanities Computing In *Proceedings of ACM SIGAPP Applied Computing*.

Schloen, D., 2001. *XSTAR: XML System for Textual and Archaeological Research.* Technical Report University of Chicago, Oriental Institute. November 2001.

Vezzoso, A., Bocchini, O., Locati, S. and Romagnoli, A. 2001. *Managing XML Data sets through an XML native database Proceeding of VAST2001*. Athens. http://www.w3.org/XML