Guidelines for the Archaeological Map of Tuscany: some issues and experiences

Marinella Pasquinucci (*) *and Oreste Signore* (#)

(*) Università degli Studi di Pisa, Dipartimento di Scienze Storiche del Mondo Antico, Via Galvani 1, 56126 Pisa (Italy)
E-mail: pasquinucci@sta.unipi.it
(#) CNR - Istituto CNUCE - Area della Ricerca di Pisa - San Cataldo, Via G. Moruzzi, 1 – 56124 Pisa (Italy)
E-mail: oreste.signore@cnuce.cnr.it

ABSTRACT
Regione Toscana recently delivered guidelines for the archaeological map of Tuscany in the aim of fulfilling both administrative and research needs. Format proposed in 1995 by CIDOC has been slightly modified and enhanced, especially to represent chronological data in a more compact and formal way, and to add room for additional information. Wherever possible, normalisation of language is achieved by controlled dictionaries or thesauri. Main goals have been portability, scientific correctness, easy of use, coherence with national and international standards. Putting information in the right spatio-temporal context has been addressed as an important goal. In light of portability of data, and considering some limitations of traditional GIS, we decided to separate basic cartographic information from description of sites, that are georeferenced and have their own geometric properties stored as attributes. Comparison with national standards showed that there are many similarities, but some information is more adequately represented in our layout. Anyway, the two formats can generally be reconciled. Document is conceived as a structured document. To accommodate multiple or uncertain values, and to reduce atomization of information, in some cases syntactic rules drive writing of field content, so making the layout more flexible, readable and usable. The whole system is based upon well-established technologies and methods (DBMS, IRS, GIS), while Hypermedia approach and XML are used as unifying technologies. A test has been conducted upon data from Provincia di Livorno.

KEYWORDS: archaeological map, documentation form.

INTRODUCTION

Archeological maps have been widely addressed as a relevant tool for understanding and planning the territory. However, scholars' efforts and financial investments can be wasted as proprietary software solutions or differences in data organization often make impossible to effectively exchange, understand and use collected data. Hence the need of bringing together flexibility, scholars' and administrations' requirements, portability.

Methodological issues

Regione Toscana delivered guidelines for the archaeological map of Tuscany in the aim of fulfilling both administrative and research needs ([5]).

Archaeological maps are produced for study/research, conservation and administration purposes. These thematic maps are valid and predictive of archaeological sites when worked out on the ground of methodologically correct fieldwork, and detailed sites- and artefacts study. Data are to be collected in a diachronical perspective, using a multidisciplinary integrated approach, including:

- geologic, geomorphologic, paleoenvironmental and pedological research;
- non destructive techniques applied to landscape and maritime archaeology (remote sensing, geophysical and geochemical research);
- archaeological surveys and monitoring;
- stratigraphic excavations in at risk or scientifically relevant sites;
- archaeological and archaeometric study of finds;
- study of epigraphic and literary sources, documents and maps;
- bioarchaeologic and environmental analyses;
- study of toponyms.

Thanks to this research method sites and finds (including the underwater ones) are documented in their environmental and stratigraphic setting. Moreover, settlement patterns, territory exploitation (off site and intersite, agricultural and pastoral activities, woodland use), manufacturing activities and infrastructure works are diachronically identified and can be properly mapped. The importance of the geographic and geomorphologic setting of dated sites and finds has to be stressed, as prerequisite for the archaeological prediction. The archaeological map of Tuscany is also planned to include data about not georeferenced finds mentioned in documents and/or bibliography, in order to enhance its archaeological prediction potential.

TECHNOLOGICAL ISSUES

Main goal has been to implement a framework *easy to use*, coherent with international and national *standards*, *scientifically correct*. Software and hardware platform independence was an additional and relevant issue.
Many projects in the field are somehow similar, but there are differences that make results incompatible, so wasting scholars' efforts. Differences are sometimes in the technological platform, but the most relevant, difficult to overcome, arise from differences in *structuring* and *representing information*. It must be noted that in many cases there is an agreement on the semantics of single information fields, but there are differences in terminology and in document structure.
A relevant issue is the emphasis that many projects put on the cartography. Cartography is a basic and irrenonceable instrument for archaeological research, but its importance mainly results in the processing and tuning of models. Therefore, maps should be seen as one, among many others, interaction metaphors. In many cases, we can get similar results using low cost, platform independent tools. In essence, to produce an archaeological map, we need to correctly represent and describe object having *descriptive*, *geometrical* and *spatial* properties. To adequately represent descriptive information, we need an appropriate information structure and a clearly defined vocabulary, where every information item has its own semantics. GIS are adequate to represent spatial and geometrical properties, but lack in representing temporal and 3D data, while the spatio-temporal reasoning is

essential in the field ([4]).

To adequately support interchange of information among different project, at national or international level, *standards* play an essential role. Conformance to the standards also helps in saving of investments. However, we must stress that a *conceptual* standard (that is, an agreement on the semantics of information fields) is the real key point, and assures effective information exchange and unambiguous interpretation ([7]). In this project, we deeply considered both the standard issued by Istituto Centrale per il Catalogo e la Documentazione (ICCD) of Italian Ministry "Beni e Attività Culturali" ([3]) as well as the format ([1]) proposed by CIDOC (Comité international pour la documentation). A strong point is that the ICCD standard is obviously a must for any national project. The two resulted essentially similar in the organisation: the card is articulated in three levels of elements, each of them can be repeated several times, as needed. The main difference is in the high level of fragmentation in the ICCD schema, while CIDOC is a little more agile. Many information items are in common, but sometimes they are grouped in a different way.

As the project was aimed to administrative purposes, too, we remained a little more adherents to the CIDOC schema, certainly checking conformance to the ICCD schema. As final result we got a schema that is compatible with the ICCD one:

- some fields are absolutely equal;
- date fields are expressed in a much more simple and expressive way (for reference, see [11]);
- the more articulated fields in the ICCD format can be extracted from our schema with an appropriate post-processing;
- where we have real differences, maybe the ICCD schema should be amended (multiple occurrences of some fields, need for representing geometrical info).

We consider particularly significant some aspects:

- in the proposed schema, for each site we have the site geometry, too, specifying its shape (point, line, polygon) and recording the relevant coordinates (x,y,z). In this way, we have a sharp distinction between the basic cartographic information and the geometric properties of the archaeological evidence. The net result is high portability across different HW/SW platforms and openness to emerging new standards and technologies (e.g. SVG).
- Language normalisation is achieved making use of dictionaries and thesauri. Considering the difficulty in obtaining an effective communication when terms are too specific, requiring an in depth knowledge and, perhaps more important, a complete agreement on the conceptual organisation of information, we limited the thesauri depth to a maximum of three levels.

The card structure

Usually, sites are described using a "*card*", which contains the fields relevant to store the appropriate information. Local cultural and scholar standards normally dictates card layout. The card used in this project is organised as a *structured document*, ready to be marked according the new emerging XML standard. Some sections are mandatory, others are optional; in many cases, this property is dependent on the context, as a section is considered mandatory only if it is absolutely impossible

to omit the information. A mandatory field in an optional section just means that the field is required if the section is going to be filled. This approach gives the opportunity of rescuing data from previous databases, and possibly having an incremental filling of the card, as information will be available.
It is easy to understand that this choice has been done in light of having a flexible environment, avoiding the common problem of unfilled cards just because some detail was not available at the first compilation phase.
There are cases when it is impossible to fill in a single value, but we need two or more values from the relevant dictionary or thesaurus. In other cases, it can be necessary to express uncertainty between two values. In both cases, appropriate syntactic writing rules accommodate these needs.
The card is articulated in seven sections, which in turn are structured in subsections and fields:

1. *Site name and identification*. This section is mandatory, and contains site identification codes, reference to other cards, archaeological collections, objects, interventions.

2. *Localisation*. This section is mandatory, and is organised in several subsections. They will be filled or not, in different combinations, depending on the possibility of having an exact geographic localisation or an approximate one.

3. *Type and chronology*. Mandatory section, can be repeated in case of diachronic evolution of the site. Compiler can specify exact or generic (periods) dates. Site is interpreted according to its shape or function, using terms semantically arranged in a thesaurus including the expression "undetermined".

4. *Conservation status*. Optional, can be repeated, and is used to document the conservation status and date of information gathering.

5. *Juridical/administrative condition*. Optional, can be repeated, and is used to document juridical and administrative condition, its originator, type and dates.

6. *Toponym/Site name*. Optional, can be repeated, and is used to report ancient toponyms and their existence time range.

7. *Site description*. Section used to synthetically describe the site, using terms included in a controlled dictionary. Description will contain: geographic context, archaelogical evidence, interpretation, chronological clues and any additional info.

This last section will be object of new and hopefully interesting developments, as we plan to introduce in it a "*light markup*" to keep all the flexibility and richness of free text, clearly identifying, in the same time, relevant semantic items. These items could be used to facilitate the hypertextual association among documents, triggering appropriate interaction paradigms, like map, classification models, time ([[8], [9], [10]).

Technologies

There are some well-established and robust technologies that constitute the frame of reference for every application in the field. In short, they are DBMS, IRS and GIS.
DBMS is really powerful in representing complex relationships and structures, and is the hearth of any information system. However, this approach requires high formalisation and atom-

ised information items. This can produce forcing that result in semantic poorness, evident in a semantically rich field as the cultural heritage. IRS keep the semantic richness of natural language, but some lexical problems contribute to maintain a certain level of ambiguity, causing low precision and recall factors. Perhaps more important, they are unable to manage complex relationships among information items, as the IRS document model is poorly structured. GIS are the most suitable tools to manage geometric and spatial information, supporting queries on these properties. They also allow the processing of spatial models. Their counterpart is that they make use of proprietary models and software tools, and compatibility among them is scarce or lacking ([2], [6]).

Anyway, each one of these technologies exhibits strong and weak points: no system relying on only one of them can be completely satisfactory. On the other hand, there are some *unifying technologies* that can help in finding an appropriate solution, especially in the current technology trends, that aim to the implementation of totally web based information systems. We can briefly refer to them as hypermedia and XML technology family.

Hypermedia approach is the most natural user interface. It can be seen as an application interface towards underlying systems. In this way, it is possible to implement complex systems that make use of the basic technologies, and emphasise the associative aspect, so relevant in the specific field. We must therefore carefully consider the increasing diffusion of XML, the basic component of the next generation web. Markup of hypermedia documents can preserve the semantic richness of natural language, reducing, in the same time, some of its intrinsic ambiguities. In this way, it is possible to overcome some of the constraints typical of the DBMS approach, even if DBMS maintains its importance as the "glue" of information. It must be stressed that an appropriate markup (or DTD, in XML jargon) requires an in depth analysis of information items, and a clear understanding of their semantics and relationships.

The proposed schema is fully compatible with this technological trend.

A first test

In a first test phase, supported by the Provincia di Livorno, 445 complete cards concerning its territory, out of more than 600 available in not machine readable format, have been recorded, adding georeferencing data.

In Figure 1, 2 and 3 we report some examples that can help in understanding basic functionality and interaction style.

Figure 1 shows the usage of a dictionary in entering Section 1 data about a site that can't be checked again, because of the site has been cancelled by some human interventions ("*Sito asportato da eventi antropici*").

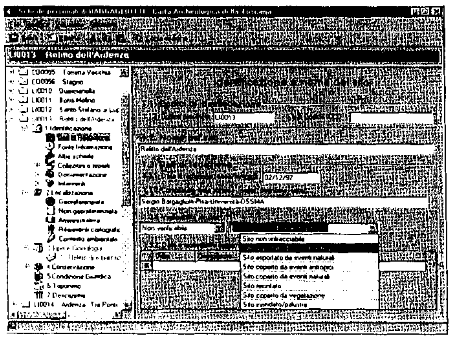

Figure 1: Entering data with help from a dictionary

Figure 2 shows as location data can be entered filling different combinations of subsections, and georeferenced data include the z coordinate (see fields 2.1.6, containing the geometry and 2.1.7 where co-ordinates can be entered).

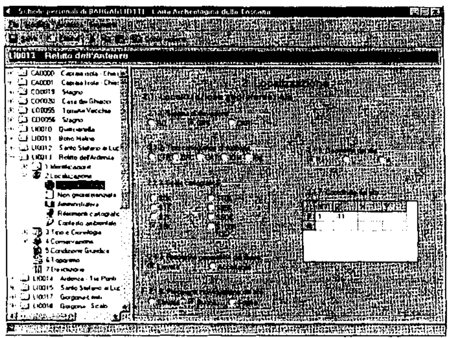

Figure 2: Coordinates include z.

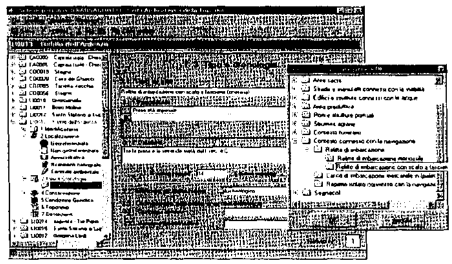

Figure 3: Browsing a thesaurus

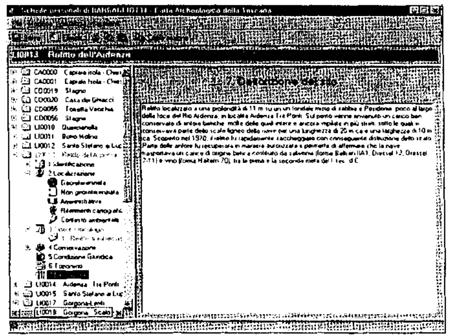

Figure 4: Free text section

In Figure 3 user can find the appropriate term browsing a thesaurus, looking for the exact name of a shipwreck type (*"Relitto di imbarcazione con scafo a fasciame (oneraria")*).
We stress that Section 7 of the card accommodates for a free text containing controlled terms, to semantically enrich the site description without constraining the user to fill a lot of fields (Figure 4).
During the experimental phase it was possible to verify the the model was adequate, obviously after some minor amendments of the first version. Therefore, during this first experimental phase it was possible to draw some indications about the adopted approach.
First of all, taking survey data and putting them into a common schema is useful to rethink previously collected data.

However, some effort is related to need of deeply entering into the system schema: this requires entering some tenth of "cards", to gather adequate experience.
In general, integration to the thesauri were very limited, as they were conceived with a large amount of terms, to accommodate for the foreseable scholars' needs. In essence, we had to insert some options to add to the term "*Other*" to enter information were data about the archaeological evidences do not allow to fully identify the appropriate term (e.g., sometimes the compiler is unable to state if an archaeological evidence is a "villa" or "necropoli", in such cases, (s)he can enter: "*villa/necropoli*").
Finally, the cultural dating (e.g. late republic, late empire, etc.) was used in conjunction with the absolute dating (expressed according [11]). However this task, mandatory for building appropriate historical maps, is really difficult, and is felt by the scholars as somehow too formal.

CONCLUSIONS

In defining the guidelines for the archaeological map of Tuscany, we considered both methodological and technological issues. We resulted in an approach that:

- Makes a sharp distinction among basic cartography and descriptive information, which includes all the elements needed to setting up position and geometry of the archaeological evidence.
- Emphasises role of dictionaries and thesauri, seen as the knowledge base shared between indexer and user.
- Defines a schema were number of fields is reasonable, scientific information can easily fit in, leaving room for insertion of semantically rich free (but somehow controlled) text.
- Is conformant to national and international standards.
- Is up to date with technology trends.

Is totally open to additional enhancements, and is portable in the Open Source Software environment.

Results from a first test phase were encouraging. We are going to check costs involved in migrating data gathered in other experiences to the new format, that can vary depending on the level of tightening to proprietary hardware and software environments.

ACKNOWLEDGEMENT

We warmly acknowledge members of the working group R. Francovich (Università di Siena), G. Nenci (Scuola Normale Superiore, Pisa), R. Palma di Cesnola and subsequently R. Gambassini (Università di Siena), M. Torelli (Università di Perugia), and staff members from ICCD, Soprintendenza Archeologica per la Toscana, Regione Toscana, who contributed to the work with their personal effort and were very effective in raising many methodological issues.
We also wish to thank the Regione Toscana, which promoted the initiative, and Provincia di Livorno, which supported the test phase.
S. Bargagliotti, A Del Rio and S. Menchelli where active partners in the whole development of the guidelines and their tuning after first tests.
Finally, G. Barletta, L. Cherubini, F. Cibecchini, M. Corsini, F. Fedeli, S. Meini and M. Vallebona contributed to record data.

REFERENCES

1. CIDOC (Comité international pour la documentation – Conseil

international des Musées, Groupe de travail sur les sites archéologiques (CIDOC-ICOM, Maison del l'Unesco) *Project de norme documentaire internationale pour les sites achéologiques*, 1995

2. Gillings M, Mattingly D., van Dalen J. *Geographical Information Systems & Landscape Archaeology*, Oxford Books, 1999

3. Papaldo S., Ruggeri Giove M., Gagliardi R., Matteucci D.R., Romano G.A., Signore O. *Proposta di strutturazione dei dati del catalogo: Beni mobili archeologici e storico-artistici.* (Edizione riveduta e aggiornata), Istituto Centrale per il Catalogo e la Documentazione (Roma) - Istituto CNUCE (Pisa), Multigrafica Editrice, Roma, 1988

4. Papaldo S., Signore O. *Un approccio metodologico per la realizzazione di una banca dati storico-geografica (A methodological approach to producing a historical/geographical databank)*, Multigrafica Editrice, Roma, 1989, ISBN 88-7597-105-6

5. Regione Toscana *Linee guida per la redazione della Carta Archeologica della Toscana*, (1998)

6. Sanjuán L. G., Wheathley D. W. The state of the Arc: differential rates of adoption of GIS for European heritage management. In *European Journal of Archaeology*, 2, (2), August 1999, 201-228

7. Signore O. From data structuring to data exchange: a simple path. In Marker, H. J. and Pagh, K. Eds. *Yesterday, Proceedings from the 6th International Conference Art History and Computing* (Odense, Denmark) August 28-30, 1991, Odense University press, 1994, 48-54, ISBN 87 7838 022 7

8. Signore O. Modelling Links in Hypertext/Hypermedia, in *Multimedia Computing and Museums, Selected papers from the Third International Conference on Hypermedia and Interactivity in Museums - (ICHIM'95 - MCN'95)* (San Diego, California, USA), October 9-13, 198-216, ISBN 1-88-5626-11-8

9. Signore O. Exploiting Navigation Capabilities in Hypertext/Hypermedia. In *HICSS-29 Annual Hawaii International Conference on System Science* (Maui, Hawaii) January 3-6, 1996, 165-175, ISBN 0-8186-7327-3, ISSN 1060-3425

10. Signore O., Bartoli R., Fresta G., Loffredo M. Implementing the Cognitive Layer of a Hypermedia. In Bearman, D. and Trant, J. Eds. *Museum Interactive Multimedia 1997: Cultural Heritage Systems Design and Interfaces – Selected papers from ICHIM 97 the Fourth International Conference on Hypermedia and Interactivity in Museums* (Paris, France) 3-5 September 1997, Archives & Museum Informatics 1997, 15-22, ISBN 1-885626-14-2

11. Signore O., Bartoli R., Fresta G., Marchetti A. Issues on historical geography. In Bearman, D. and Trant, J. Eds. *Museum Interactive Multimedia 1997: Cultural Heritage Systems Design and Interfaces – Selected papers from ICHIM 97 the Fourth International Conference on Hypermedia and Interactivity in Museums* (Paris, France) 3-5 September 1997, Archives & Museum Informatics 1997, 32-37, ISBN 1-885626-14-2

ABOUT THE AUTHORS

Marinella Pasquinucci is full professor of Ancient Topography, professor of Submarine Archaeology, professor of Ancient Topography in the PhD in Ancient History and in the PhD in Archaeology at the University of Pisa.
Her scientific interests are in landscape archaeology, ancient town planning, ports and calls, manufacturing activities and trade in Italy, from the hellenistic period to late Antiquity.
She published about 120 books and papers.
E-mail: *pasquinucci@sta.unipi.it*

Oreste Signore is full time staff member of CNUCE-CNR since 1973. His scientific interests are in the area of Information systems, covering hypermedia, user interface, web applications, information retrieval. He published more than 90 books and papers in national and international journals and conference proceedings. He had a relevant role in defining national standards for the cultural heritage. His present interests are in Semantic Web and adaptive systems.
E-mail: *oreste.signore@cnuce.cnr.it*